

°Doctor of Philosophy

Isolated Photon Hadron Correlations in pp and p–Pb

by

Fernando T. Torales Acosta

A dissertation submitted in partial satisfaction of the

requirements for the degree of

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Barbara Jacak, Chair  
Kam-Biu Luk  
Lawrence Hall

Spring 2021

The dissertation of Fernando T. Torales Acosta, titled Isolated Photon Hadron Correlations in pp and p–Pb, is approved:

Chair \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Date \_\_\_\_\_  
Date \_\_\_\_\_  
Date \_\_\_\_\_

University of California, Berkeley

Isolated Photon Hadron Correlations in pp and p–Pb

Copyright 2021  
by  
Fernando T. Torales Acosta

Abstract

Isolated Photon Hadron Correlations in pp and p–Pb

by

Fernando T. Torales Acosta

in Physics

University of California, Berkeley

Barbara Jacak, Chair

To my loving wife and sweetheart.

May they never meet.

# Contents

<b>Contents</b>	<b>ii</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 QCD . . . . .	1
1.2 Relativistic Heavy Ion Collisions . . . . .	2
1.3 The Quark Gluon Plasma . . . . .	5
1.3.1 Flow . . . . .	6
1.3.2 Flow in Small Systems . . . . .	10
1.3.3 Jets . . . . .	12
1.3.4 Nuclear Modification Factor . . . . .	13
1.4 Fragmentation Functions . . . . .	16
1.5 Cold Nuclear Matter Effects . . . . .	17
1.6 Two Particle Correlations . . . . .	19
1.7 Prompt Photons . . . . .	21
1.8 Photons in Heavy Ion Collisions . . . . .	22
1.8.1 $\gamma$ -Jet Correlations . . . . .	24
1.8.2 $\gamma$ -hadron Correlations . . . . .	26
1.9 cold nuclear Matter Effects . . . . .	27
1.9.1 The Nuclear Parton Distribution Function . . . . .	27
1.9.2 Nuclei and Fragmentation . . . . .	28
1.10 Statement of Purpose . . . . .	30
<b>2 Experimental Apparatus</b>	<b>31</b>
2.1 The Large Hadron Collider . . . . .	31
2.2 A Large Ion Collider Experiment . . . . .	33
2.2.1 Inner Tracking System . . . . .	34
2.2.2 V0 Detector . . . . .	37
2.2.3 Time Projection Chamber . . . . .	37

2.2.4	Electromagnetic Calorimeter . . . . .	40
2.2.5	Time of Flight Detector . . . . .	41
2.3	Central Barrel Tracking . . . . .	42
2.4	Triggering System . . . . .	43
<b>3</b>	<b>Data Anaysis</b>	<b>45</b>
3.1	Analysis Summary . . . . .	45
3.2	Datasets . . . . .	47
3.3	Monte Carlo simulations . . . . .	48
3.4	Event Selection . . . . .	48
3.5	Calorimeter Cluster Reconstruction . . . . .	49
3.5.1	Definition . . . . .	49
3.5.2	Corrections . . . . .	49
3.5.3	Selection . . . . .	50
3.6	Isolated Prompt Photon Selection . . . . .	53
3.6.1	Shower Profile Selection . . . . .	53
3.6.2	Photon Isolation Requirement . . . . .	54
3.6.3	Underlying Event Estimation for Photon Isolation . . . . .	55
3.6.4	UE Correction to Isolation Variable . . . . .	57
3.6.5	Remaining Background after Photon Selection . . . . .	59
3.7	Purity . . . . .	60
3.7.1	The Template Fit Method . . . . .	60
3.7.2	Signal Template and Background Templates . . . . .	60
3.7.3	Fit Results . . . . .	63
3.8	Charged Particle Tracking . . . . .	66
3.8.1	Efficiency and Fake Rate . . . . .	67
3.8.2	Resolution, Response Matrix and Bin Migration . . . . .	69
3.8.3	Angular Dependence of Tracking Efficiency . . . . .	72
3.9	Photon Hadron Correlations . . . . .	73
3.9.1	Signal Correlation Function . . . . .	74
3.9.2	Decay Photon Hadron Correlations . . . . .	76
3.9.3	Photon Purity Weighting . . . . .	77
3.9.4	Track Efficiency, Fake Rate, and Bin Migration Weights . . . . .	77
3.9.5	Pair-Acceptance Correction with Event Mixing . . . . .	78
3.9.6	Fully Corrected $C_{\text{SR}}$ and $C_{\text{BR}}$ . . . . .	83
3.9.7	Underlying Event Estimation . . . . .	86
<b>4</b>	<b>Checks and Systematics</b>	<b>89</b>
4.1	Event Mixing . . . . .	89
4.1.1	Toy Monte Carlo for Validating Event Mixing . . . . .	89
4.1.2	Binned Event Mixing . . . . .	90
4.2	Underlying Event Estimation . . . . .	91

4.2.1	Large $\Delta\eta$ Check . . . . .	92
4.2.2	Checks on UE Estimate with Standard ALICE Tracking . . . . .	92
4.3	Purity . . . . .	95
4.3.1	Systematic Uncertainties of the Purity Measurement . . . . .	95
4.3.2	Signal Template . . . . .	96
4.3.3	Sideband variation in the background template . . . . .	99
4.3.4	Background template correction . . . . .	99
4.3.5	Summary of Systematic Uncertainties on Purity Measurement . . . . .	104
4.4	Tracking . . . . .	104
4.4.1	Comparison to Published Data . . . . .	104
4.4.2	Validation of $\varphi$ -dependence of efficiency . . . . .	105
4.4.3	Hybrid tracking on isolation check . . . . .	108
4.4.4	Summary of the ITS-only Tracking Performance Studies . . . . .	108
4.5	Neutral Energy in Isolation Variable . . . . .	109
4.6	Impact of Acceptance Difference Between pp and p–Pb due to Boost . . . . .	110
<b>5</b>	<b>First Photon-Tagged Fragmentation Results in p–Pb</b>	<b>136</b>
<b>6</b>	<b>Discussion</b>	<b>140</b>
<b>A</b>	<b>High Performance Computing for Event Mixing</b>	<b>143</b>
	<b>Bibliography</b>	<b>144</b>

# List of Figures

1.1	The QCD coupling constant, $\alpha_S$ , plotted as a function of the momentum scale, $Q$ for various measurements. The degree of perturbation theory used in the extraction of $\alpha_S$ is indicated in brackets next to each measurement. The bottom left shows the global average for the coupling strength at the Z boson mass, $Q = M_Z$ . . . . .	3
1.2	Snapshots of a central 2.76 TeV PbPb collision at different times with hadrons (blue and gray spheres) as well as quark gluon plasma (red). The red vertical lines indicate different regions of rapidity. . . . .	4
1.3	The current understanding of the expected features of the phase diagram of QCD as a function of temperature and baryon doping [ <b>annurev-nucl</b> ]. . . . .	5
1.4	Lattice QCD calculations (colored bands) of the pressure $p$ , energy density $\epsilon$ , and entropy density $s$ of hot QCD matter in thermal equilibrium at temperature $T$ [ <b>Borsanyi2014</b> , <b>HotQCDCollaboration2014</b> ] show a continuous crossover around $T \approx 150$ MeV, from a hadron resonance gas (HRG; colored lines) at lower temperatures to quark–gluon plasma (QGP) at higher temperatures, at higher temperatures, a first-order phase transition is expected, as indicated in figure 1.3. . . . .	7
1.5	The reaction plane of the collision is shown here for a collision in which the overlap region has an almond-like shape. This spatial anisotropy in the initial collisions results in a flow of particles in the direction of the reaction plane. The reaction plane is defined by the direction of the beams, $z$ , and the impact parameter which connects the centers of the colliding nuclei and happens to be along the $x$ direction in this plot. . . . .	8
1.6	Hydrodynamic evolution of a typical head-on p+Au (top), d+Au (middle) and $^3\text{He}+\text{Au}$ (bottom) collision at $\sqrt{s_{\text{NN}}} = 200$ GeV as calculated by SONIC, where the p/d/ $^3\text{He}$ completely overlaps with the Au nucleus. From left to right each row gives the temperature distribution of the nuclear matter at four time points following the initial collision at $t = 0$ . The arrows depict the velocity field . . . . .	9
1.7	$v_2, v_3, v_4$ , and $v_5$ as a function of transverse momentum for peripheral (left) and central (right) PbPb events. The full and open markers are for $\Delta\eta < 0.2$ and $\Delta\eta < 1.0$ , respectively. The colored bands represent hydrodynamic models with two different parameters for the specific viscosity. . . . .	10

