



How to discover QCD Instantons at the LHC

Simone Amoroso¹ , Deepak Kar², Matthias Schott^{3,a}

¹ DESY, Hamburg, Germany

² University of Witwatersrand, Johannesburg, South Africa

³ Johannes Gutenberg-University, Mainz, Germany

Received: 8 June 2021 / Accepted: 5 July 2021 / Published online: 17 July 2021
© The Author(s) 2021

Abstract The Standard Model of particle physics predicts the existence of quantum tunnelling processes across topological inequivalent vacua, commonly known as Instantons. In Quantum Chromodynamics, these Instantons play a fundamental role in explaining much of the theory long-distance behaviour. However, they have not yet been observed experimentally. Their direct observation would mark a breakthrough in modern particle physics, shedding light on our fundamental understanding of the non perturbative dynamics in the Standard Model. Recently, new calculations for QCD Instanton processes in proton–proton collisions became public, suggesting sizeable cross sections as well as possible experimental signatures at the LHC. In this work, we explore possible analysis strategies for the LHC experiments to discover small-size QCD Instanton induced processes. Moreover, we derive a first limit on the Instanton production cross section using published data of Minimum Bias processes at $\sqrt{s} = 13$ TeV at the LHC.

Contents

1 Introduction	1
2 QCD Instanton processes at the LHC	3
2.1 Production of the instanton pseudo-particle in proton–proton collisions	3
2.2 Decay of the instanton pseudo-particle and experimental observables	4
3 MC samples, detector simulation and experimental observables	4
3.1 Soft QCD processes	4
3.2 Hard QCD and other processes	5
3.3 Signal simulation	6
3.4 Experimental observables	6
4 Search strategies	8
4.1 Very low instanton masses: the soft QCD regime	8

4.2 Medium instanton masses: the hard QCD regime	10
4.3 High instanton masses: the top quark regime	12
5 Limits on instanton processes in proton–proton collisions	15
5.1 How to mimic QCD instanton signatures	15
5.2 LHC sensitivity projections	15
5.3 First limit on instanton processes from proton–proton collision data	17
6 Conclusion	18
References	18

1 Introduction

Yang–Mills theories [1], embedded in the Standard Model (SM) of particle physics, form the basis of our understanding of the strong and electroweak interactions. The beauty and success of the SM lies in its predictive power, achieved in the weakly coupled regime by calculations based on perturbative approaches. Perturbation theory, developed in order to describe hadron collisions at high energies, relies on this small size of the strong coupling constant at high momentum transfer and short distance. The study of Quantum Chromodynamics (QCD) in the perturbative regime has seen tremendous advancements in the last decades. Hard scattering cross sections have been calculated up to the third order in the strong coupling [2, 3], and perturbative QCD predictions have been verified to incredible precision over many orders of magnitude of momentum transfer at high energy colliders [4]. At the same time a fundamental understanding of Yang–Mills theories in the strongly coupled limit is still lacking, and remains one of the biggest challenges for particle physics to date.

Yang–Mills theories exhibit a rich and non-trivial vacuum structure. In particular, they admit semi-classical solutions corresponding to fluctuations of the gauge fields across topologically non-equivalent vacua, known as Instantons [5]. These inherently non-perturbative phenomena are of great

^a e-mail: matthias.schott@cern.ch (corresponding author)

theoretical interest and have been linked to many fundamental aspects of QCD (introductory reviews on the physics of instantons can be found in [6–9]). Even though Instanton processes are a core prediction of the SM, they have never been experimentally observed.

The Yang–Mills vacuum structure is depicted in Fig. 1, showing the energy density of the gauge field as a function of the Chern–Simons (or winding) number, N_{CS} , which characterises the topological charge of a system. Instantons describe tunnelling transitions in Minkowski spacetime between classically degenerate vacua, which only differ by their winding number by one unit, i.e. $\Delta N_{CS} = 1$. Instanton solutions are not only localised in time, but also in space, i.e. they have a certain spatial extension. A second type of classical solutions, known as Sphalerons, corresponds to transitions from one vacuum by a half-integer winding number on top of the energy barrier (also shown in Fig. 1), where its static energy corresponds to the barrier height.

These tunnelling solutions differ significantly from ordinary solutions obtained in perturbation theory, where only field configurations corresponding to small changes of the vacuum field at $N_{CS} = 0$ are accessible and minima which cannot be obtained by continuous transformation of the gauge field are ignored. Instanton and Sphaleron solutions provide crucial ingredients for an understanding of a number of non-perturbative issues in the SM. In the electroweak theory, Instanton and Sphaleron transitions are associated with $B + L$ violation. They become highly relevant at high temperatures [7, 10, 11] and have a crucial impact on the evolution of the baryon and lepton asymmetries in the early universe (see also Ref. [12, 13] for a review). In QCD, these topological solutions have been argued to play an important role in various long-distance aspects of the theory. They provide a possible solution to the axial $U(1)$ problem [14] and are associated with chiral symmetry breaking [15–17].

The height of the energy barrier between two vacua, called Sphaleron mass M_{Sp} , in the electroweak theory is of the order $M_{Sp} \sim \frac{\pi}{\alpha \rho_{\text{eff}}} \sim \pi \frac{M_W}{\alpha_W} \sim 10 \text{ TeV}$ [18], where α_W is the weak coupling constant and ρ_{eff} the effective Instanton size. As the energy barrier is below the LHC center of mass

energy, one might think that electroweak Sphalerons should be produced, and could be observed at the LHC. The question whether manifestations of such topological fluctuations can be directly observed in high-energy experiments was already raised in the 1980s, in the context of the electroweak sector [19–22]. However, the difficulty of obtaining a coherent state makes these processes likely to remain unobservable at current and future colliders [23, 24]. The situation is different for QCD Instanton processes, for which the energy barrier height, $M_{Sp} \sim \frac{3\pi}{4\alpha_s \rho_{\text{eff}}} \sim Q$ [25], with α_s the strong coupling and the parameter Q related to the energy scale of the underlying process, can be as low as a few GeV. The phenomenology of Instanton production at colliders has been first developed in the context of Deep Inelastic ep Scattering at the HERA collider [26, 27]. Searches for Instanton processes have been performed by the ZEUS and H1 Collaborations [28–30] excluding the lower range of the predicted cross sections. It is then interesting to understand if these processes could also be measured at the LHC. Recent works have provided first calculations for LHC cross sections [31], and some discussions on the expected phenomenology [32]. In this work we explore in further details suitable analysis strategies at the LHC, in particular exploring the (relatively) small-size regime with Instanton masses of few tenths of a GeV, where the cross section is the highest. In this regime the challenge lies in finding suitable observables that, while retaining sensitivity to the soft decay products of the Instanton, can also be described to an acceptable level of accuracy by the phenomenological models of soft QCD activity.

The paper is structured as follows: In Sect. 2 we briefly provide a review on Instanton processes at the LHC, covering their expected production cross sections and experimental signature. This is followed by an overview of the Monte Carlo samples used in Sect. 3. Possible search strategies and the optimisation of the event selection are described in Sect. 4. The expected sensitivity of the proposed analysis, as well as first limit on QCD Instanton processes are presented in Sect. 5. The paper concludes in Sect. 6.

Fig. 1 Instanton and Sphaleron processes in the topology of a Yang–Mills vacuum; energy density of the gauge field (y-axis) vs. winding number N_{CS} (x-axis)

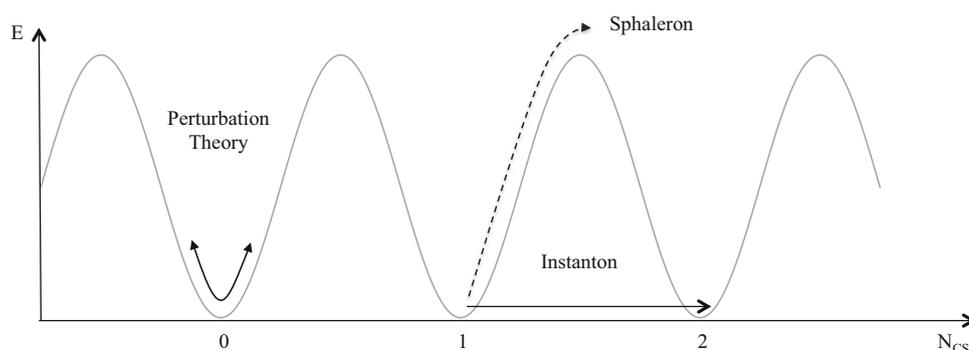
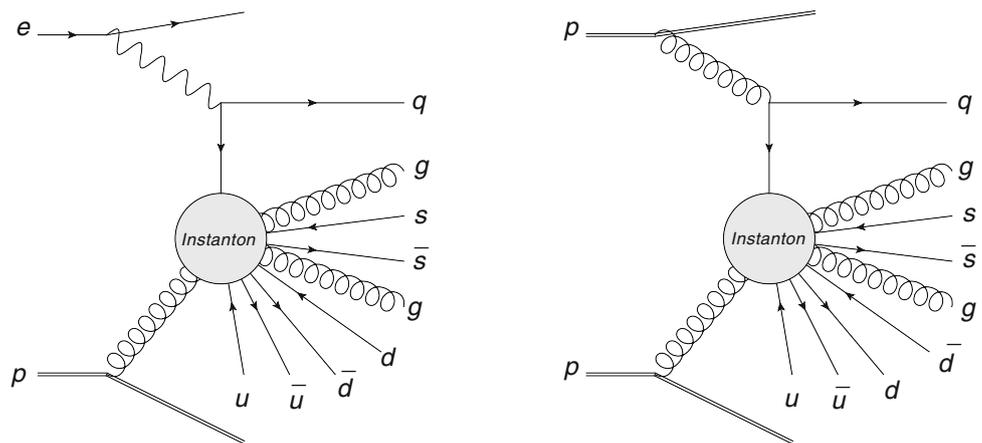


Fig. 2 Depiction of a QCD Instanton processes in electron–proton (left) and proton–proton (right) collisions, where an external scale parameter Q' is required



2 QCD Instanton processes at the LHC

2.1 Production of the instanton pseudo-particle in proton–proton collisions

Following the work of [25,33] can write the inclusive partonic cross section of QCD Instanton-induced processes in Instanton perturbation theory as

$$\sigma_{parton,parton}^{(I)} \sim \int_0^\infty \int_0^\infty D(\rho)D(\bar{\rho})d\rho d\bar{\rho} \cdot e^{-(\rho+\bar{\rho})Q'} \cdot e^{-\frac{4\pi}{\alpha} \Omega(E/M_I)} \cdot (\text{further terms}), \tag{2.1}$$

where $D(\rho)$ and $D(\bar{\rho})$ denote the Instanton and anti-Instanton size distributions, E the available energy of the process and Ω describes the Instanton anti-Instanton interaction, with $\Omega(x) = 1$ for $x \rightarrow 0$ and $\Omega(x) = 0$ for $x \rightarrow \infty$. The Instanton size distribution is proportional to $D(\rho) \sim \rho^{11-2/3n_f-5}$ [15,25,34–36], thus an integral over ρ would diverge. However, it was shown that the additional term $e^{-(\rho+\bar{\rho})Q'}$ has to be taken into account [25], where Q' describes a generic hard scale of the Instanton process. This form factor effect renders the ρ integration convergent. In order to make reliable calculations of cross sections in QCD, Instanton perturbation theory [25] has to be applied. This requires the validity of the diluted gas approximation [25], which requires that the extensions of Instantons and anti-Instantons are not overlapping. Therefore, the validity of Instanton perturbation theory requires Instantons to be sufficiently small, i.e. localised in space-time. In QCD, a generic hard scale Q of the underlying process can be chosen, reducing the Instanton size and justifying the diluted gas approximation and the Instanton perturbation theory approach. In this scenario, the cross sections can become sizeable at high energies. The reason for the large cross section can be intuitively understood [37] by changing the picture from a tunnelling between vacua at $E = 0$ to that of the actual creation of a Sphaleron-like configuration [18] on top of the potential barrier of height. Therefore, in a naive (but not fully correct)

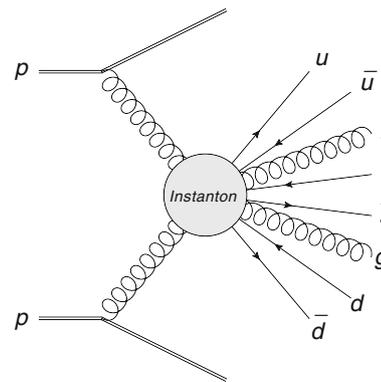


Fig. 3 Depiction of a QCD Instanton processes in proton–proton (right) collisions without the requirement of an external scale parameter Q'

picture, the Instanton process can be interpreted as the creation and the decay of a Sphaleron pseudo-particle, where the pass of the pseudo-particle depends directly on the height of potential barrier.