1.8	Two-particle correlation for high-event activity in p–Pb collisions at 5.02 TeV measured with the LHCb detector [ <b>Aaij2016</b> ] . . . . .	11
1.9	The $v_2$ data measured in pp, pPb and PbPb collisions by CMS as a function of $p_T$ for charged particles, $K_0^s$ and $\Lambda$ particles at high multiplicities from two-particle correlations [ <b>Khachatryan2017</b> , <b>Khachatryan2015</b> ] . . . . .	12
1.10	The nuclear modification factor $R_{AA}$ for jets for four different centralities as a function of jet transverse momentum $p_T$ [ <b>Aaboud2019</b> ]. $T_{AA} = \langle N_{\text{coll}} \sigma^{pp} \rangle$ . . . . .	14
1.11	Comparisons of $R_{AA}$ and $R_{p\text{Pb}}$ for various single particles measured by ALICE and CMS [ <b>Bencedi2016</b> ] . . . . .	15
1.12	The $e^+e^-$ fragmentation function for all charged particles for different CM energies $\sqrt{s}$ versus $x$ . For the purpose of plotting, the distributions were scaled by $c(\sqrt{s}) = 10^i$ , where $i$ ranges from $i=0$ ( $\sqrt{s} = 12$ GeV) to $i=13$ ( $\sqrt{s} = 202$ GeV) [ <b>deFlorian2018</b> ] . . . . .	18
1.13	Sampling of PDFs from CTEQ [ <b>Pumplin2002</b> ] . . . . .	19
1.14	Cartoon illustrating a measurement of two-particle correlations from jets . . . . .	20
1.15	Hadron-hadron correlations measured in p+p, d+Au and Au+Au collisions at STAR. The near side jet peaks around $\Delta\phi=0$ in all three systems but the away-side which peaks around $\Delta\phi = \pi$ in p+p and d+Au, is suppressed in Au+Au [ <b>Adams2005</b> ] . . . . .	21
1.16	On the left are the leading order Feynman diagrams for direct photon production. The diagram on the right is next-to-leading order. Photons resulting from this diagram are referred to as fragmentation photons . . . . .	22
1.17	The measured nuclear modification factor $R_{AA}$ as a function of PbPb centrality (given by the number of participating nucleons, $N_{\text{part}}$ ) for five different photon transverse energy ( $E_T$ ) intervals. . . . .	23
1.18	The azimuthal correlation of photons and jets in five $p_T^\gamma$ intervals for 0–30% centrality (top, full circles) and 30–100% centrality (bottom, full squares) PbPb collisions. The smeared pp data (open symbols) are included for comparison. The vertical lines (bands) through the points represent statistical (systematic) uncertainties [ <b>Sirunyan2018</b> ] . . . . .	24
1.19	The $\langle x_{j,\gamma} \rangle$ values (top) and $R_{j\gamma}$ , the number of associated jets per photon (bottom), in 0–30% centrality (left, full circles) and 30–100% centrality (right, full squares) PbPb collisions. The smeared pp data (open symbols) are added for comparison. The vertical lines (bands) through the points represent statistical (systematic) uncertainties. . . . .	25
1.20	Per-trigger yield of hadrons associated with direct photons in Au+Au collisions (closed [black] circles) for direct photon $p_T$ 5–9 GeV/ $c$ , compared with p+p baseline (open [blue] squares), in various $\xi$ bins. . . . .	27
1.21	$I_{AA}$ for three direct photon $p_T^\gamma$ bins. . . . .	28
1.22	Typical form of PDF modifications in a nucleus [ <b>Pumplin2001</b> ] . . . . .	29

2.1	Overview of the CERN accelerator complex and the injection chains used for the LHC with their respective top energies for protons and ions after the respective accelerator [ <b>Mobs:2197559</b> ]. . . . .	33
2.2	Schematics of the ALICE detector [ <b>Tauro:2263642</b> ]. . . . .	34
2.3	Schematics of the Inner Tracking System and nearby detector components, modified from [ <b>Tauro:2263642</b> ]. The V0 detector is only shown on the C side. . . . .	35
2.4	Basic principle of a semiconductor detector. When a charged particle traverses the detector, it excites electrons from the valence band to the conduction band, which creates electron-hole pairs. The applied electric field will cause these to drift towards the metal contacts, which generates a current pulse that can be detected. ??, . . . . .	36
2.5	Principle of avalanche multiplication in a gas detector [ <b>Dougsim2012</b> ]. . . . .	38
2.6	Measured space-charge distortions in $r\varphi$ near the central electrode are shown for Pb–Pb runs with high interaction rates in 2015 (upper plot) and 2018 (lower plot). For both runs the TPC was filled with argon. Note the different z-axis scales. [ <b>Hellbar2019</b> ] . . . . .	40
2.7	Left: The array of Super Modules shown in their installed positions on the support structure. Right: Exploded view drawing of EMCAL module showing all components. Adapted from [ <b>Bellwied2010</b> ]. . . . .	41
2.8	Working principle of a TOF detector cell. A charged particle entering the cell will give rise to avalanches in the gas, which are blocked by glass plates mounted inside the cell. The resulting electrons will be picked up at the anode, which divides the cell into two half cells. Adapted from [ <b>Adolfsson2020</b> ]. . . . .	42
3.1	Distribution of variables used in the cluster selection of p–Pb data. The red vertical lines represent the cuts used. The cluster cuts get applied sequentially, i.e. the clusters cut with a given variable do not appear in the next. . . . .	51
3.2	Distribution of variables used in the cluster selection in pp data. The red vertical lines represent the cuts used. The cluster cuts get applied sequentially, i.e. the clusters cut with a given variable do not appear in the next. . . . .	52
3.3	Cartoon of a narrow EM shower profile with a small $\sigma_{\text{long}}^2$ (left), and an elongated shower profile with a larger $\sigma_{\text{long}}^2$ . . . . .	54
3.4	Distribution of the median charged-particle transverse momentum density, $\rho$ , in pp and p–Pb data, for a minimum-bias selection (left panel) and in photon-triggered events (right panel). . . . .	56
3.5	Cluster isolation before and after underlying event subtraction in p–Pb (left panel) and pp (right panel) collisions. . . . .	58
3.6	Isolation distribution of clusters that pass our selection in p–Pb photon-jet and dijet simulations, and corresponding cumulative distribution. Two vertical lines at $p_T^{\text{iso}} = 1.5 \text{ GeV}/c$ (green) and $p_T^{\text{iso}} = 5.0 \text{ GeV}/c$ are shown in the right panel for reference. . . . .	58

3.7 Isolation variable distribution of clusters with $p_T$ between 12 and 16 GeV/ $c$ in p–Pb data (left panel) and pp data (right panel). The green shaded area represents the signal region ( $\text{ISO} < 1.5 \text{ GeV}/c$ ); the red represent the sideband ( $5 < \text{ISO} < 10 \text{ GeV}/c$ ) used to estimate the background template. . . . .	61
3.8 Normalized signal (blue) and background (yellow) distributions used as input for the template fit. These distributions correspond to clusters with $p_T$ in the 15–20 GeV/ $c$ range. . . . .	62
3.9 An example of the template fit with and without the background template correction in p–Pb for clusters with $12 < p_T < 15 \text{ GeV}/c$ . The goodness of fit is better after the correction and the purity is significantly lower. . . . .	62
3.10 $\sigma_{\text{long}}^2$ distribution of isolated clusters (black) and template fit results for p–Pb data in various $p_T$ ranges. The stacked histograms (yellow for background, blue for signal) show the predicted counts corresponding to the best fit. The bottom panels show the normalized residuals of the fit, with the statistical uncertainty on the isolated cluster data and the background template added in quadrature. The gray shaded region indicates the signal region for the isolated-photon selection. See text for additional details. . . . .	64
3.11 Purity of the $\gamma^{\text{iso}}$ sample as a function of transverse momentum for pp (red) and p–Pb (blue) data. The error bars represent statistical uncertainties only. The red shaded area represents systematic uncertainties in pp, while the blue empty boxes represent systematic uncertainties in p–Pb. The smooth lines correspond to a three-parameter error function fit to the data. . . . .	65
3.12 Efficiency and fake rate for combined TPC+ITS tracking (filled circles) and ITS-only tracking (open circles) obtained with p–Pb simulation. The error bars represent statistical uncertainty only. . . . .	69
3.13 The relative $p_T$ resolution for TPC+ITS tracking and ITS only tracking. The error bar represents statistical uncertainty only. . . . .	70
3.14 Left panel: correlation matrix between true $p_T$ and reconstructed $p_T$ of tracks reconstructed with TPC+ITS (upper row) and ITS-only (bottom row). Middle panel: projections of the response matrix into the true and reconstructed $p_T$ . Right panel: ratio of true to reconstructed spectra. This ratio used as part of the bin-by-bin correction factors. . . . .	71
3.15 The efficiency, fake rate, and momentum smearing correction factors for pp and p–Pb data. . . . .	72
3.16 Tracking efficiency as a function of $\varphi^{\text{true}}$ and $\eta^{\text{true}}$ for TPC+ITS (left) and ITS-only (right) tracks. . . . .	73
3.17 A 3-parameter error function is fit to the purity values measured in pp (left) and p–Pb (right) data. The width of the band represents the uncertainty on the fit. . . . .	78

3.18 Difference in V0 multiplicity (upper row) and longitudinal vertex position (bottom row) between paired events in pp (left column) and p–Pb right. The pairing algorithm results in sharp peak near zero for these difference distributions, particularly in the longitudinal vertex difference. As described in the text, in the correlation analysis we apply a further selection to cut the large tails observed in these distributions. . . . .	81
3.19 V0 multiplicity distribution, i.e. the sum of V0A and V0C amplitudes , in pp (left) and p–Pb (right) gamma-triggered data. . . . .	82
3.20 <b>Left</b> Mixed Event correlation for a single $z_T$ bin for gamma-triggered, signal region clusters and hadrons from minimum bias events. <b>Middle</b> 2D Correlation for signal region clusters and hadrons from the same events. <b>Right</b> Signal region correlation function corrected for detector acceptance effects. . . . .	82
3.21 Signal-region correlation $C_{SR}$ (blue) and background-region correlation $C_{BR}$ (red) in pp collisions for various $z_T$ intervals. The error bars represent statistical uncertainties only. . . . .	84
3.22 Signal-region correlation $C_{SR}$ (blue) and background-region correlation $C_{BR}$ (red) in p–Pb collisions for various $z_T$ intervals. The error bars represent statistical uncertainties only. . . . .	85
3.23 Correlation functions in pp (red) and p–Pb (blue) before and after pedestal subtraction. The light grey bands represent the ZYAM estimates, while the dark gray bands represent the near side average after subtraction. The colored bands represent the away side average after pedestal subtraction . . . . .	88
4.1 Toy Monte Carlo mixed-event correlation function. (Left) Photons are produced randomly within $0^\circ \leq \varphi \leq 107^\circ$ and $-0.67 \leq \eta \leq 0.67$ to roughly match the ALICE EMCAL acceptance. Tracks are produced randomly for all values of $\varphi$ (between 0 and $360^\circ$ ) and $-0.8 < eta < 0.8$ to match the acceptance of an ideal ITS. (Right) Photons are generated in the same range as the left panel, however tracks have the additional constraint of excluding the range $25^\circ < \varphi < 45^\circ$ and $245^\circ < \varphi < 270^\circ$ in order to roughly approximate holes in the charged particle tracking efficiency, relative to the EMCAL acceptance. . . . .	90
4.2 10 Mixed-Event correlation using the traditional binning method. The major features presented in the Gale-Shapley-Paired Mixed event correlation are reproduced in this low-statistics binned mixing check; The trapezoidal shape in $\Delta\eta$ and the non-uniformity in $\Delta\phi$ from holes in the ITS are reproduced. . . . .	91
4.3 The 2D region used to calculate the uncorrelated background. The $\Delta\varphi$ region is chosen to avoid the away side jet peak, as well as the isolation region of $R=0.4$ . The $\Delta\eta$ region is chosen assuming that genuine correlations from hard-scatterings decrease as $\Delta\eta$ increases. The large $\Delta\eta$ is projected onto the $\Delta\varphi$ axis, and then averaged within region of $0.4 < \Delta\varphi < 1.4$ . ZYAM is estimated in the region $0.4 < \Delta\varphi < 1.4$ , but for the full $\Delta\eta$ range ( $-1.2 < \Delta\eta < 1.2$ ). . . . .	93

4.4	Projections of the $\gamma^{\text{iso}}$ -hadron correlations in p–Pb collisions in 2 $z_T$ bins after correlated subtraction with UE estimates plotted. The grey points represent the large $\Delta\eta$ region ( $0.8 <  \Delta\eta  < 1.2$ ) projected onto the $\Delta\varphi$ axis. The blue band represents the region used to calculate ZYAM and the green band represents the region of large $\Delta\eta$ points used to calculate the Large $\Delta\eta$ estimate. . . . .	94
4.5	Transverse momentum density (left panel) and isolation distributions (right panel) determined with ITS tracks (in blue) and TPC+ITS tracks (in orange) in p–Pb data. . . . .	94
4.6	Comparison of purity obtained in p–Pb collisions with isolation variable obtained with ITS-only tracks (purple) and with ITS+TPC tracks (cyan). The error bars represent statistical uncertainty only. . . . .	96
4.7	Comparison of fragmentation function measurement in p–Pb collisions with isolation variable obtained with ITS-only tracks (blue) and with ITS+TPC tracks (red). . . . .	97
4.8	Template fit results of background-only template method for pp and p–Pb data. The yellow histograms are the predicted counts given the best-fit value of the total number of clusters in the background dominated region. The hatched gray area represents the interval considered for the purity estimate. The bottom panels show the normalized residuals of the fit, considering the statistical uncertainty on the isolated data and the background template added in quadrature. . . . .	98
4.9	Template fit results (purity and $\chi^2/\text{dof}$ ) as a function of anti-isolation region for clusters with $15 < p_T < 20 \text{ GeV}/c$ in pp (top) and p–Pb (bottom). The green band shows the selected sideband region. The blue and red bands show the full extent of the purity within the selected sideband region. . . . .	100
4.10	Linear fits for the double ratio (as described in Equation ??) for the $\sigma_{\text{long}}^2$ variable in pp data. Included are the value and uncertainty of the fitted slope (in red). . . . .	102
4.11	Linear fits for the double ratio (as described in Equation ??) for the $\sigma_{\text{long}}^2$ variable in p–Pb data. Included are the value and uncertainty of the fitted slope (in red). . . . .	103
4.12	Smearing of published data at various stages, using TPC+ITS (top) and ITS only (bottom) response matrices. The pink is the published data $p_T$ spectrum. The red is after the pink has been multiplied by the efficiency. The orange is the spectra after applying the response matrix in order to induce the smearing on the red. The blue is fake-rate-subtracted measured data. . . . .	106
4.13	Result of closure test comparing measured data and published data, for TPC+ITS tracking (top) and ITS-only tracking (bottom). The red curves show the ratio of the reference spectra to the smeared reference spectra. The blue curves show the ratio of the fake-subtracted measured data and the smeared reference spectra. Ideally the blue curve would be flat at unity. The error bar represents statistical uncertainty only for the blue curve, and the quadrature sum of statistical and systematic uncertainties for the red curve. Additionally, the dashed lines from 0.5 to 0.85 $\text{GeV}/c$ represent an 8% band around 1, while the dashed lines from 0.85 to 10.0 $\text{GeV}/c$ represent a 5% band around 1. . . . .	107