In deep inelastic scattering processes, the hard scale Q' was defined by a highly virtual momentum transfer by a photon, emitted by the incoming electron, yielding a highly energetic final state quark in addition to the Instanton process, as shown in Fig. 2. This concept can be easily transferred to proton–proton collisions where the photon exchange is simply replaced by a gluon as shown in Fig. 2. A first calculation of the latter processes was completed in [32], but requiring the associated adronic jet to be sufficiently energetic lead to negligible cross sections at LHC energies.

An alternative strategy to calculate cross sections was also recently published [31]. Here, a second independent kinematic scale, as the DIS highly virtual momentum scale Q , is not required (Fig. 3). In this approach, only small Instantons contribute to the scattering processes in QCD and potentially problematic contributions of Instantons with large size are automatically cut-off by the inclusion of quantum effects due to interactions of the hard initial states that generate the factor $e^{-\alpha_s \rho^2 s' \log s'}$. The latter reasoning was pre-

sented in 1991 [38,39] and provides a dynamical solution to the problem of IR divergences arising from Instantons of large scale-sizes in QCD. The scale invariance of the classical Yang–Mills theories is broken by those quantum effects which lead to a suppression of all but small Instantons with sizes $\rho \lesssim (10 - 30)/\sqrt{s'}$. The characteristic QCD Instanton size is therefore inversely proportional to the centre-of-mass (CoM) energy of two colliding partons $\sqrt{s'}$. Table 1 shows the proton–proton cross sections for Instanton processes at a center of mass energy of 13 TeV for various choices of minimal values of $\sqrt{s'_{\min}}$ as calculated in [31]. The cross section at $\sqrt{s'_{\min}} = 0$ GeV already contributes several percent to the total proton–proton cross section and at even lower values would saturate. This implies that the calculation breaks down at some small value of $\sqrt{s'_{\min}}$.

In the following it is assumed that the cross section estimate is reliable for $\sqrt{s'_{\min}} > 20$ GeV, keeping in mind that the uncertainty on this prediction could be of several order of magnitudes. The cross section dependence can be interpolated by a phenomenological formula $\sigma \sim e^{a \cdot e^{b \cdot x} + c \cdot x^2 + d \cdot x + e}$, implying an exponential decrease vs. $\sqrt{s'_{\min}}$. In the pseudo-particle picture of the Instanton process, the available energy $\sqrt{s'_{\min}}$ can be interpreted as the Instanton mass m_I . The production of Instantons is therefore not a resonant but a continuous processes, with large production rates for Instantons of low mass and small production rates for Instantons of high masses.

2.2 Decay of the instanton pseudo-particle and experimental observables

A QCD Instanton tunnelling process between $\Delta N_{CS} = 1$ vacua leads to the creation of exactly one quark–antiquark pair of different chirality for each flavour, N_f in association with a number n_g of additional gluons. Following the approach of [31] we only consider the dominant contribution from gluon–gluon initiated transition, which proceeds through the process:

$$g + g \rightarrow \sum_{f=1}^{N_f} (q_R^f + \bar{q}_L^f) + n_g g \quad (2.2)$$

In the pseudo-particle picture, this can be interpreted as a decay process of a Instanton pseudo-particle with a mass m_I . For low Instanton masses, e.g. in the 50 GeV range, we expect therefore an isotropic decay into up to 5 quarks, 5 anti-quarks as well as 5–10 gluons. The number of gluons is assumed to be Poisson distributed around an average, $\langle n_g \rangle$, which has been calculated in [31], and which in turn depends on m_I and varies between 5 and 13 over the mass range $10 \text{ GeV} < m_I < 4 \text{ TeV}$. As a consequence, QCD Instanton-induced scattering

processes produce *soft bombs* – a high-multiplicity spherically symmetric distributions of relatively soft particles [40]. These generic properties of QCD Instanton decays can be used to design experimental searches for these processes. The Instanton production cross section, reported in Table 1, falls rapidly with increasing mass of the Instanton pseudo-particle, m_I , or equivalently with the center of mass energy at parton level, $\sqrt{s'}$ of the Instanton process. The cross section dependence on $\sqrt{s'}$ is different from that of other SM processes. Experimentally, m_I can be approximated by the 4-vector sum of all reconstructed charged particles with a certain minimal transverse momentum.

Depending on the range of m_I considered, different regimes can be identified with different processes contributing as background and different signal to background ratios. While the decay of low mass Instantons, e.g. $m_I = 30$ GeV, results in events with a high multiplicity of low energetic charged particle tracks, for masses larger than 200 GeV jets of hadronic particles are produced. In the low mass regime, *soft-QCD* events originating from inelastic, non-diffractive processes constitute the dominant background due to their large production cross section. For higher Instanton masses, high- p_T jet production processes (*hard-QCD*), as well as vector-boson and top anti-top quark pair production in the hadronic decay channels contribute.

3 MC samples, detector simulation and experimental observables

In the following section we describe the event generation and detector simulation configuration used for the signal and background sample, summarised in Table 2, as well as the experimental observables investigated.

For all generated samples a typical detector response has been simulated through the DELPHES framework [41] with settings corresponding to the ATLAS experiment, and without considering additional pile-up interactions. Pile-up activity could become a non-negligible source of background when selecting events with high multiplicity final states, and will require dedicated studies in LHC analyses that cannot be performed with parametrised simulations.

3.1 Soft QCD processes

Soft and diffractive QCD events constitute the largest part of the hadronic cross-section when no high p_T object is identified. Due to their non-perturbative nature, *soft-QCD* processes are described by phenomenological models, typically based on a multiparton interaction (MPI) description, which have been tuned to data using a wide variety of reference measurements. Within this study, the *soft-QCD* processes in the PYTHIA8 [42] generator with the Monash tune [43] and the

Table 1 Overview of expected cross sections of Instanton processes with different values of $\sqrt{s'_{\min}}$ [31]

$\sqrt{s'_{\min}}$ (GeV)	10	20	50	100	200	500
$\sigma(pp \rightarrow I \rightarrow X)$ (pb)	1.7×10^{11}	6.3×10^9	4.1×10^7	8.0×10^4	1.1×10^2	3.5×10^{-3}

Table 2 Overview of the MC samples used to model the Instanton signal process and the SM background processes

Process	Generator	Main generator setting	# Events
QCD-Instanton (low-mass regime)	SHERPA	Instanton_MIN_MASS: 25.	10,000
QCD-Instanton (low-mass regime)	SHERPA	Instanton_MIN_MASS: 50.	10,000
QCD-Instanton (medium-mass regime)	SHERPA	Instanton_MIN_MASS: 100.	10,000
QCD-Instanton (medium-mass regime)	SHERPA	Instanton_MIN_MASS: 200.	10,000
QCD-Instanton (high-mass regime)	SHERPA	Instanton_MIN_MASS: 300.	10,000
QCD-Instanton (high-mass regime)	SHERPA	Instanton_MIN_MASS: 500.	10,000
QCD-Instanton (high-mass regime)	SHERPA	Instanton_MIN_MASS: 1000.	10,000
<i>softQCD</i>	PYTHIA8	SOFTQCD:ALL = ON - Monash Tune	1,000,000
<i>softQCD</i>	PYTHIA8	SOFTQCD:ALL = ON - A14 Tune	1,000,000
<i>softQCD</i>	SHERPA		1,000,000
<i>softQCD</i>	HERWIG		1,000,000
$qq \rightarrow X, qg \rightarrow X, gg \rightarrow X$ (<i>hardQCD</i>)	PYTHIA8	HARDQCD:ALL = ON PHASESPACE:P THATMIN = 5.	1,000,000
$qq \rightarrow X, qg \rightarrow X, gg \rightarrow X$ (<i>hardQCD</i>)	PYTHIA8	HARDQCDALL = ON PHASESPACE:P THATMIN = 100.	1,000,000
$qq \rightarrow X, qg \rightarrow X, gg \rightarrow X$ (<i>hardQCD</i>)	PYTHIA8	HARDQCDALL = ON PHASESPACE:P THATMIN = 300.	1,000,000
$W \rightarrow q\bar{q} + X$	PYTHIA8	WEAKSINGLEBOSON:FFBAR2W = ON	1,000,000
$Z \rightarrow q\bar{q} + X$	PYTHIA8	WEAKSINGLEBOSON:FFBAR2W = ON	1,000,000
$WW \rightarrow q\bar{q}q\bar{q} + X$	PYTHIA8		1,000,000
$t\bar{t} \rightarrow bq\bar{q} + \bar{b}q\bar{q} + X$	PYTHIA8	TOP:ALL = ON	1,000,000

NNPDF23LO PDF set [44] are used as the baseline. In total, one million events have been generated for proton–proton collisions at a center of mass energy of 13 TeV.

We also consider *softQCD* production in PYTHIA8 using a different shower and multi-parton interaction tune, A14 [45], as well as the predictions of *softQCD* processes from the HERWIG7 [46] and the SHERPA [47,48] event generators. We compare normalised distributions for a number of observables (defined in Sec. 3.4) of interest for Instanton processes in Fig. 4 for events with a reconstructed invariant mass based on reconstructed tracks between 20 GeV and 40 GeV. A good agreement is observed between the different predictions for most distributions, with the possible exception of the $N_{\text{Displaced}}$ distribution, where the SHERPA predictions differs by about 20% from the other generators. In the following, the maximal difference between the various *softQCD* samples is taken as systematic uncertainty on the nominal *softQCD* prediction taken from PYTHIA8.

3.2 Hard QCD and other processes

High- p_T jet production processes can be predicted with high accuracy in perturbation theory, with inclusive jet production known up to NNLO [1]. We use PYTHIA8 with the NNPDF23LO PDF set to simulate LO *hardQCD* di-jet production, with additional jets included through additional parton scatterings. The transition between *softQCD* and *hardQCD* processes is ambiguous and not well defined. In our study, we use the *softQCD* samples for all events, which have no jet at particle level with a transverse momentum above 20 GeV, while the simulation of *hardQCD* processes is used for all other events.

In addition to multi-jet final states, the production of top-quark pairs and of W and Z bosons and di-boson processes can also lead to final states with high multiplicity of particles, in particular in their fully hadronic decay channels. These processes are also simulated at LO with PYTHIA8 and the

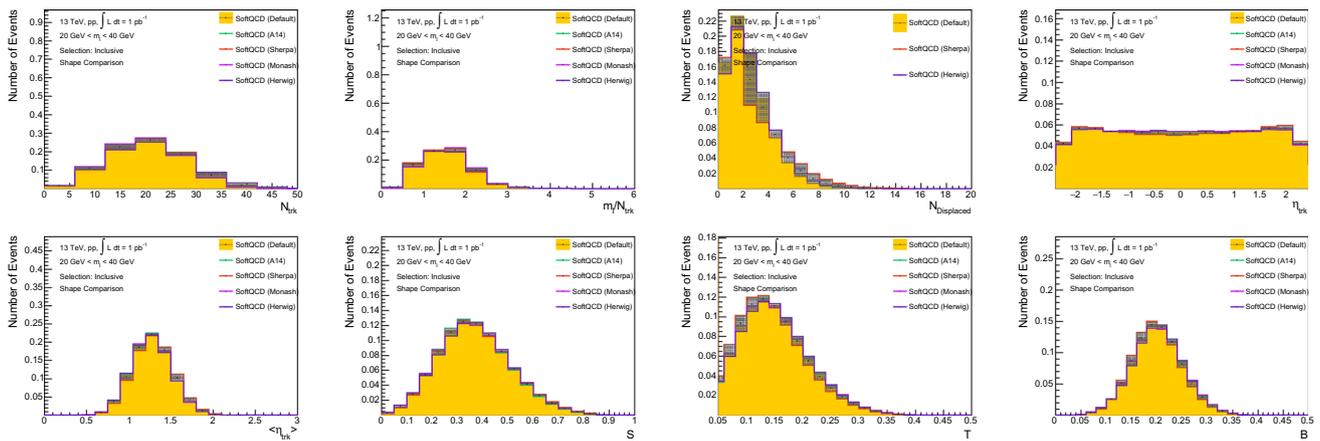


Fig. 4 Normalized distributions for *softQCD* processes with an invariant mass based on all tracks between 20 GeV and 40 GeV, predicted by different MC generators or generator settings: N_{Trk} , m_I/N_{Trk} , $N_{Displaced}$, η_{Trk} , $\langle \eta_{Trk} \rangle$, S , B , T

NNPDF23LO PDF set. An uncertainty of 10% on all predictions of multi-jet, $t\bar{t}$ and vector boson processes are assumed in the following, to account for all theoretical uncertainties from missing higher perturbative orders and the modelling of effects from shower, hadronisation and MPI. However, it should be noted that it is possible to simulate these processes at higher perturbative accuracy. In addition, a realistic data analysis could use the leptonic decay channels of vector boson and $t\bar{t}$ production to validate the theoretical predictions in dedicated control regions, thus reducing the model uncertainties.