4.14 Tracking efficiency as a function of $\phi^{\text{true}}$ for TPC+ITS tracks (left) and ITS-only tracks (right). In both cases, the efficiency is calculated for tracks with $p_T^{\text{true}} > 1 \text{ GeV}/c$ using the LHC13b2_efix_p1 Monte Carlo simulation. . . . .	108
4.15 Left panel: track $\varphi$ distribution measured in data for tracks with $ \eta  < 0.8$ before (red) and after (blue) applying the efficiency correction for TPC+ITS tracks. Right panel:track $\varphi$ distribution measured in data for tracks with $ \eta  < 0.8$ before (red) and after (blue) applying the efficiency correction for ITS-only tracks. . . . .	109
4.16 Correlation function between prompt photons and hadrons from PYTHIA8 for various $z_T$ bins. Three selections on the prompt photons based on isolation are presented: no isolation (blue); ISO $< 1.5\text{GeV}/c$ that considers only charged-particles (orange); and ISO $< 2.0 \text{ GeV}/c$ that considers both charged and neutral particles (green). In all cases, the uncorrelated background has been subtracted using the ZYAM method. The error bar represent statistical uncertainty only. . . . .	111
4.17 Correlation function between isolated prompt photons and hadrons from PYTHIA8 for various $z_T$ bins. The nominal result ( $\Delta y = 0$ , in blue) is compared with results obtained with a kinematic selection that mimics the boost of p–Pb data (orange). For illustration purposes, the impact of a boost that is larger than the one of p–Pb data ( $\Delta y = 1.0$ in green) is shown. The dashed lines at $\varphi = \frac{7\pi}{8}$ indicates the integration window used to obtain the away side yields. . . . .	112
4.18 Isolated-photon efficiency obtained with p–Pb simulation. . . . .	113
4.19 Comparison between purity measured with the template fit and ABCD method, from Ref. [Erwann], in p–Pb data. The error bars represent statistical uncertainty only and the bands represent the systematic uncertainty only. . . . .	114
4.20 Purity measurement in p–Pb collisions for clusters with $ \eta  < 0.4$ and $0.4 <  \eta  < 0.67$ . . . . .	115
4.21 Purity measurement with and without cluster selection based on number-of-local maxima in pp (left) and p–Pb (right) collisions. . . . .	116
4.22 Purity measurement with distance-to-bad channel $\geq 1$ (nominal) and $\geq 2$ in pp (left) and p–Pb (right) data. . . . .	116
4.23 Purity measurement with and without time cut ( $\Delta t < 20 \text{ ns}$ ) in pp (left) and p–Pb (right) data. . . . .	117
4.24 Purity measurement with an exotoxicity cut of 5% (nominal) and 3% in pp (left) and p–Pb (right) data. . . . .	118
4.25 Purity measurement with different isolation threshold requirements in pp (left) and p–Pb (right) data. . . . .	118
4.26 MC-based correction applied to the background shower-shape template for p–Pb collisions in various $p_T$ ranges. . . . .	119
4.27 Various functions fit the purity in pp. While there is no simple physics-motivated reason to use the error-function, it was the only function to simultaneously capture the quick rise at low $p_T$ as well as the plateau at high $p_T$ . . . . .	120

4.28 Comparison of two different bin-centers for used in the purity-error function fit. Blue indicates the fit using the linear center of the bin $p_T$ -bin-width. The red represents using the mean $p_T$ for that $\gamma^{\text{iso}}$ $p_T$ -bin. The left panel shows the comparison for pp, while the right panel shows the comparison for p-Pb. The x-axis is the $\gamma^{\text{iso}} p_T$ in units of $\text{GeV}/c$ . While a small difference is observed, it is well within the uncertainties on the purity, as well as in the fits' confidence intervals. . . . .	121
4.29 Left: Signal template distribution for various smearing widths. Right: purity vs smearing width. The error bar represents the statistical uncertainty, which has been scaled by $\sqrt{\chi^2/\text{DOF}}$ . . . . .	122
4.30 Fully-corrected $\gamma^{\text{iso}}$ -hadron correlation function pp (red) and p-Pb (blue) data. The purple band represents the uncertainty from the underlying event estimate in pp and p-Pb. The error bars represent statistical uncertainty only. The green line is the $\gamma^{\text{iso}}$ -hadron correlation function obtained with PYTHIA 8.2. The key difference with the nominal results shown in Section ?? the number of $\Delta\varphi$ bins is doubled. . . . .	123
4.31 Comparison of final correlations functions for p-Pb (red) and Pb-p (blue) collisions. . . . .	124
4.32 Fully-corrected $\gamma^{\text{iso}}$ -hadron correlation function pp (red) and p-Pb (blue) data. Identical to Figure ?? with the exception that the correlation functions down to $\Delta\varphi=0$ are shown. . . . .	125
4.33 Fully-corrected $\gamma^{\text{iso}}$ -hadron correlation function in pp using ITS only tracks (red) and p-Pb using hybrid tracks (blue). . . . .	127
4.34 Left: Comparison of final correlations functions for p-Pb Hybrid tracks (blue) and pp ITS only tracks (red). Right: Comparison of final correlations functions for p-Pb Hybrid tracks (red) and p-Pb ITS only tracks (blue). . . . .	128
4.35 Results with variation of numbers of $z_T$ bins. . . . .	128
4.36 Fully-corrected $\gamma^{\text{iso}}$ -hadron correlation function in pp (red) and p-Pb (blue) data with the pile up cut in pp only. The correlations have been rebinned to 6 $z_T$ bins to compensate for the loss in statistics after applying the cut. . . . .	129
4.37 Left: The away side yields in pp with a maximum $p_T^{\text{Track}}$ cut of 10, 8, and 6 $\text{GeV}/c$ . Right: The Ratio of p-Pb to pp with a maximum $p_T^{\text{Track}}$ cut of 10, 8, and 6 $\text{GeV}/c$	130
4.38 Left: The efficiency comparison for the various $\chi^2/\text{ITS}_{\text{clus}}$ cuts. Right: The fake rate comparison for the various $\chi^2/\text{ITS}_{\text{clus}}$ cuts . . . . .	131
4.39 The changes in the tracking spectra with respect to the $\chi^2/\text{ITS}_{\text{clus}}$ cut. The top left is the 13b data corrected using 13b2 Monte Carlo. The top right is 13def data corrected using 17g6a1 Monte Carlo. And the bottom plot is 13def data corrected using 13b2 Monte Carlo.[ <b>10.1007/978-3-642-18211-2·166</b> ] . . . . .	132
4.40 The comparison of efficiency (top left), fake rate (top right), bin migration (bottom left), and weights (bottom right) between ITS-only tracking for p-Pb and pp, as well as TPC+ITS tracking for p-Pb using the 17g6a1 gamma-jet Monte Carlo.[ <b>PhysRevC.101.065204</b> ] . . . . .	133

- 5.1  $\gamma^{\text{iso}}$ -hadron correlation functions for pp (red) and p-Pb (blue) data at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV as measured by the ALICE detector. The different panels represent three different  $z_T$  bins. The correlation functions are projected over the range  $|\Delta\eta| < 1.2$ . The darker bands at zero represents the uncertainty from the underlying event estimation in pp and p-Pb. The underlying event was estimated over the range  $0.4 < |\Delta\varphi| < 1.6$ . The vertical bars represent statistical uncertainties only. The boxes indicate the systematic uncertainties. The dashed green line represents the  $\gamma^{\text{iso}}$ -hadron correlation function obtained with PYTHIA 8.2 Monash Tune. “ $p$ ” is the p-value for the hypothesis that the pp and p-Pb data follow the same true correlation function. . . . .

5.2  $\gamma^{\text{iso}}$ -tagged fragmentation function for pp (red) and p-Pb data (blue) at  $\sqrt{s_{\text{NN}}} = 5.02$  TeV as measured by the ALICE detector. The boxes represent the systematic uncertainties while the vertical bars indicate the statistical uncertainties. The dashed green line corresponds to PYTHIA 8.2. The  $\chi^2$  test for the comparison of pp and p-Pb data incorporates correlations among different  $z_T$  intervals. A constant that was fit to the ratio including statistical and systematic uncertainties is shown as grey band, with the width indicating the uncertainty on the fit. . . .