3.3 Signal simulation

The Instanton signal samples have been produced with a modified version of the SHERPA event generator [31,47,48]. The predicted cross sections for Instanton production at different Instanton masses, m_I , have been implemented based on the calculations in [31] and are shown in Fig. 5. The figure also shows the predicted dependence of the Instanton cross section on the center of mass energy of proton–proton collisions for two different values of $\sqrt{s'_{min}}$, where also cross sections for soft- and hard-processes are shown in comparison. It is interesting to note that the Instanton cross sections exhibit a different dependence on \sqrt{s} than *softQCD* and *hardQCD* processes.

The decay of the Instanton pseudo-particle in the SHERPA implementation proceeds as follows [31]: firstly, the particle content of the final state is determined, where quark–anti-quark pairs $q\bar{q}$, starting from the lowest mass, are added as long as the mass of the quark m_q is smaller than a kinematics dependent threshold μ_q , $m_q < \mu_q$ and as long as the combined mass of all pair-produced quarks is smaller than the Instanton mass. In a second step, the number of additional gluons is determined according to a Poissonian distribution

with mean $\langle n_g \rangle$. The *Rambo* algorithm [49] is then used to distribute isotropically momenta to all decay products in the rest-frame of the Instanton pseudo-particle and boosted back to the lab-frame. The subsequent showering and hadronisation is based on the standard SHERPA implementation. The shapes of selected experimental observables for Instanton processes is shown in Fig. 6 for a mass range of 500 GeV to 800 GeV. For comparison, the same distribution for the *hardQCD* processes with a minimal energy s' of 500 GeV are overlaid.

To efficiently populate the full range of Instanton masses, a total of six Instanton signal samples have been produced, each covering an exclusive mass range between 20 GeV and 600 GeV. An overview is given in Table 2.

In order to validate the main properties of the Instanton decay in the SHERPA implementation, additional samples of the similar decay in the HERWIG7 generator [50] have been produced, where isotropic decays of pseudo-particles with masses between 500 GeV and 800 GeV into $2 \cdot N_f$ -quarks and n_g gluons have been simulated and reasonable agreement has been observed.

3.4 Experimental observables

Several observables sensitive to Instanton processes have been studied at detector level. The basic input quantities used are the four-vectors of charged particles reconstructed as particle tracks as well as particle jets, reconstructed using an anti- k_T [51,52] algorithm with a radius parameter of 0.4. Tracks are assumed to be massless and hence their four-vectors are defined by their transverse momentum, p_T , in the $x - y$ plane,¹ the polar angle θ measured from the positive z axis, as well as the azimuthal angle ϕ in the $x - y$ plane. The polar

¹ Transverse to the beam axis.

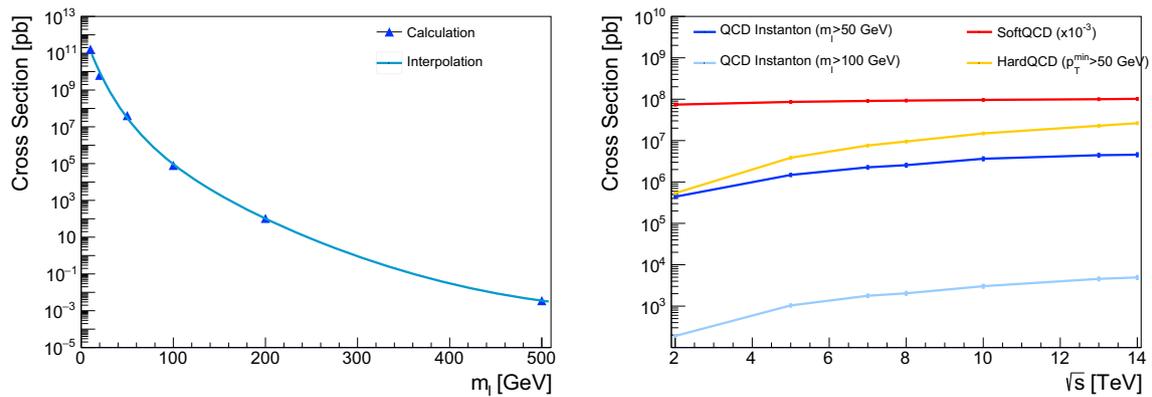


Fig. 5 The production cross section of Instanton processes in proton proton collisions at $\sqrt{s} = 13$ TeV as a function of their mass (left) and a comparison of the cross section dependence on \sqrt{s} between the Instanton signal and various background processes (right)

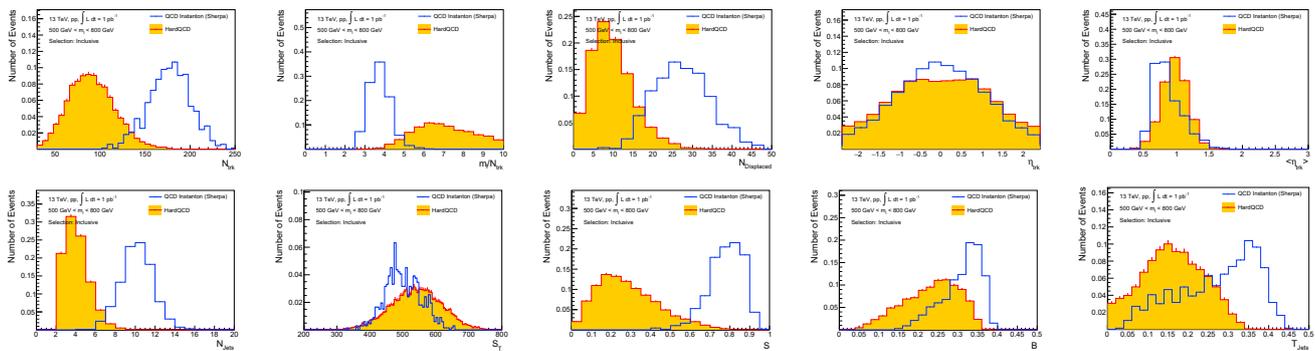


Fig. 6 Normalized distributions of the Instanton signal sample, generated by Sherpa with Instanton masses between 500 GeV and 800 GeV as well as for *hardQCD* processes in a similar mass range: N_{Trk} , m_I/N_{Trk} , $N_{Displaced}$, η_{Trk} , $\langle|\eta_{Trk}|\rangle$, N_{Jets} , S_T^{tracks} , S , B , T

angle is expressed in terms of the pseudorapidity η , defined by $\eta = -\ln(\tan \theta/2)$. We require transverse momenta greater than 500 MeV for reconstructed tracks, as well as a maximal absolute pseudo-rapidity of 2.5. Particle jets are required to have at least a transverse momentum of 20 GeV and $\eta < 2.5$.

A first experimental observable for the selection of Instanton processes is the number of reconstructed tracks, N_{Trk} , as well as the number of reconstructed jets N_{jet} . Since many charged decay particles are expected from the decay of an Instanton pseudo-particle with a given mass, the ratio between the mass and the number of tracks, m_I/N_{Trk} is also of interest. Similarly, the scalar sum of the transverse momenta p_T of all charged tracks (or particle-jets), $S_T = \sum |\vec{p}_T^i|$ in dependence of m_I is studied. Isotropic decays of resonances are expected to have more central than forward activity, i.e. the pseudo-rapidity distribution of all charged tracks, η_{Trk} , as well as the average pseudo-rapidity of charged tracks per event $\langle\eta_{Trk}\rangle$ are expected to be sensitive. Due to the presence of *c*- and *b*-quarks as decay products of the Instanton and the relatively long life-times of the corresponding hadronised mesons, one might expect a higher number of charged particles displaced vertices compared to other SM processes. The number of reconstructed charged

particles tracks with a production vertex that has a distance in the transverse plane of more than 0.02 mm to the primary vertex of the collision, $N_{Displaced}$, is therefore also studied.

We also consider variables which are directly related to the expected isotropy of Instanton decays. One such variable is the event-sphericity S , defined via the tensor S ,

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2},$$

where the indices denote the *x*, *y*, and *z* components of the momentum of the particle *i* in its rest-frame. The sphericity of the event is constructed using the two smallest eigenvalues of this tensor, λ_2 and λ_3 , i.e. $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ and takes values between 0 and 1. A fully balanced dijet events leads to a sphericity of $S = 0$, while a fully isotropic event has a sphericity of $S = 1$. A similar event shape variable is the thrust T , defined as

$$T = 1 - \max_{\vec{n}} \frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|}, \tag{3.1}$$

where \vec{n} is a unit vector. Fully spherical symmetric events yield $T = 0.5$, while fully balanced dijet events have $T = 0$. The definition of thrust also defines the thrust axis \vec{n} , which

maximises the value of \mathcal{T} . The thrust axis defines a left \mathcal{L} and right \mathcal{R} hemisphere for each event, which can be used to define the jet broadening of an event. The left and right broadening is defined as

$$\mathcal{B}_{\mathcal{L}} = \sum_{i \in \mathcal{L}} \frac{|\vec{p}_i \times \vec{n}|}{\sum_i |\vec{p}_i|} \quad \text{and} \quad \mathcal{B}_{\mathcal{R}} = \sum_{i \in \mathcal{R}} \frac{|\vec{p}_i \times \vec{n}|}{\sum_i |\vec{p}_i|}. \quad (3.2)$$

The total jet broadening \mathcal{B} is then defined as $\mathcal{B} = \mathcal{B}_{\mathcal{L}} + \mathcal{B}_{\mathcal{R}}$, and behaves similar as \mathcal{T} , i.e. is 0 and 0.5 for dijet and spherically symmetric events, respectively.

The sphericity \mathcal{S} , the thrust \mathcal{T} as well as the total jet broadening \mathcal{B} are calculated twice, using all reconstructed tracks and using all reconstructed jets in an event. The calculation is based on the code provided in [53]. It should be noted that these three observables are significantly correlated. An additional correlation is observed between \mathcal{S} and the η_{Trk} distribution as events with large values of \mathcal{S} tend to enhance the number of tracks in the central region, i.e. with $|\eta| < 1.0$.

4 Search strategies

In contrast to most searches for new particles, no resonance behaviour is expected for Instanton induced processes, but rather a continuous, rapidly falling excess in the spectrum of invariant mass of all hadronic final state objects. This provides significant challenges in the search for Instanton-induced processes. While sizeable cross sections are expected for small Instanton masses, the experimental signatures in this energy range might be difficult to distinguish from soft QCD activity. At high luminosities, the large amount of expected pile-up events would further complicates such a search. In the high energy regime, the experimental signatures of Instanton-induced processes would be striking, their cross sections are however highly suppressed and hence difficult to observe in the first place.