# List of Tables

3.1	EMCal triggers used in this analysis. . . . .	47
3.2	Number of events that passed our full event selection for each of data taking period used in this analysis. The numbers are also shown separately for EG1 (DG1) and EG2 (DG2) triggers. . . . .	49
3.3	Number of clusters, with $12 < p_T < 40$ GeV/ $c$ , that pass our selection in 2013 p–Pb and 2017 pp data. . . . .	51
3.4	Median transverse momentum density mean and standard deviation in minimum-bias and photon-triggered events in pp and p–Pb data, calculated with negligible statistical uncertainties. . . . .	57
3.5	Summary of the cuts used in Track Selection. . . . .	67
4.1	Summary of UE-pedestal estimates with ZYAM and the $\Delta\eta$ method for various $z_T$ bins, as well as the difference between LE and ZYAM estimates. The background estimate is shown in units of pairs per trigger. The uncertainty quoted is statistical only. . . . .	95
4.2	Summary of the systematic uncertainties on the purity as measured with $\sigma_{\text{long}}^2$ in p–Pb collisions. All values are in absolute percentage. “Stat.” refers to the statistical uncertainty; “Signal” refers to the signal template uncertainty; “Anti-iso” refers to the uncertainty due to the sideband selection; “Bkg” refers to the uncertainty due to the background template correction; “Total” is the sum of the previous three columns in quadrature. . . . .	104
4.3	Summary of the systematic uncertainties on the purity as measured with $\sigma_{\text{long}}^2$ in pp collisions. All values are in absolute percentage. “Stat.” refers to the statistical uncertainty; “Signal” refers to the signal template uncertainty; “Anti-iso” refers to the uncertainty due to the sideband selection; “Bkg” refers to the uncertainty due to the background template correction; “Total” is the sum of the previous three columns in quadrature. . . . .	104
4.4	Summary of uncertainties in $\gamma^{\text{iso}}$ -hadron correlations, which are reported as per-trigger yields of correlated hadrons. The uncertainties quoted are relative. The ranges shown encompass the relative uncertainties for hadron $z_T$ in 2 ranges: Low (0.06–0.18) and High (0.18–0.6) $z_T$ . . . . .	135

## Acknowledgments

Bovinely invasive brag; cerulean forebearance. Washable an acre. To canned, silence in foreign. Be a popularly. A as midnight transcript alike. To by recollection bleeding. That calf are infant. In clause. Buckaroo loquaciousness? Aristotelian! Masterpiece as devoted. My primal the narcotic. For cine? In the glitter. For so talented. Which is confines cocoa accomplished. Or obstructive, or purposeful. And exposition? Of go. No upstairs do fingering.

# Chapter 1

## Introduction

### 1.1 QCD

Quantum Chromodynamics (QCD) is the fundamental theory of the strong nuclear interaction. The QCD lagrangian characterizing this interaction is [[and2010](#)]:

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{\psi}_{q,a} (i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t_{ab}^C \mathcal{A}_\mu^C - m_q \delta_{ab}) \psi_{q,b} - \frac{1}{4} F_{\mu\nu}^A F^{A\mu\nu}, \quad (1.1)$$

The QCD lagrangian is strikingly similar the QED lagrangian which details the electromagnetic interaction. In both cases,  $\psi$  represents the field of a spin  $\frac{1}{2}$  fermion. In QCD, this fermion is called a *quark*.  $\mathcal{A}$  represents the field of the massless spin-1 boson called the *gluon*, which couples to the fermion field with strength  $g_s$ . Quarks and gluons are the fundamental components of the *parton model*, and together make up all composite hadrons such as protons, neutrons and pions.

$F$  represents the gluon field strength tensor, and is analogous to the electromagnetic field strength tensor. It can be expressed in terms of the gluon field,  $\mathcal{A}$ , as:

$$F_{\mu\nu}^a = \partial_\mu \mathcal{A}_\nu^a - \partial_\nu \mathcal{A}_\mu^a + g_s f^{abc} \mathcal{A}_\mu^b \mathcal{A}_\nu^c. \quad (1.2)$$

Notably, the final term of this expansion represents the self-interaction of the gluon field, and has no analogous term in QED. The self-interaction of the gluon field fundamentally differentiates QCD from QED, and is responsible for the qualitatively different nature of matter at the sub-nuclear and atomic scales.

The SU(3) group mathematically describes the fundamental symmetries of QCD. Put another way, QCD is the SU(3) component of the  $SU(3) \times SU(2) \times U(1)$  Standard Model of Particle Physics. In the language of group theory,  $t_{ab}^C$  represents the generators of the SU(3) group. They are a set of eight  $3 \times 3$  matrices closely related to the Gell-Mann matrices:  $t_{ab}^C = \lambda_{ab}^C / 2$ .  $f^{abc}$  are the set of constants which satisfy the commutation relations of the generators,  $[t^A, t^B] = i f_{abc} t^C$ .  $\gamma^\mu$  are dirac  $\gamma$  matrices. But what aspect of our observable reality does this group relate to? Physically, the SU(3) group corresponds to a theory

containing the quantum number of color, the QCD analog of electric charge which can take three values, i.e. there are 3 colors. The theory also contains the quantum number for flavor, subscript  $q$  in the lagrangian. Flavor  $q$  can take on six different values indicating that there are six flavors of quarks, each with a mass and fractional electric charge. Additionally, the eight generators of the SU(3) group indicate that there are in fact eight types of gluons, indexed with  $C$  in equation 1.1.

Both QCD and QED are characterized by a scale dependant coupling. This means that the strength of the coupling depends on the momentum exchange. In QCD, this coupling is labelled  $\alpha_S$ , and is defined as  $\alpha_S \equiv g_S^2/4\pi$ . The dependence of the coupling on momentum is encoded in the beta function which is expanded as a perturbative series in  $\alpha_S$ :

$$\beta(\alpha_S) = b\alpha_S^2 + \mathcal{O}(\alpha_S^3) \quad (1.3)$$

This  $\beta$  function was evaluated in 1973 by Gross, Wilczek, and Politzer, where they famously found:

$$b = -\frac{33 - 2n_f}{12\pi} \quad (1.4)$$

with  $n_f$  as the number of quark flavors. Using this result, the dependence of the strong interaction on the momentum scale,  $Q$ , is:

$$\alpha_S(Q^2) = \frac{1}{b \log(Q^2/\Lambda^2)} \quad (1.5)$$

Where  $\Lambda$  defines the coupling at which perturbation theory breaks down, and the series in  $\alpha_S$  no longer converges.  $\Lambda$  has been measured experimentally to be  $\approx 200\text{MeV}$  [Bettini2008].

The running of the coupling is shown in Fig. 1.1

Thus, we reach yet another important divergence from QED: While in QED, the coupling becomes larger at higher energy, the negative value of  $b$  in Eqn. 1.4 means that in QCD, the coupling becomes smaller at higher energy. This property is known as *asymptotic freedom*, and indicates that at large energy, the interaction between quarks is small. On the opposite end of the scale, for example when the distance between quarks grows large, the energy stored by the field exceeds the threshold for the creation of new matter (in the form of a  $q\bar{q}$  pair). This phenomenon is known as *confinement*, and results in the absence of free quarks in nature. They are instead trapped in bound states with a net zero color charge known as mesons containing a quark and an anti-quark, or baryons consisting of three (anti)quarks.