The expected invariant mass distribution of reconstructed tracks is shown in Fig. 7 for the Instanton processes and the SM backgrounds, scaled to the expected event yields for an integrated luminosity of $\int L dt = 1 \text{ pb}^{-1}$. At low invariant masses, the *softQCD* processes are the dominating background, while for high invariant masses *hardQCD* processes as well as top-quark pair and electroweak boson production become relevant. The signal over background ratio falls rapidly with increasing mass, suggesting a higher chance of observing of Instanton processes in the low mass regime. As expected, the high multiplicity final state for the Instanton processes make the number of reconstructed charged particle tracks and the number of reconstructed jets in the event powerful discriminants against background. Figure 7 shows the invariant mass distribution for signal and background processes when requiring at least eight reconstructed jets with

a minimal momentum of 20 GeV. The expected signal to background ratio increases by several orders of magnitude, remaining below 10^{-4} with the expectation of 1 Instanton event for an integrated luminosity of $\int L dt \approx 0.1 \text{ fb}^{-1}$. This highlights the challenge of observing of Instanton processes for masses of several hundred GeV.

In the following, we will present and discuss possible analysis strategies. To this aim, we found useful to distinguish four different mass ranges: $20 < \sqrt{s'}_{\min} < 40$ GeV and $40 < \sqrt{s'}_{\min} < 80$ GeV for the *low mass regime*, where *softQCD* processes dominate, $200 < \sqrt{s'}_{\min} < 300$ GeV for the *medium mass regime* where *hardQCD* processes dominate and $300 < \sqrt{s'}_{\min} < 500$ GeV for the *high mass regime*, where also top-quark pair productions becomes relevant.

Different signal selections have been studied, optimised on the signal to background ratio. In addition, at least two control regions are defined for each mass range. These regions are designed to have only a small signal contribution and could therefore be used for the validation of the modelling of background processes. In a real analysis, the control regions can be used for an ABCD-based background estimation technique, i.e. to determine the background contribution in the signal regions in a fully data-driven way.

4.1 Very low instanton masses: the soft QCD regime

The very low Instanton mass regime is defined for two regions: the first requires the invariant mass of reconstructed tracks, m_I , between 20 and 40 GeV, the second between 40 and 80 GeV. The average m_I values for Instantons in both regions are 24 GeV and 46 GeV, respectively. A veto on 20 GeV jets at reconstruction level is applied for both mass ranges. This requirement is applied to keep these regions orthogonal to the regions where *hardQCD* processes dominate.²

The lower mass region is discussed first: Fig. 8 shows the predicted distributions of the event sphericity, \mathcal{S} and the pseudo rapidity, η_{Trk} , of reconstructed charged particles for the various processes considered. The distributions are scaled to the expected event yields for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$. While *softQCD* processes dominate the background, the Instanton signal is enhanced at large values of the sphericity and predicts a more central η tracks distribution.

It is illustrative to compare the signal and background shapes for various observables. An overview of eight relevant observables, previously introduced, is shown in Fig. 9. Instantons processes are expected to have larger track multiplicities and hence smaller values of m_I/N_{Trk} . As expected, the observables related to the event topology indicate more

² Only *hardQCD* events with at least one jet on particle level with $p_T > 20$ GeV, that has either not been reconstructed or reconstructed with a smaller value of p_T , pass the signal selection

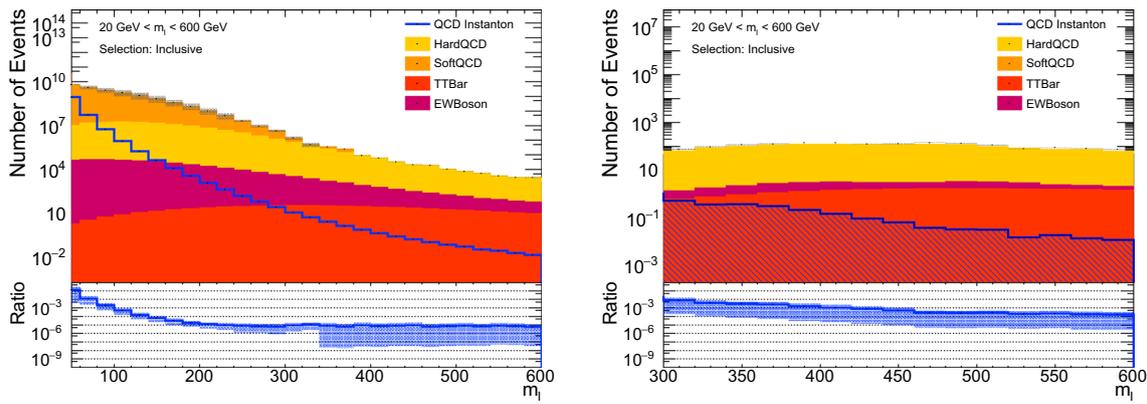


Fig. 7 Invariant mass distribution of all reconstructed tracks for standard model background processes and Instanton processes (left) as well as the same distribution for events with at least ten reconstructed particle jets with a $p_T > 20 \text{ GeV}$ (right). The events correspond to an integrated

luminosity of $L = \int 1 \text{ pb}^{-1}$ and the distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands. The lower plots show the signal over background ratio corresponding to the upper row

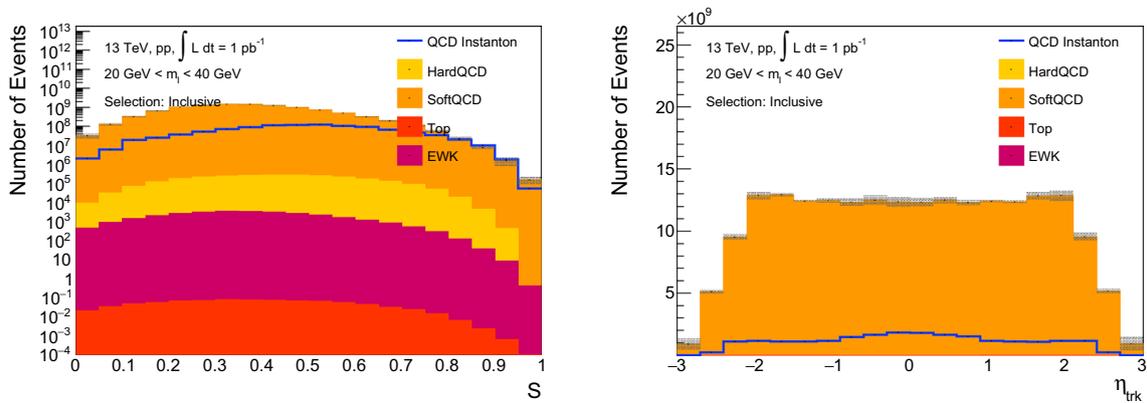


Fig. 8 Predicted distributions of the event sphericity, S (left), and pseudo rapidity, η , of reconstructed charged particles (right) for various processes, normalised to the expected event yields for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$. The invariant mass of all reconstructed

tracks is required to be between 20 and 40 GeV (very low mass regime). The distributions from all SM processes except for the Instanton are stacked. The model uncertainties are indicated as bands

spherical events compared to the background processes. Highly interesting is the distribution of $N_{\text{Displaced}}$, i.e. the number of tracks with a displaced origin, as it differs significantly for *soft*- and *hard*QCD processes and the signal process. This behaviour might be explained by the fact that more heavy quarks in the final state of the Instanton decays are expected, which typically hadronise to long(er) lived mesons and baryons. Based on these distributions, a few signal selection scenarios have been developed and are summarised in Table 3. No requirements are made specifically for the event sphericity as well as the pseudo-rapidity of tracks. The idea behind this approach is, that these distributions could then be used as final discriminating variables in a combined fit of signal and background templates to data in order to extract a limit on the Instanton signal-strength.

The *standard signal selection* applies requirements on the N_{Trk} , m_l/N_{Trk} and N_{Jets} distributions, where the latter is required to be 0 to reject *hard*QCD processes. The resulting sphericity and track η distributions for the signal and background processes is shown in Fig. 10. An improvement by a factor of two in the signal over background ratio becomes visible after these selections are applied. In particular, the expected number of Instanton events becomes larger than the total SM background for event with sphericities $S > 0.85$, which is used to define the signal region selections in the following. The *event-shape* signal selection adds additional requirements on \mathcal{B} and \mathcal{T} , hence affecting also the S distribution. The observables after this selection are shown in Fig. 11. The signal over background ratio improves further when also a requirement of $N_{\text{Displaced}} > 6$ is applied, which

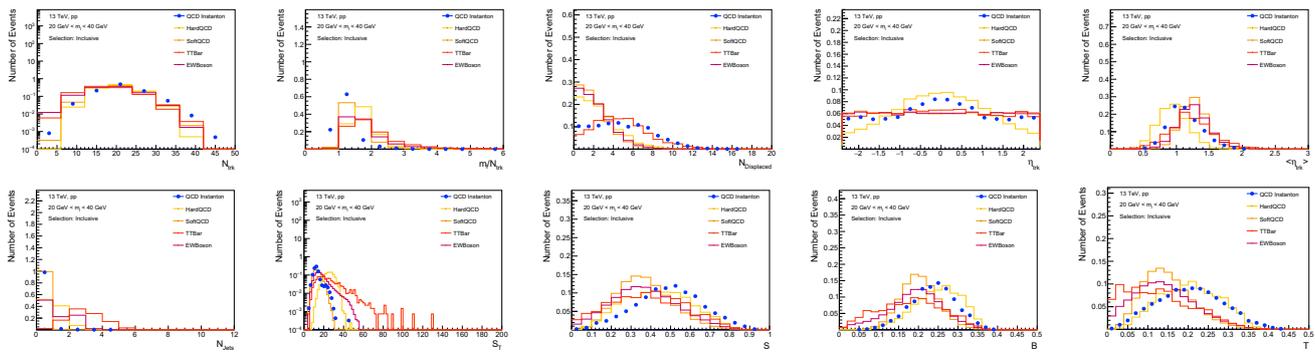


Fig. 9 Normalized distributions in the very low mass regime for various processes: N_{Trk} , m_I/N_{Trk} , $N_{Displaced}$, η_{Trk} , $\langle|\eta_{Trk}|\rangle$, N_{Jets} , S_T^{tracks} , S , B , T

defines our *tight signal selection* (Table 3). The resulting sphericity distribution for the *tight signal selection* is shown in Fig. 11. A very clean Instanton signal is expected for this tight selection.

Two possible definitions of control regions, called A and B, are summarised in Table 3. Both exhibit a signal contamination smaller than 10%. The $N_{Displaced}$ distribution for control region A, as well as the sphericity distribution for control region B, are shown in Fig. 12.

The calculation of Instanton cross sections relies on a semiclassical approximation which breaks down when the coupling is very large. This effect is exemplified by the fact that the Instanton cross section rises beyond the total proton–proton cross section for very low values of s_{min} . We thus also study a higher range of Instanton masses, $40 < m_I < 80$ GeV, where we expect the cross section predictions for Instanton processes should be more reliable. In this regime, the *softQCD* background is still the dominant one in most regions of the phase space. Analogously to the previous case, three different signal selection scenarios and two control regions are defined, which is summarised in Table 4.

The sphericity distribution for the signal and background processes for an inclusive selection³ in this mass range, as well as the three signal region definitions (*standard*, *event-shape*, *tight*) are shown in Fig. 13, together with the expected signal and background events in the signal region in Table 4. The *standard* and *event-shape* selections yield signal to background ratios below 1 and are dominated by *softQCD* processes. The *tight* selection significantly enhances the signal and would allow for a clear observation. In this selection the dominant background contribution in the signal region comes from *hardQCD* processes. We note however that the negligible contribution from *softQCD* could just be a consequence of the limited statistics employed in this study, and it is possible that the actual background

from *softQCD* processes is larger. The interplay between *softQCD* and *hardQCD* processes could be experimentally studied, by applying various requirements on the number of reconstructed jets and thus define different control regions.

4.2 Medium instanton masses: the hard QCD regime

Increasing the range of Instanton masses considered, one enters the regime of perturbative QCD, and the background prediction becomes less uncertain. A medium mass range with $200 < m_I < 300$ GeV with an average Instanton mass of 220 GeV was studied. The S_T^{tracks} distribution and the number of reconstructed jets with $p_T > 20$ GeV in this mass range is shown in Fig. 14 for the signal and background processes. The Instanton processes are expected to peak for $3 \leq N_{jet} \leq 6$, and one can see events with $S_T > 150$ GeV have only a negligible contribution from *softQCD* processes.

A *standard signal selection* imposes requirements on N_{Trk} , m_I/N_{Trk} and the N_{Jets} distribution. The *event-shape* selection applies in addition a minimum requirement on B and T . Similarly to the low mass scenarios, a requirement on $N_{Displaced}$ is made for the *tight signal selection*. All cuts for the signal selection as well as the definitions of the two control regions are summarized in Table 5. The *event-shape* selection yields 1 signal event and approximately 10 background events for an integrated luminosity of $\int Ldt = 1 \text{ pb}^{-1}$. For the *tight selection* 0.5 signal events and 0.6 background events are expected. An observation would therefore be possible with an integrated luminosity of about $\int Ldt \sim 10 \text{ pb}^{-1}$.

In this context the ability of the LHC experiments to efficiently trigger on these event topologies becomes relevant. For the studies of *softQCD* processes at the LHC, special triggers are used, which record collision events even with limited activity in the detector, i.e. are nearly free of any bias towards a certain physics signature. Given the enormous rates of such *minimum bias triggers*, only a small fraction of

³ Inclusive is defined here as only applying selection cuts on the reconstructed invariant mass m_I of the Instanton

Table 3 Overview of the standard and tight signal selection as well as the definition of two control regions aiming at very low Instanton masses ($20 \text{ GeV} < m_I < 40 \text{ GeV}$)

	Signal region			Control region	
	Standard	Event-shape	Tight	A	B
Invariant mass of rec. tracks (Instanton mass), m_I	$20 \text{ GeV} < m_I < 40 \text{ GeV}$				
Selection requirements					
Number of rec. tracks, N_{Trk}	>20	>20	>20	>15	>20
Number of rec. tracks/Instanton mass, m_I/N_{Trk}	<1.5	<1.5	<1.5	>2.0	<1.5
Number of Jets, N_{Jets}	$=0$	$=0$	$=0$	$=0$	$=0$
Broadening, $\mathcal{B}_{\text{Tracks}}$		>0.3	>0.3	>0.3	>0.3
Thrust, $\mathcal{T}_{\text{Tracks}}$		>0.3	>0.3	>0.3	>0.3
Number of displaced vertices, $N_{\text{Displaced}}$			>6		<4
Expected events for $\int L dt = 1 \text{ pb}^{-1}$ in the signal region ($S > 0.85$)					
N_{Signal}	1.1×10^7	8.9×10^6	5.9×10^6	<1	6.8×10^5
$N_{\text{Background}}$	6.2×10^6	4.3×10^6	1.8×10^5	3×10^5	3.3×10^6

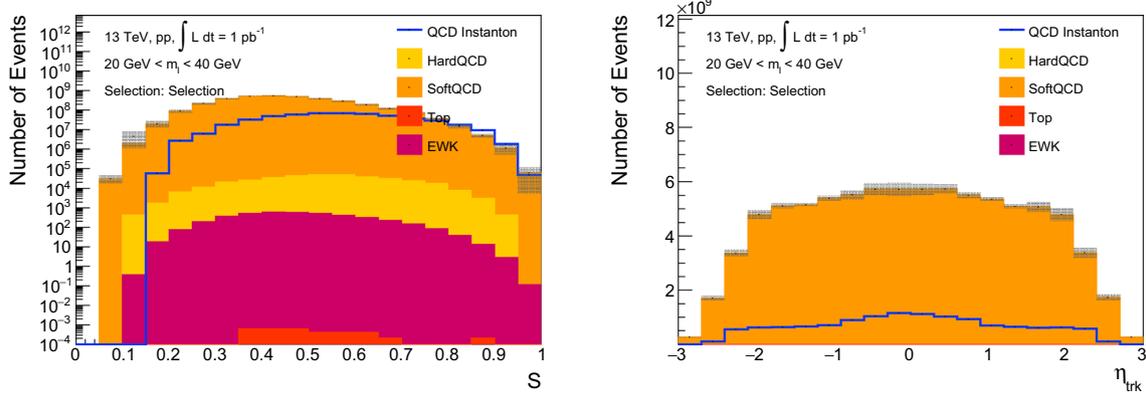


Fig. 10 Predicted distributions of the event sphericity (left) and pseudo-rapidity, η , of reconstructed charged particles (right) for various processes, weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$ after the nominal selection. The invariant mass

of all reconstructed tracks is required to be between 20 GeV and 40 GeV (very low mass regime). The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

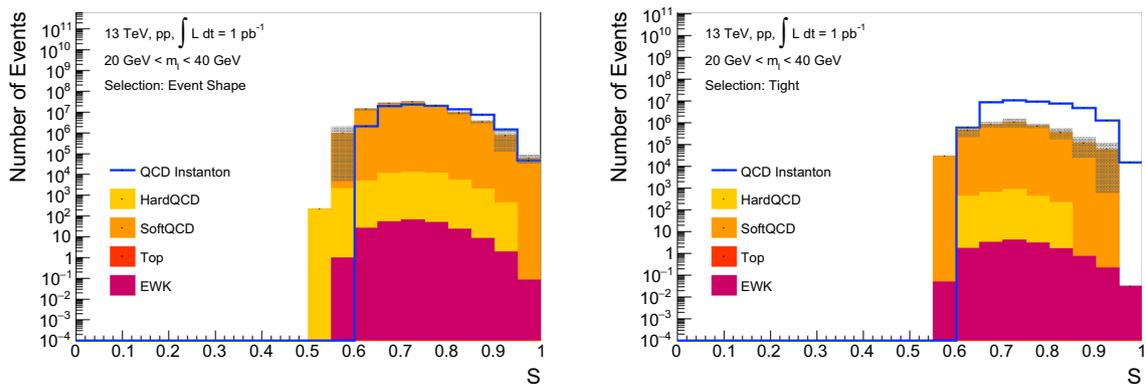


Fig. 11 Predicted distributions of the event sphericity for various processes, weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$ after the *event-shape* based selected (left) and tight selection (right). The invariant mass of all reconstructed tracks is

required to be between 20 and 40 GeV (very low mass regime). The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

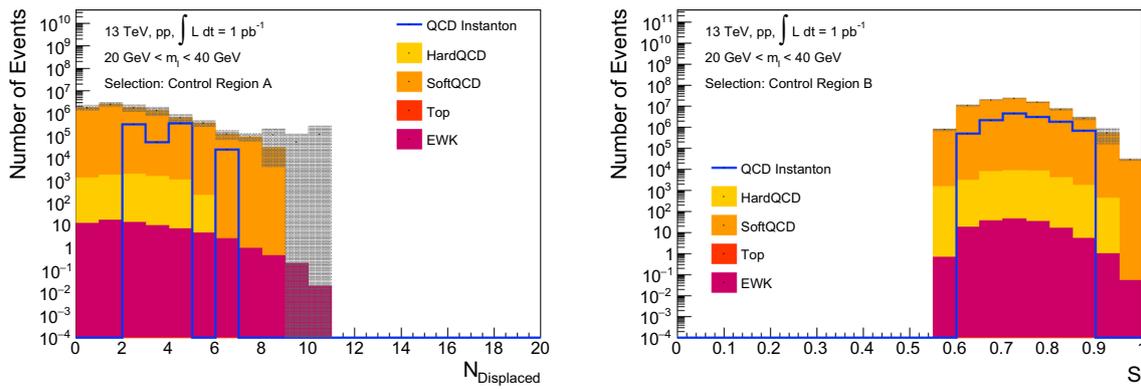


Fig. 12 Predicted distributions of the number of displaced tracks in the control region A (left) and the event sphericity in control region B (right), weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$. The invariant mass of all reconstructed tracks is

required to be between 20 and 40 GeV (very low mass regime). The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

Table 4 Overview of the standard and tight signal selection as well as the definition of two control regions aiming at very low Instanton masses ($40 \text{ GeV} < m_I < 80 \text{ GeV}$)

	Signal region			Control region	
	Standard	Event-shape	Tight	A	B
Invariant mass of rec. tracks (Instanton mass), m_I	$40 \text{ GeV} < m_I < 80 \text{ GeV}$				
Selection requirements					
Number of rec. tracks, N_{Trk}	>30	>30	>30	<20	>30
Number of rec. tracks/Instanton mass, m_I/N_{Trk}	<2.0	<2.0	<2.0	>2.0	<2.0
Number of Jets, N_{Jets}	$= 0$	$= 0$	$= 0$	$= 0$	$= 0$
Broadening, $\mathcal{B}_{\text{Tracks}}$		>0.3	>0.3	>0.3	>0.3
Thrust, $\mathcal{T}_{\text{Tracks}}$		>0.3	>0.3	>0.3	>0.3
Number of displaced vertices, $N_{\text{Displaced}}$			>12		<8
Expected events for $\int L dt = 1 \text{ pb}^{-1}$ in the signal region ($S > 0.85$)					
N_{Signal}	4.8×10^5	3.7×10^5	52000	0	1.3×10^5
$N_{\text{Background}}$	1.2×10^6	8.7×10^5	432	3000	7.5×10^5

these events can be actually stored on tape and subsequently analysed. However, the published *softQCD* analysis at the LHC indicate that sufficient statistics has been already collected to allow for Instanton searches in the low mass regime. This is not obviously the case for the medium and high Instanton mass regime, as the integrated luminosity required for an observation increases significantly. Typically, jet and multi-jet triggers require minimal transverse jet energies of 50 GeV or more and hence are of no use. New trigger strategies might be needed the upcoming LHC runs to be able to record enough Instanton events, as pointed out independently in [32].

The predicted distributions of the event sphericity for the signal and background processes, scaled to the expected event yields for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$ is shown in Fig. 15 for the *event-shape* and the *tight* selection. The dominant background are multijet events from

hardQCD processes. Figure 16 shows the predicted number of displaced tracks in the control region A as well as the event sphericity in control region B, which again can be used to validate the modelling of the background processes.

4.3 High instanton masses: the top quark regime

The $300 < m_I < 500 \text{ GeV}$ mass range is also dominated by *hardQCD* multi-jet background. This can be seen in Fig. 14, which shows the event sphericity and the number of reconstructed particle jets with $p_T > 20 \text{ GeV}$ for signal and background events. However, it is interesting to apply a dedicated event selection that promotes another process as dominant background, not suffering from the same model uncertainties as the background in the medium mass regime. Hence we focus our selection here on top-quark pair events. While an obvious approach to enhance the top-quark (as well as the

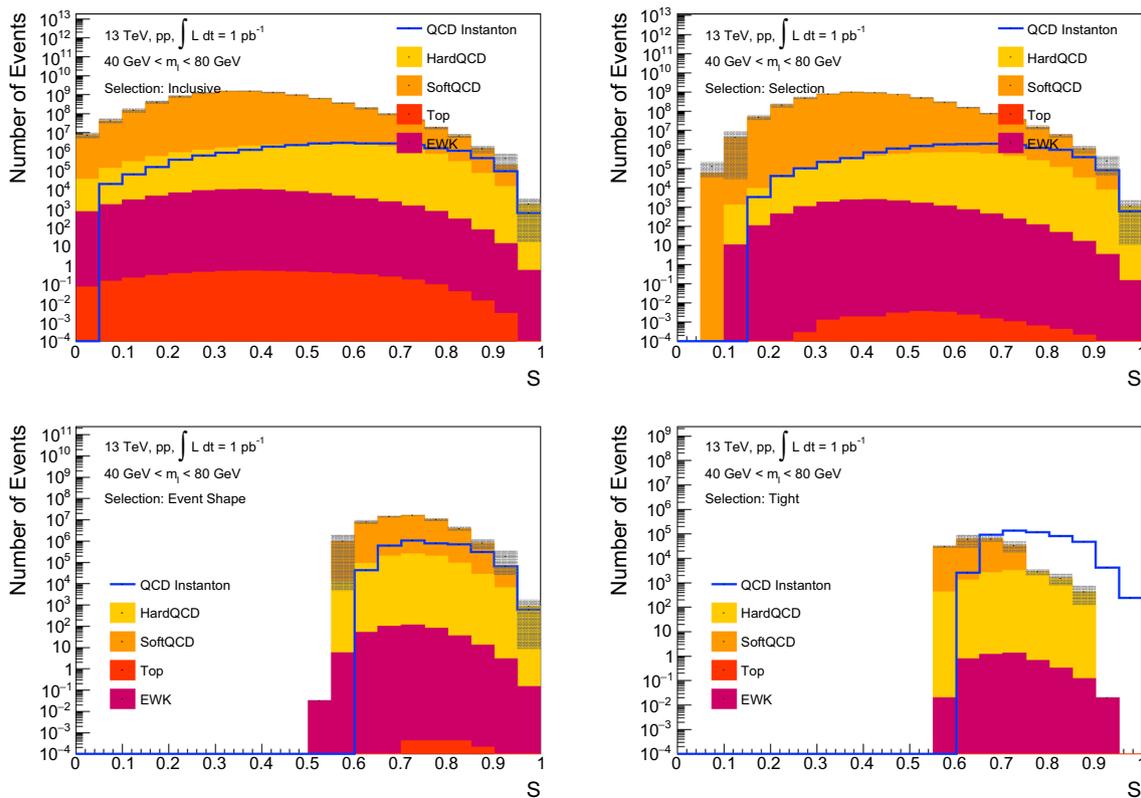


Fig. 13 Predicted distributions of the event sphericity for various processes, weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$ for an inclusive selection (upper left), the nominal selection (upper right), the *event-shape* based selection (lower left)

and the tight selection (lower right). The invariant mass of all reconstructed tracks is required to be between 40 and 80 GeV. The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