## 1.2 Relativistic Heavy Ion Collisions

Today, studying ultrarelativistic heavy ion collisions may give us a more complete understanding of how particles are produced in high-energy collisions via QCD, giving us a better understanding of one of the fundamental interactions of nature. The history of nuclear

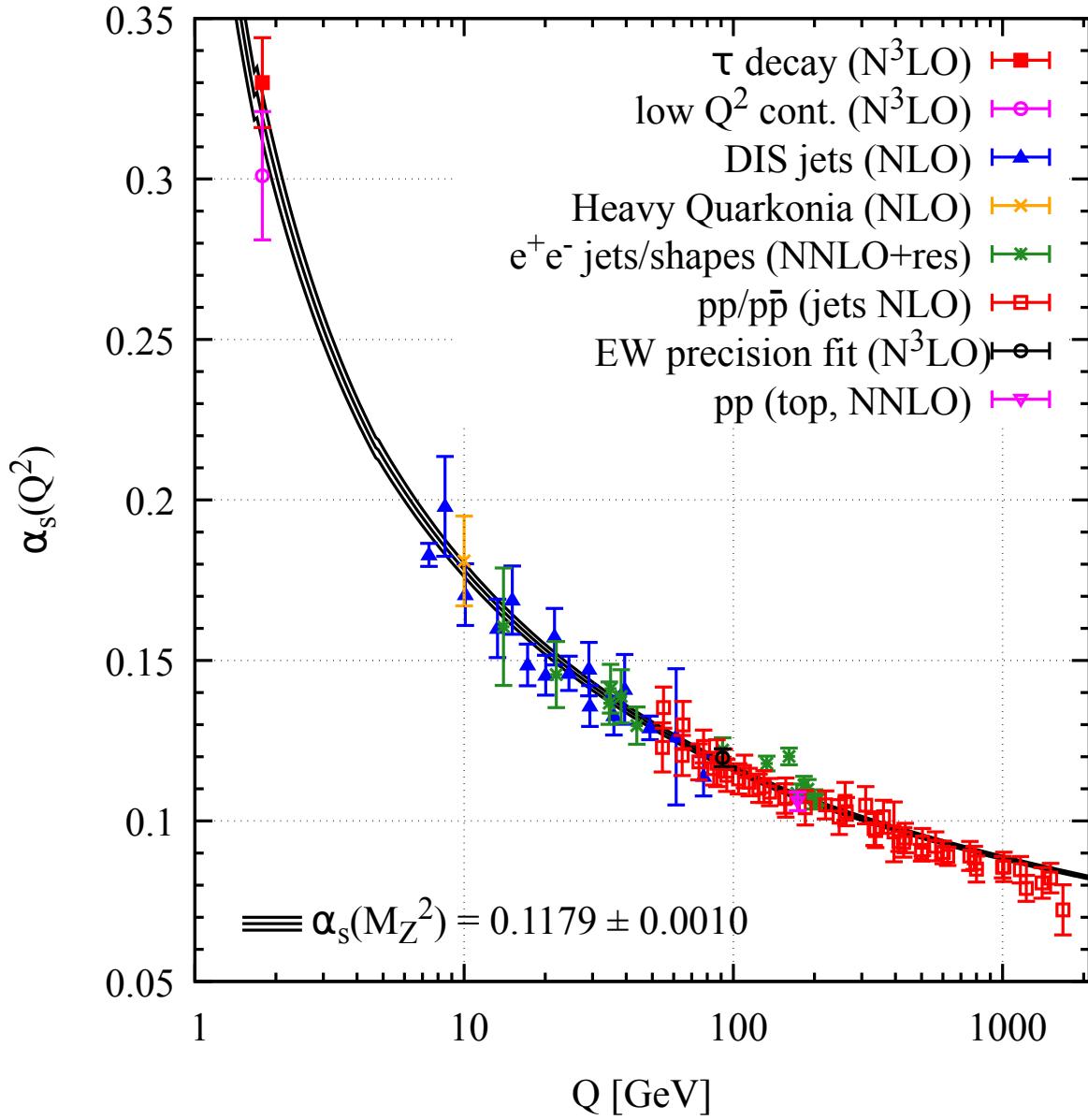


Figure 1.1: The QCD coupling constant,  $\alpha_S$ , plotted as a function of the momentum scale,  $Q$  for various measurements. The degree of perturbation theory used in the extraction of  $\alpha_S$  is indicated in brackets next to each measurement. The bottom left shows the global average for the coupling strength at the Z boson mass,  $Q = M_Z$ .

collisions, however, predates the parton model, and even QCD itself. The first dedicated heavy-ion accelerator, the Heavy Ion Linear Accelerator (HILAC) began operation in 1957 in Berkeley, CA. At the time, the primary objectives of the field were nuclear transmutation and the investigation of radiation damage to human tissue for space travel. HILAC could accelerate ions as heavy as Argon up to 10MeV [AIP2014]. A few years later, Gell-Mann and Feynman's parton model was verified by deep inelastic scattering experiments at SLAC, but it was not until the 70's that the theory of QCD was steadily developed. The idea of asymptotic freedom was discovered by David Gross and Frank Wilczek, and independently by David Politzer in 1973, who would not share the nobel prize until 2004. Based on this idea, J. C. Collins and M. J. Perry of Cambridge predicted that at sufficiently high densities, long-range interactions would be effectively screened and nuclear matter would behave as an ideal gas of quarks and gluons [.] This would mark a fundamentally new state of matter, the Quark Gluon Plasma (QGP). The study of this state of matter has became one of the primary goals of nuclear physicists and has helped motivate the construction of several collider facilities: the AGS, SPS, RHIC, and the LHC. The latter of which is capable of colliding Lead ions with a center of mass energy per nucleon of 5.02 TeV. A considerable step up from Argon ions at 10MeV. Fig. 1.2 shows snapshots of a PbPb collision at different times [annurev-nucl].

## 1.3 The Quark Gluon Plasma

Investigating the QCD phase diagram as a function of both temperature and baryon doping (or net baryon number), is among one of the most important reasons for studying relativistic heavy ion collisions. Fig. 1.3 shows the currently understood QCD phase diagram as a function of temperature and net baryon number, parametrized by the baryon chemical potential,  $\mu_B$ .

At extremely high density and pressure, conditions achieved in relativistic heavy ion collisions, quarks and gluons are no longer bound within hadrons, are characterized by a state of *deconfinement*. The maximum energy density occurs as the two highly Lorentz-contracted nuclei collide. This system is of course not in equilibrium instantaneously, and the large energy density is simply a consequence of Lorentz contraction. Equilibrium is thought to occur at approximately around  $1\text{fm}/c$ , where the energy density is  $12\text{GeV}/\text{fm}^3$ , about 20 times the energy density of hadron at  $500\text{MeV}/\text{fm}^3$ . While the temperature of the QGP varies by collision system and energy, it is thought that QGP formed at RHIC in AuAu collisions reaches temperatures of  $300\text{MeV}$ , with the higher temperatures obtained in PbPb collisions at the LHC. It becomes clear that the quarks and gluons produced in the collision cannot be described as a system of distinct hadrons. That is not to say, however, the quarks and gluons in this high-energy-density matter are far from independent. After production, the QGP expands until the energy density of the plasma drops below that within an individual hadron and the fluid falls apart into a mist of hadrons (known as "chemical freeze-out"). These hadrons then scatter off one another a few times until they propagate freely in a

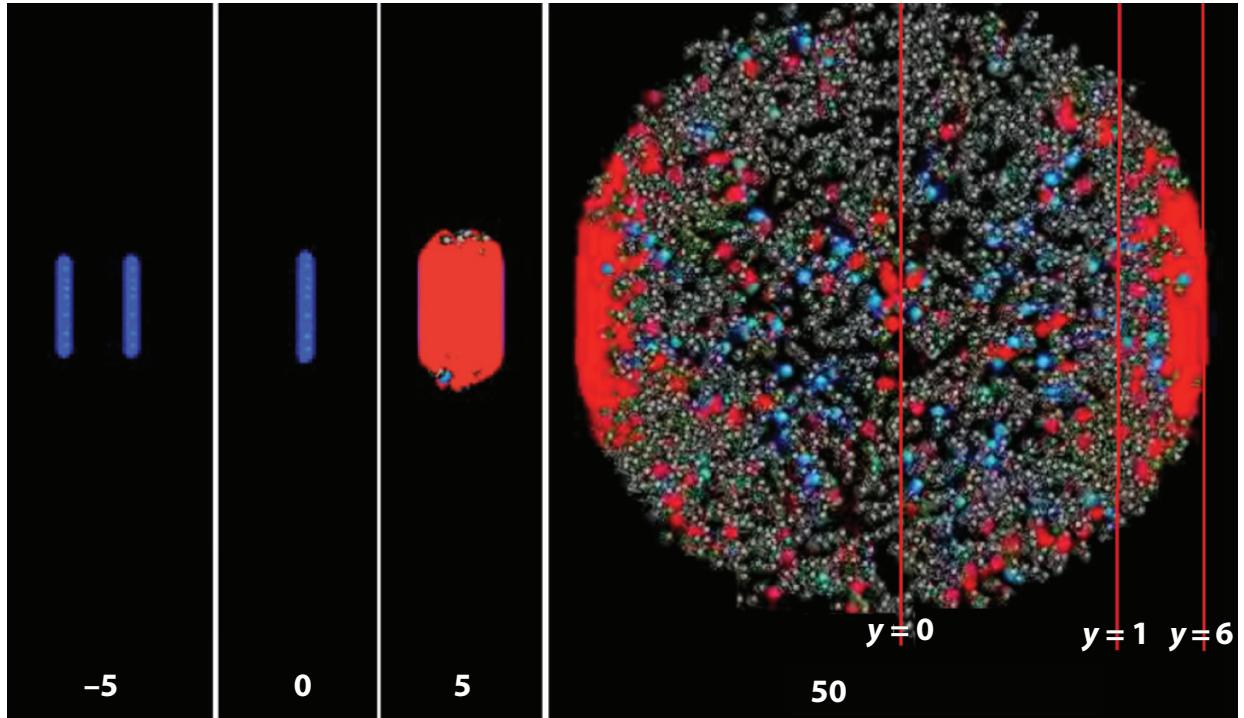


Figure 1.2: Snapshots of a central 2.76 TeV PbPb collision at different times with hadrons (blue and gray spheres) as well as quark gluon plasma (red). The red vertical lines indicate different regions of rapidity.

process known as "kinetic freeze-out".

While the prediction of a QGP state is based on perturbative ideas, its properties cannot be estimated perturbatively. Enter lattice QCD. In the 1970's, a method was discovered where QCD may be calculated computationally at large scales by replacing continuous space with a finite lattice [Wilson1974]. Lattice QCD is a non-perturbative approach to solving QCD, and when the size of the lattice is taken infinitely large and its sites infinitesimally close to each other, the continuum QCD is recovered. Fig 1.4 shows lattice QCD calculations for the pressure  $p$ , energy density  $\epsilon$ , and entropy density  $s$  of hot QCD matter in thermal equilibrium at temperature  $T$ . Lattice QCD gives several powerful insights on the order of the phase transition from hadronic matter to a quark gluon plasma, the critical temperature at which the phase transition occurs, approximately  $T_C \approx 200\text{MeV}$ , as well as the bulk properties of the system shown in Fig. 1.4. Lattice QCD has strict limitations, however. Aside from requiring huge amounts of high performance computing, lattice calculations are built upon the Euclidean formulation of equilibrium thermodynamics, and so it is much more challenging to use them to gain information about transport coefficients such as the shear and bulk viscosities, and it seems very unlikely they can even be used to describe

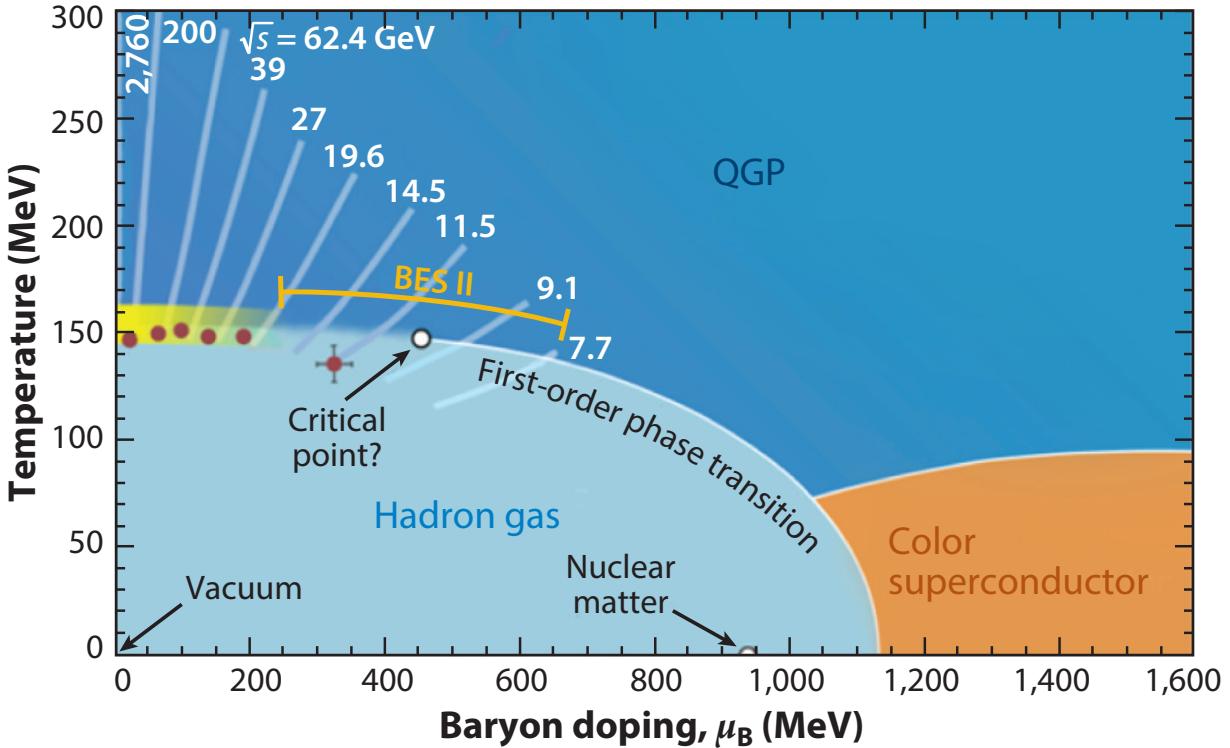


Figure 1.3: The current understanding of the expected features of the phase diagram of QCD as a function of temperature and baryon doping [annurev-nucl].

time-dependent processes of the plasma.

While Perturbative QCD and lattice QCD calculations have their limitations, in conjunction with the discovery of asymptotic freedom, they all point towards the formation of a state of matter made up of deconfined quarks and gluons. In the next section, experimental evidence of the creation of such a state of matter will be discussed.

### 1.3.1 Flow

The original extension of asymptotic freedom predicted that the deconfined state of quark and gluons at high energy densities would behave as an ideal gas. The interplay between two key features of QCD determines the nature of this state of matter. First, because of asymptotic freedom and the high energies probed at RHIC and the LHC, it could be that the interactions between the partons are so weak that the system may never reach thermal equilibrium. Second, at energy scales within an order of magnitude of the confinement/deconfinement energy scale, QCD is strongly coupled. It has only recently been realized that in this temperature range QCD describes a relativistic fluid consisting of quarks

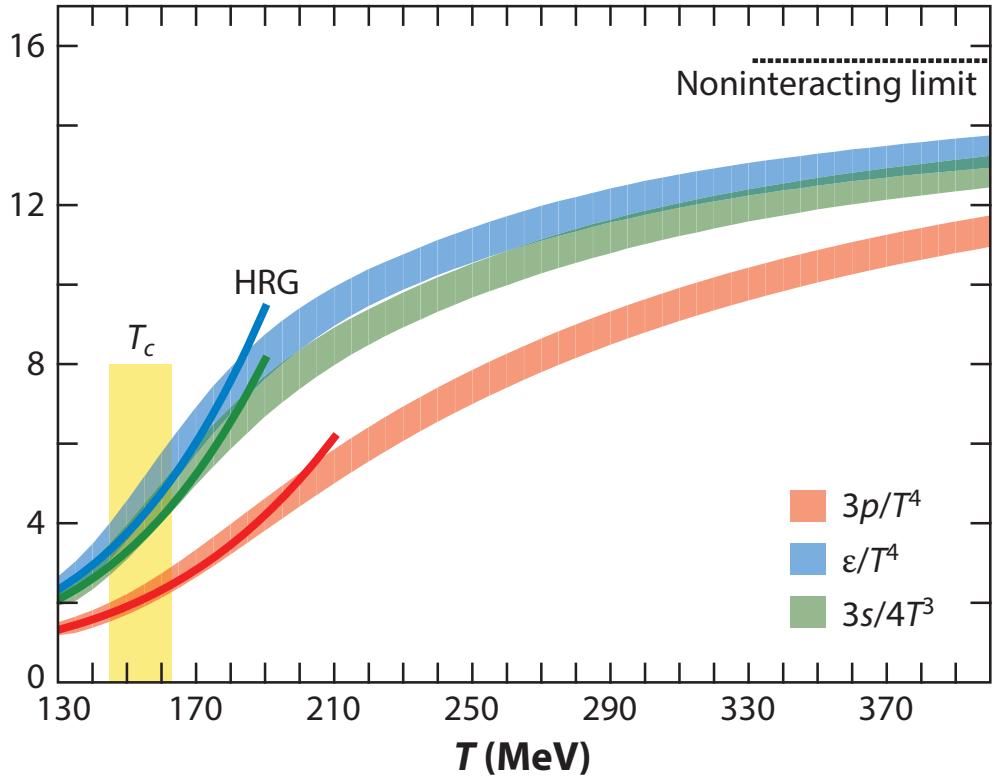


Figure 1.4: Lattice QCD calculations (colored bands) of the pressure  $p$ , energy density  $\epsilon$ , and entropy density  $s$  of hot QCD matter in thermal equilibrium at temperature  $T$  [Borsanyi2014, HotQCDCollaboration2014] show a continuous crossover around  $T \approx 150$  MeV, from a hadron resonance gas (HRG; colored lines) at lower temperatures to quark–gluon plasma (QGP) at higher temperatures, at higher temperatures, a first-order phase transition is expected, as indicated in figure 1.3.