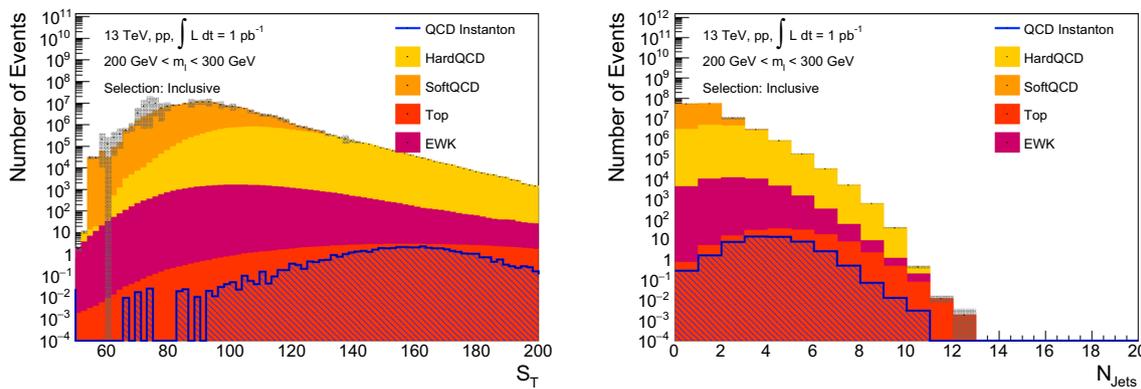


Fig. 14 Predicted distributions of the ST (left) and number of reconstructed jets (right) for various processes, weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$. The invariant mass of all reconstructed tracks is required to be between 200 and

300 GeV (medium mass regime). The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

Table 5 Overview of the standard and tight signal selection as well as the definition of two control regions aiming at very low Instanton masses ($200 \text{ GeV} < m_I < 300 \text{ GeV}$)

	Signal region			Control region	
	Standard	Event-shape	Tight	A	B
Invariant mass of rec. tracks (Instanton mass), m_I	$200 \text{ GeV} < m_I < 300 \text{ GeV}$				
Selection requirements					
Number of rec. tracks, N_{Trk}	>80	>80	>80	>80	>80
Number of rec. tracks/Instanton mass, m_I/N_{Trk}	<3.0	<3.0	<3.0	>3.0	<3.0
Number of Jets, N_{Jets}	3–6	3–6	3–6	3–6	3–6
Broadening, $\mathcal{B}_{\text{Tracks}}$		>0.3	>0.3	>0.3	>0.3
Thrust, $\mathcal{T}_{\text{Tracks}}$		>0.3	>0.3	>0.3	>0.3
Number of displaced vertices, $N_{\text{Displaced}}$			>15		<10
Results					
Expected events for $\int L dt = 1 \text{ pb}^{-1}$ in the signal region ($S > 0.85$)					
N_{Signal}	5.6	1.0	0.54	0.04	0.21
$N_{\text{Background}}$	1900	9.6	0.64	200	1100

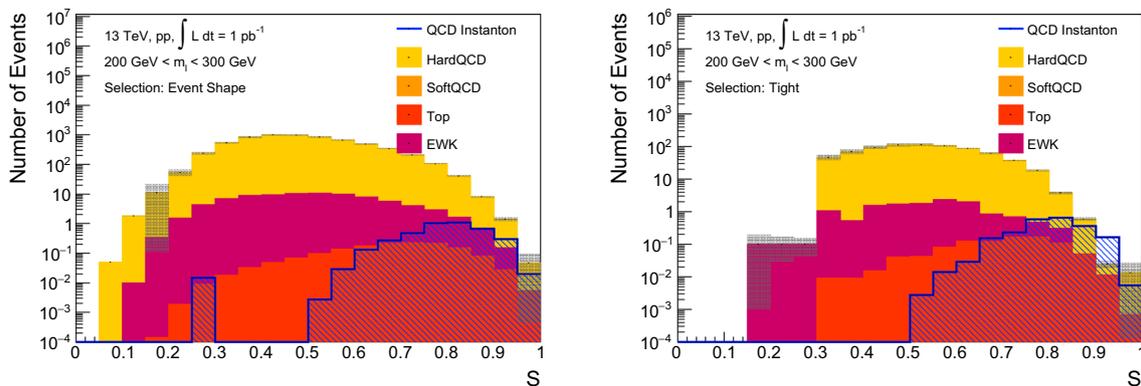


Fig. 15 Predicted distributions of S for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$ for the *event-shape* based selection (left) and the *tight* selection (right). The invariant mass of all reconstructed tracks is required to be between 200 and 300 GeV (medium mass regime)

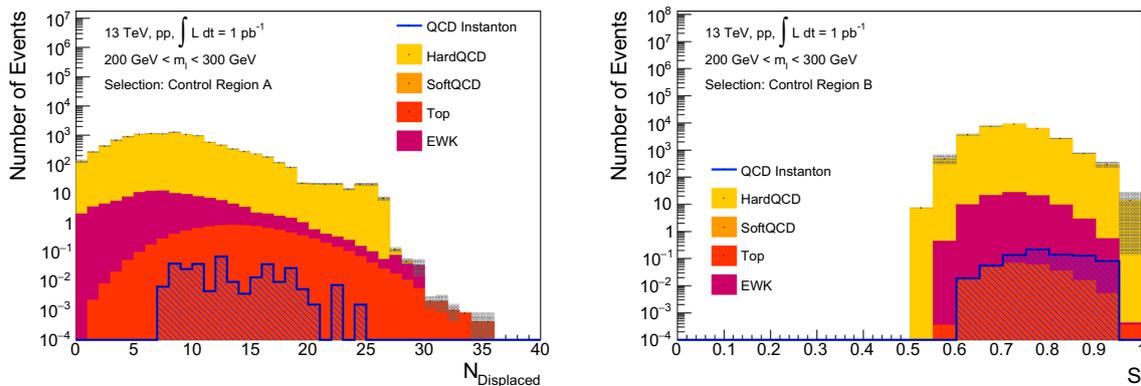


Fig. 16 Predicted distributions of the number of displaced tracks in the control region A (left) and the event sphericity in control region B (right), weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$. The invariant mass of all reconstructed tracks

is required to be between 200 and 300 GeV (medium mass regime). The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

Instanton) contribution is to require reconstructed jets tagged to originate from b -hadrons, this was not required within this study due to the uncertainties in the signal simulation. Our results on the expected sensitivity are therefore conservative (Fig. 17).

The large track multiplicity in events within $300 < m_I < 500$ GeV does not allow anymore for a clean separation between signal and background processes. Hence no signal selection cut involving N_{Trk} is applied. The *standard* signal selection requires only $m_I/N_{\text{Trk}} < 3.0$ as well as more than seven reconstructed jets with $p_T > 20$ GeV. The latter requirement allowing to enhance the top-quark background contribution. The *event-shape* and *tight* signal selections follow the same lines as for the lower mass ranges, i.e. impose cuts on the event topology as well as on $N_{\text{Displaced}}$. A summary of all signal selection criteria is given in Table 6. The corresponding event sphericity distributions for the *event-shape* and the *tight* selection is shown in Fig. 18, where the top-quark background starts to dominate large values of \mathcal{S} . written before b -tagging is not applied. A further enhancement of the top-quark contribution can be achieved by an additional b -tagging requirement without impacting significantly the signal yield.

The advantage of this signal selection relies on the experimentally well understood top-quark pair production. In addition to similar control regions as in the low mass ranges, top-quark enhanced regions can be envisioned, e.g. by requiring one additional reconstructed lepton (electron or muon) in the event, originating from the leptonic decay of one top-quark (see definition of Control Region C in Table 6). This ensures large experimental constraints on the background uncertainties in the signal region. However, the signal over background ratio is only of the order of 30% and only 2 signal events are expected for an integrated luminosity of $\int L dt = 1 \text{ fb}^{-1}$. An observation with a 5σ significance would thus require an integrated luminosity of more than 80 fb^{-1} based on pure statistical considerations. Such large integrated luminosities would require dedicated triggers for high multiplicity of low- p_T jet events during the Run-3 of the LHC, or a long data-taking period with pre-scaled jet triggers during the high luminosity phase of the LHC.

5 Limits on instanton processes in proton–proton collisions

5.1 How to mimic QCD instanton signatures

As shown in the previous sections, the most promising mass range for the observation of Instanton induced processes at the LHC is below 100 GeV, where the *softQCD* background contribution dominates. It is therefore crucial to understand if the *softQCD* phenomenological models have enough free-

dom to mimic the QCD Instanton signatures. A first indication that this might not be easily achieved comes from the observation that the *softQCD* predictions from the PYTHIA, SHERPA, and the HERWIG7 generators, which implement different models, are remarkably consistent for the observables considered in this study. All of these models are however based on the same MPI description of *softQCD*, and with parameters tuned to describe the same, or similar data. It remains possible that with a suitable parameter choice the *softQCD* predictions can be made more similar to the Instanton.

As a proof-of-principle demonstration, we have tested if the *softQCD* Pythia predictions can be made to yield significantly more spherical events, even beyond what data indicates. Starting with the baseline MONASH tune [43] of PYTHIA8, we found that increasing the MULTIPARTONINTERACTIONS:ALPHASVALUE = 0.150 in *softQCD* events does lead to more spherical events, as seen in Fig. 19. However, such a tune would also alter many other event shape distributions, such as the number of charged particles vs. η , which are not supported by data, as seen for example in Fig. 20). While the latter is based on $\sqrt{s} = 7$ TeV data, the same conclusions hold for $\sqrt{s} = 13$ TeV. We also expect that tune of multiple parton interactions would not impact certain distributions, such as the number of tracks with displaced vertices $N_{\text{Displaced}}$, which could then be used to isolate Instanton events. As the cross section dependence on m_I of QCD Instanton and *softQCD* processes is very different, it is also non-trivial to tune QCD Instanton sensitive distributions for different m_I regions. This can be taken as a further motivation for studying Instanton production at low mass in different mass ranges, i.e. $20 < m_I < 40$ GeV and $40 < m_I < 80$ GeV. We remark however that to fully rule out the potential degeneracy between more detailed studies, exploring changes to the *softQCD* models themselves, would be needed.

5.2 LHC sensitivity projections

Having defined suitable signal regions, we evaluate here the expected 95% confidence level (CL) upper limits on the instanton cross section. For each range of Instanton mass, the signal and background expectations in the respective *tight* signal region selection are used to perform a counting experiment using the `pyhf` package [56]. The systematic uncertainty on the background estimate is estimated as described in the previous sections, and ranges from about 20% at low invariant masses to about 50% at high invariant masses.

Results are shown in Fig. 21 for different assumptions on the integrated luminosity of 1 pb^{-1} , 100 pb^{-1} , and 10 fb^{-1} . We can see how even with only 1 pb^{-1} , the predicted Instanton cross sections can be excluded for masses up to about 150 GeV, excluding at low masses cross sections ten times

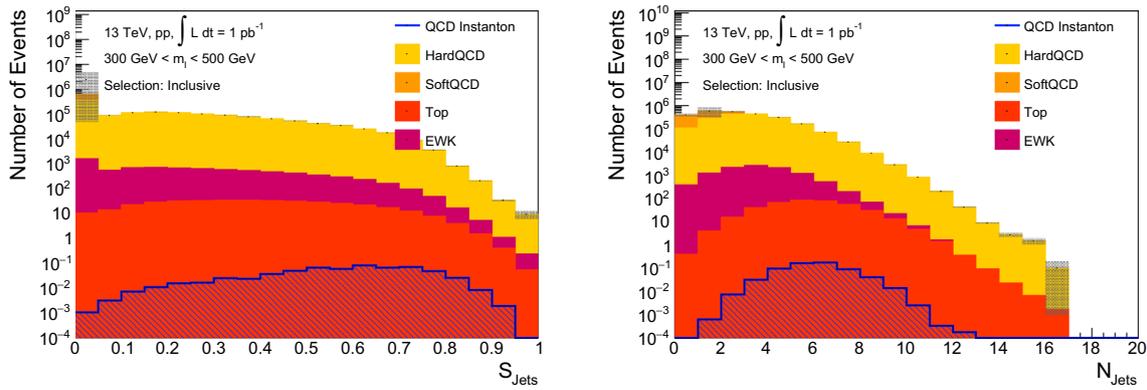


Fig. 17 Predicted distributions of the event sphericity (left) and number of reconstructed jets (right) for various processes, weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$. The invariant mass of all reconstructed tracks is required to be between

300 and 500 GeV (high mass regime). The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

Table 6 Overview of the standard and tight signal selection as well as the definition of two control regions aiming at very low Instanton masses ($300 \text{ GeV} < m_l < 500 \text{ GeV}$)

	Signal region			Control region		
	Standard	Event-shape	Tight	A	B	C
Invariant mass of rec. tracks m_l	$300 \text{ GeV} < m_l < 500 \text{ GeV}$					
Selection requirements						
Number of rec. tracks, N_{Trk}	–	–	–	–	–	–
Number of rec. tracks/inst. mass, m_l/N_{Trk}	<3.0	<3.0	<3.0	>3.0	<3.0	<3.0
Number of Jets, N_{Jets}	>7	>7	>7	>7	>7	>5
Broadening, $\mathcal{B}_{\text{Tracks}}$	–	>0.3	>0.3	>0.3	>0.3	>0.3
Thrust, $\mathcal{T}_{\text{Tracks}}$	–	>0.3	>0.3	>0.3	>0.3	>0.3
Number of displaced vertices, $N_{\text{Displaced}}$	–	–	>20	–	<15	>20
Identified Leptons	–	–	–	–	–	1
Expected events for $\int L dt = 1 \text{ pb}^{-1}$ in the signal region ($S > 0.85$)						
N_{Signal}	0.007	0.004	0.0021	0.002	0.0003	–
$N_{\text{Background}}$	0.204	0.015	0.0074	33	0.0022	0.002

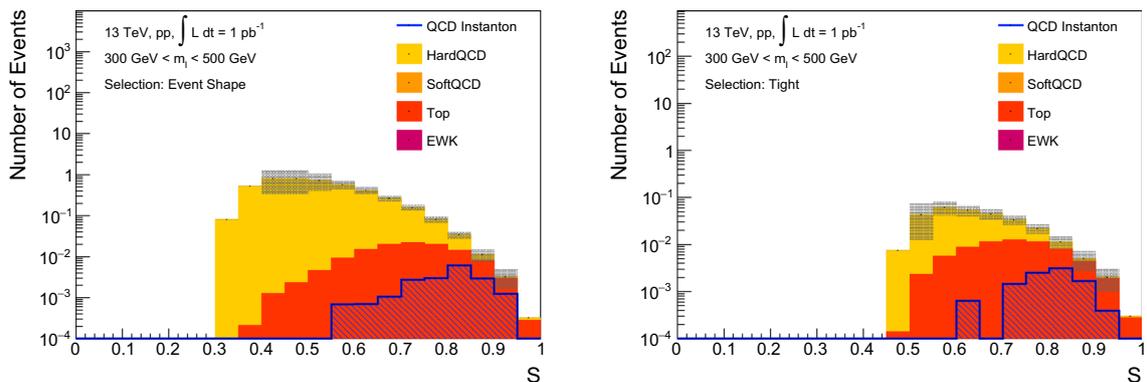


Fig. 18 Predicted distributions of the event sphericity for various processes, weighted by their predicted cross sections for an integrated luminosity of $L = \int 1 \text{ pb}^{-1}$ for the *event-shape* based selection (left) and the *tight* selection (right). The invariant mass of all reconstructed tracks is

required to be between 300 and 500 GeV (high mass regime). The distributions from all processes except the Instanton process are stacked. The model uncertainties are indicated as bands

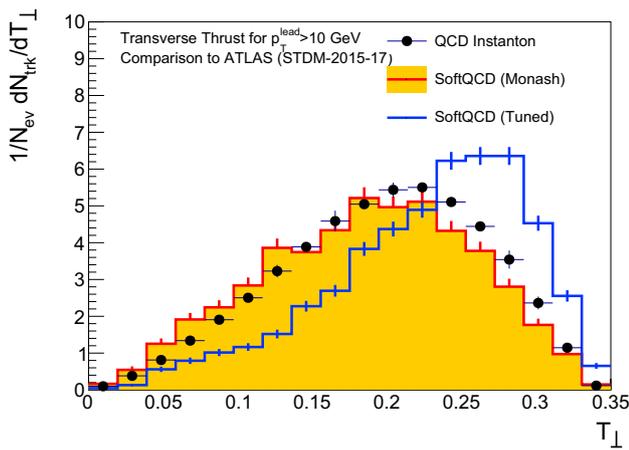


Fig. 19 Predicted distribution of the event thrust of the MONASH *softQCD* tune of PYTHIA8 as well as a modified version with significantly enhanced multiple parton interaction probability (MPI:ALPHASVALUE = 0.150) in comparison to the measurement at 13 TeV of the ATLAS Collaboration [54]

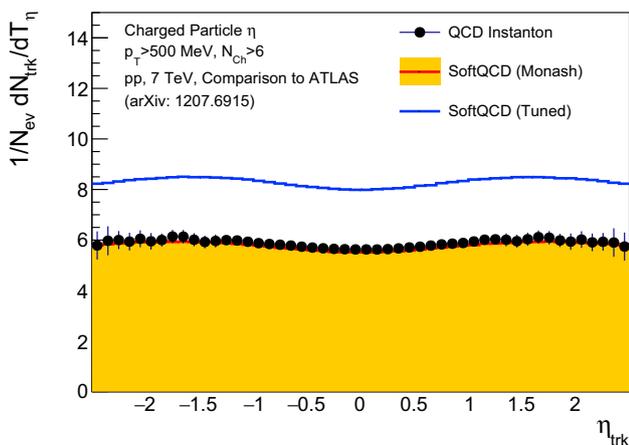


Fig. 20 Predicted distribution of the charged particle spectrum vs. η of the MONASH *softQCD* tune of PYTHIA8 as well as a modified version with significantly enhanced multiple parton interaction probability (MPI:ALPHASVALUE = 0.150) in comparison to the measurement at 7 TeV of the ATLAS Collaboration [55]

smaller than those predicted by [31]. Increasing the collected luminosity to 100 pb^{-1} would extend the limit at large invariant masses to 300 GeV, with a negligible improvement of the limit at low masses. A further integrated luminosity increase by an additional factor of 100 would push the limit at high masses beyond our last simulated mass point of 400 GeV.

5.3 First limit on instanton processes from proton–proton collision data

As already indicated in Fig. 8, the distribution of charged particles vs. their pseudorapidity is a sensitive observable for QCD Instanton processes in the low mass regime. This observable is routinely used in *softQCD* studies and has been

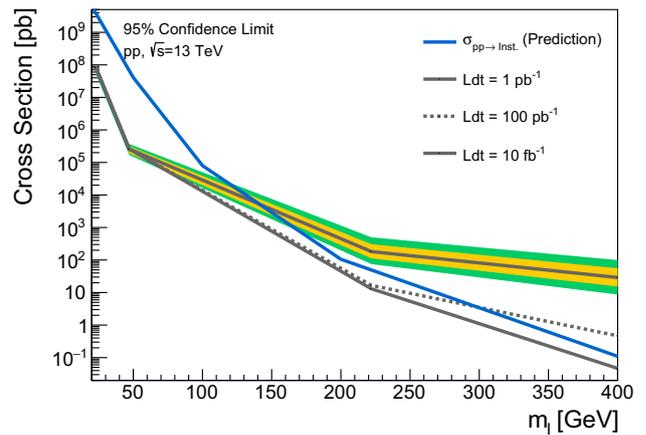


Fig. 21 Expected exclusion limits at 95% CL on the cross section for Instanton induced processes for three different assumed integrated luminosities. Conservative systematic uncertainties on the background modelling have been made.

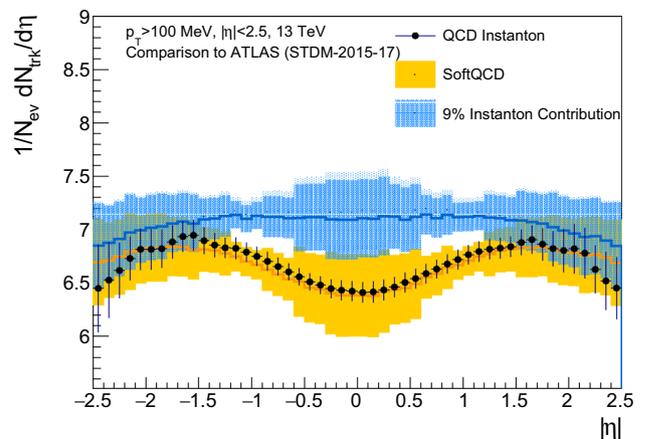


Fig. 22 Primary charged-particle multiplicities as a function of $|\eta|$ for events with at least two primary charged particles with $p_T > 100 \text{ MeV}$ and $|\eta| < 2.5$, each with a lifetime $\tau > 300 \text{ ps}$. The black dots represent the measurement of ATLAS [54], while the *softQCD* prediction is produced with PYTHIA8. Also shown is a prediction which includes a contribution of 9% on Instanton induced processes. The model uncertainties are indicated by the bands

measured at $\sqrt{s}=13 \text{ TeV}$ in pp collisions at the LHC [54,57–59]. We consider the ATLAS Collaboration measurement [54], which requires events with at least two primary charged particles with $p_T > 100 \text{ MeV}$ and $|\eta| < 2.5$, each with a lifetime $\tau > 300 \text{ ps}$, and studied to which extent these measurements can be used to constrain QCD Instanton production. The ATLAS analysis as implemented in RIVET was applied on the simulated *softQCD* sample based on PYTHIA8 as well as the Instanton signal sample with $s'_{\text{min}} > 25 \text{ GeV}$. A modelling uncertainty on the *softQCD* prediction was estimated by considering the envelope of the PYTHIA, HERWIG7 and SHERPA samples. To obtain a systematic uncertainty on the signal modelling, the differences in the η_{Trk} distribution at

$m_I \approx 500$ GeV between the SHERPA and HERWIG7 predictions have been taken as an approximation. The resulting signal and background uncertainties have been treated once fully correlated and once fully uncorrelated bin-to-bin.

The Instanton signal prediction has been added with a scaling factor α to the predicted *softQCD* distribution, normalised accordingly and this combined prediction fitted via a χ^2 minimization approach to the data distribution, as shown in Fig. 22. The fit yields a maximal value of $\alpha = 0.09$ and $\alpha = 0.03$ at 95% CL, assuming bin-to-bin correlated as well as bin-to-bin uncorrelated uncertainties on the predictions, respectively. The expected prediction for an Instanton contribution of 9% to the standard *softQCD* processes is also shown in Fig. 22. The fiducial cross section defined by the selection $p_T > 100$ MeV and $|\eta| < 2.5$, each with a lifetime $\tau > 300$ ps can be estimated to be $\sigma = 71$ mb, when taking the integrated luminosity of $\int L dt = 151 \mu\text{b}^{-1}$, the number of selected events $N = 9.3 \cdot 10^6$ and assuming a detector efficiency of $\epsilon = 0.87$. Hence, an upper limit on Instanton induced processes with $s'_{\text{min}} > 25$ GeV can be placed between 2.1 and 6.4 mb, depending on the correlation scenario assumed. In principle, further measured distributions could be used to derive more stringent limits. However, we think a dedicated analysis effort by the LHC collaborations would be the right next step to shed further light on QCD Instanton processes.

6 Conclusion

In this paper we presented detailed studies towards possible analysis strategies to observe Instanton induced processes in proton–proton collisions at the Large Hadron Collider. Several observables have been identified, which allow to effectively separate signal and background events. In order to study Instanton processes at higher energies, special triggers might have to be implemented for the upcoming LHC runs. However, the situation is different for low energies which have significantly larger cross sections and should be in principle already be recorded. It is concluded that the most promising phase-space region for an early observation is therefore at low energies, for Instanton masses below 100 GeV, where the cross section is very high. Since the dominant background in this energy regime is from *softQCD* processes, several methods to constrain and validate *softQCD* models in dedicated control regions have been discussed. We find that with an integrated luminosity of just 1 pb^{-1} , the LHC can already probe this low mass Instanton regime. With 10 fb^{-1} it would be possible to probe Instanton masses of up to 0.5 TeV. In addition, available measurements of Minimum Bias data have been used to derive a first upper limit on the cross section of Instanton processes, yielding an upper bound of 6.4 mb for Instanton masses above 25 GeV.

The methods described in this paper will hopefully boost dedicated search efforts at the LHC over the full Instanton mass range by several experiments, leading to a robust result based on different strategies.

Acknowledgements We would like to thank Frank Krauss for all his help during the generation of Instanton signal samples with the SHERPA generator, Simon Plätzer and Andreas Papaefstathiou for providing us with a preliminary implementation of Instanton production in HERWIG7, and Daniel Milne for his help in validating our event shape calculations.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This study is of theoretical nature and has no data associated to it.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³.

References

1. C.-N. Yang, R.L. Mills, Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.* **96**, 191–195 (1954)
2. B. Mistlberger, Higgs boson production at $N^3\text{LO}$ in QCD. *JHEP* **05**, 028 (2018)
3. C. Duhr, F. Dulat, B. Mistlberger, Drell–Yan cross section to third order in the strong coupling constant. *Phys. Rev. Lett.* **125**(17), 172001 (2020)
4. D. d'Enterria, Experimental QCD summary (ICHEP 2020) PoS **ICHEP2020**, 012 (2021). <https://doi.org/10.22323/1.390.0012>
5. A.A. Belavin, A.M. Polyakov, A.S. Schwartz, Y.S. Tyupkin, Pseudoparticle solutions of the Yang–Mills equations. *Phys. Lett. B* **59**, 85–87 (1975)
6. E.V. Shuryak, The role of instantons in quantum chromodynamics. I. Physical vacuum. *Nucl. Phys. B* **203**, 93 (1982)
7. P.B. Arnold, L.D. McLerran, Sphalerons, small fluctuations and Baryon number violation in electroweak theory. *Phys. Rev. D* **36**, 581 (1987)
8. H. Forkel, A primer on instantons in QCD (2000). [arXiv:hep-ph/0009136](https://arxiv.org/abs/hep-ph/0009136)
9. E. Shuryak, Lectures on nonperturbative QCD (nonperturbative topological phenomena in QCD and related theories), 12 (2018). [arXiv:1812.01509](https://arxiv.org/abs/1812.01509)
10. V.A. Kuzmin, V.A. Rubakov, M.E. Shaposhnikov, On the anomalous electroweak Baryon number nonconservation in the early universe. *Phys. Lett. B* **155**, 36 (1985)
11. G.R. Farrar, M.E. Shaposhnikov, Baryon asymmetry of the universe in the minimal Standard Model. *Phys. Rev. Lett.* **70**, 2833–2836 (1993) [**Erratum: Phys.Rev.Lett.** **71**, 210 (1993)]
12. V.A. Rubakov, M.E. Shaposhnikov, Electroweak baryon number nonconservation in the early universe and in high-energy collisions. *Usp. Fiz. Nauk* **166**, 493–537 (1996)

13. V.A. Rubakov, M.E. Shaposhnikov, Phys. Usp. **39**, 461 (1996)
14. G. 't Hooft, How instantons solve the U(1) problem. Phys. Rept. **142**, 357–387 (1986)
15. G. 't Hooft, Computation of the quantum effects due to a four-dimensional pseudoparticle. Phys. Rev. D **14**, 3432–3450 (1976) [Erratum: Phys. Rev.D**18**,2199(1978)]
16. G. 't Hooft, Symmetry breaking through Bell–Jackiw anomalies. Phys. Rev. Lett. **37**, 8–11 (1976)
17. D. Diakonov, V.Y. Petrov, A theory of light quarks in the instanton vacuum. Nucl. Phys. B **272**, 457–489 (1986)
18. F.R. Klinkhamer, N.S. Manton, A saddle point solution in the Weinberg–Salam theory. Phys. Rev. D **30**, 2212 (1984)
19. A. Ringwald, High-energy breakdown of perturbation theory in the electroweak instanton sector. Nucl. Phys. B **330**, 1–18 (1990)
20. H. Aoyama, H. Goldberg, Anomalous Baryon number nonconservation in p p collisions at 40-TeV. Phys. Lett. B **188**, 506–510 (1987)
21. O. Espinosa, High-energy behavior of baryon and lepton number violating scattering amplitudes and breakdown of unitarity in the standard model. Nucl. Phys. B **343**, 310–340 (1990)
22. A. Ringwald, Vacuum structure and high-energy scattering. Nucl. Phys. B Proc. Suppl. **121**, 145–148 (2003). [https://doi.org/10.1016/S0920-5632\(03\)01831-0](https://doi.org/10.1016/S0920-5632(03)01831-0)
23. A. Ringwald, Electroweak instantons/sphalerons at VLHC? Phys. Lett. B **555**, 227–237 (2003)
24. V.V. Khoze, D.L. Milne, Suppression of electroweak instanton processes in high-energy collisions. Int. J. Mod. Phys. A **36**(05), 2150032 (2021). <https://doi.org/10.1142/S0217751X21500329>
25. S. Moch, A. Ringwald, F. Schrempp, Instantons in deep inelastic scattering: the simplest process. Nucl. Phys. B **507**, 134–156 (1997)
26. A. Ringwald, F. Schrempp, Zooming in on instantons at HERA. Phys. Lett. B **503**, 331–340 (2001). [https://doi.org/10.1016/S0370-2693\(01\)00216-7](https://doi.org/10.1016/S0370-2693(01)00216-7)
27. A. Ringwald, F. Schrempp, QCDINS 2.0: A Monte Carlo generator for instanton induced processes in deep inelastic scattering. Comput. Phys. Commun. **132**, 267–305 (2000). [https://doi.org/10.1016/S0010-4655\(00\)00148-X](https://doi.org/10.1016/S0010-4655(00)00148-X)
28. C. Adloff et al., Search for QCD instanton induced processes in deep inelastic scattering at HERA. Eur. Phys. J. C **25**, 495–509 (2002)
29. S. Chekanov et al., Search for QCD instanton induced events in deep inelastic ep scattering at HERA. Eur. Phys. J. C **34**, 255–265 (2004)
30. V. Andreev et al., Search for QCD instanton-induced processes at HERA in the high- Q^2 domain. Eur. Phys. J. C **76**(7), 381 (2016)
31. V.V. Khoze, F. Krauss, M. Schott, Large effects from small QCD instantons: making soft bombs at Hadron colliders. JHEP **04**, 201 (2020)
32. V.V. Khoze, D.L. Milne, M. Spannowsky, Searching for QCD instantons at hadron colliders. Phys. Rev. D **103**(1), 014017 (2021). <https://doi.org/10.1103/PhysRevD.103.014017>
33. A. Ringwald, F. Schrempp, Instanton induced cross-sections in deep inelastic scattering. Phys. Lett. B **438**, 217–228 (1998)
34. C.W. Bernard, Gauge zero modes, instanton determinants, and QCD calculations. Phys. Rev. D **19**, 3013 (1979)
35. T.R. Morris, D.A. Ross, C.T. Sachrajda, Higher order quantum corrections in the presence of an instanton background field. Nucl. Phys. B **255**, 115–148 (1985)
36. A. Ringwald, F. Schrempp, Confronting instanton perturbation theory with QCD lattice results. Phys. Lett. B **459**, 249–258 (1999)
37. F. Schrempp, A. Utermann, QCD instantons and high-energy diffractive scattering. Phys. Lett. B **543**, 197–207 (2002)
38. A.H. Mueller, Leading power corrections to the semiclassical approximation for gauge meson collisions in the one instanton sector. Nucl. Phys. B **353**, 44–58 (1991)
39. V.V. Khoze, A. Ringwald, Nonperturbative contribution to total cross-sections in nonAbelian gauge theories. Phys. Lett. B **259**, 106–112 (1991)
40. R. Harnik, T. Wizansky, Signals of new physics in the underlying event. Phys. Rev. D **80**, 075015 (2009)
41. J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaitre, A. Mertens, M. Selvaggi, DELPHES 3, a modular framework for fast simulation of a generic collider experiment. JHEP **02**, 057 (2014)
42. T. Sjostrand, S. Mrenna, P.Z. Skands, A Brief Introduction to PYTHIA 8.1. Comput. Phys. Commun. **178**, 852–867 (2008)
43. P. Skands, S. Carrazza, J. Rojo, Tuning PYTHIA 8.1: the Monash 2013 Tune. Eur. Phys. J. C **74**(8), 3024 (2014)
44. R.D. Ball et al., Parton distributions with LHC data. Nucl. Phys. B **867**, 244–289 (2013)
45. ATLAS Pythia 8 tunes to 7 TeV data, Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva (2014)
46. M. Bahr et al., Herwig++ physics and manual. Eur. Phys. J. C **58**, 639–707 (2008)
47. E. Bothmann, et al., Event Generation with Sherpa 2.2. SciPost Phys. **7**(3), 034 (2019). <https://doi.org/10.21468/SciPostPhys.7.3.034>
48. E. Bothmann et al., Event generation with Sherpa 2.2. SciPost Phys. **7**(3), 034 (2019)
49. R. Kleiss, W. James Stirling, S.D. Ellis, A new Monte Carlo treatment of multiparticle phase space at high-energies. Comput. Phys. Commun. **40**, 359 (1986)
50. J. Bellm et al., Herwig 7.2 release note. Eur. Phys. J. C **80**(5), 452 (2020)
51. M. Cacciari, G.P. Salam, G. Soyez, The anti- k_t jet clustering algorithm. JHEP **04**, 063 (2008)
52. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012)
53. C. Bierlich et al., Robust independent validation of experiment and theory: Rivet version 3. SciPost Phys. **8**, 026 (2020)
54. M. Aaboud, et al., Charged-particle distributions at low transverse momentum in $\sqrt{s} = 13$ TeV pp interactions measured with the ATLAS detector at the LHC. Eur. Phys. J. C **76**(9), 502 (2016). <https://doi.org/10.1140/epjc/s10052-016-4335-y>
55. G. Aad, et al., Measurement of charged-particle event shape variables in $\sqrt{s} = 7$ TeV proton–proton interactions with the ATLAS detector. Phys. Rev. D **88**(3), 032004 (2013). <https://doi.org/10.1103/PhysRevD.88.032004>
56. L. Heinrich, M. Feickert, G. Stark, K. Cranmer, pyhf: pure-Python implementation of HistFactory statistical models. J. Open Source Softw. **6**(58), 2823 (2021). <https://doi.org/10.21105/joss.02823>
57. B. Abelev et al., Underlying event measurements in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV with the ALICE experiment at the LHC. JHEP **07**, 116 (2012)
58. G. Aad et al., Measurement of distributions sensitive to the underlying event in inclusive Z-boson production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Eur. Phys. J. C **74**(12), 3195 (2014)
59. S. Chatrchyan et al., Measurement of the underlying event in the Drell–Yan process in proton–proton collisions at $\sqrt{s} = 7$ TeV. Eur. Phys. J. C **72**, 2080 (2012)