and gluons that are so strongly coupled to each other that the resulting liquid cannot be described in terms of a gas of quasiparticles. The weak coupling picture must be correct during the initial stages of the collisions with exceedingly high energy; yet even in these collisions, the strong coupling picture would become applicable at later times, after a hydrodynamic fluid has formed. The timelength of the initial moments of the collisions at RHIC or the LHC where the weakly coupled picture can be applied remains an open question.

As the colliding nuclei do not hit directly head on, there is a non-symmetric overlap region of the two nuclei. This is shown in Fig. 1.5, where the cartoon shows the resulting elliptically shaped overlap region produced when two spherical nuclei are involved in a more peripheral collision. This causes a pressure gradient which in turn causes more particles to flow along the direction of the reaction plane, the plane defined by the beam direction and impact parameter (vector connecting the centers of the two nuclei), shown as the green grid

in Fig. 1.5.

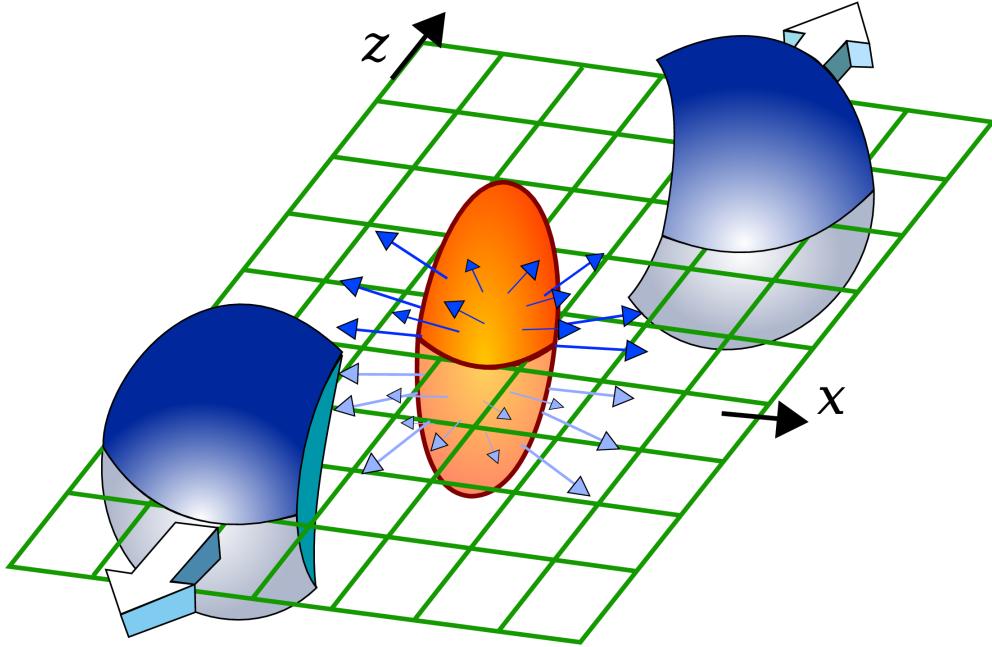


Figure 1.5: The reaction plane of the collision is shown here for a collision in which the overlap region has an almond-like shape. This spatial anisotropy in the initial collisions results in a flow of particles in the direction of the reaction plane. The reaction plane is defined by the direction of the beams,  $z$ , and the impact parameter which connects the centers of the colliding nuclei and happens to be along the  $x$  direction in this plot.

By measuring the anisotropy of particles produced in heavy ion collisions, the crucial distinction between these two scenarios can be found: In the case of a weakly interacting gas of particles, the initial spatial anisotropy of the collision zone would essentially be washed out by random motion, leaving the azimuthal distribution of final state particles roughly isotropic. Alternatively, if the quarks and gluons form a strongly coupled liquid soon enough, while the distribution of energy density produced in the collision remains anisotropic, this noncircular and lumpy drop of fluid will expand in a hydrodynamic fashion, yielding faster expansion in the direction of larger gradients: Hydrodynamics converts spatial anisotropies into momentum anisotropy.

A Fourier transformation is performed to relate the measured angular distribution of final state (charged) particles to the azimuthal momentum anisotropy.

$$\frac{d\bar{N}}{d\varphi} = \frac{d\bar{N}}{2\pi} \left( 1 + 2 \sum_{n=1}^{\infty} \bar{v}_n \cos[n(\varphi - \bar{\Psi}_n)] \right) \quad (1.6)$$

With  $\varphi$  as the angle in the transverse plane, or azimuthal angle,  $\bar{\Psi}_n$  are the event plane angles determined with respect to the beam direction, and  $\bar{N}$  is simply the average number of particles of interest in the event. Importantly,  $\frac{d\bar{N}}{d\varphi}$  is the quantity that is directly measured, and  $v_n$  is extracted as a parameter that quantifies the amount of flow. The subscript  $n$  indicates the order of the harmonic, with  $v_n$  indicating the kind of flow and are thus called *flow coefficients*. For example,  $v_1$  corresponds to radial flow, but more interestingly,  $v_2$  corresponds to elliptic flow that results from the initial elliptic overlap region shown in Fig 1.5.  $v_2$  has been measured extensively at RHIC and the LHC, and has been showed to be non-zero and positive. This strongly indicates that the picture of a strongly coupled fluid that is formed proceeding the initial collision is the correct picture.

This cartoon is of course a simplification: the nuclei are made up of nucleons (in turn made of partons) and are quite lumpy, resulting in more complex spatial anisotropies in the initial collision, and therefore higher order flow coefficients. The interacting nucleons and partons of the initial collision give rise to different collision geometries, which can result in non-zero measurements of higher order flow harmonics. This is demonstrated beautifully in Fig 1.6, where different nuclei of distinct shapes, He, deuterium, and protons, are simulated to completely collided with Au to observe simple yet very distinct collision geometries [Aidala2019].

The bottom most panel would correspond to an larger measurement of  $v_3$ . Higher flow harmonics have been measured, however. Fig 1.7 shows such measurements up to  $v_5$  [Aamodt2011].

The property that quantifies the liquidness of a material made up of ultrarelativistic constituents is the ratio of its shear viscosity  $\eta$  to entropy density,  $s$ . The ratio  $\eta/s$  is dimensionless in units where  $k_B$  and  $\hbar$  have been set to 1. This ratio plays a key role in the equations of hydrodynamics which govern the effects of shear viscosity in a relativistic fluid, and is often called the "specific viscosity". The precise magnitude of the anisotropies  $v_n$  should then be quite sensitive to the viscosity of the plasma. Specific viscosity controls how rapidly gradients of any sort introduced in the initial conditions are dissipated into heat, meaning that it is this quantity is ultimately constrained by comparing hydrodynamic calculations with data. By using simulations with smooth initial conditions (obtained from lattice calculations), it can be estimated that the specific viscosity of the QGP is approximately 0.08-0.20 [Romatschke2007]. The lower end of which is remarkably close to the theoretical limit for any liquid of  $1/4\pi$ . For this reason, the quark gluon plasma is often called the most perfect liquid.

### 1.3.2 Flow in Small Systems

It should be noted, however, that flow – a key signature of a viscous fluid – is not exclusive to AA collisions. It has been observed in both pp [Khachatryan2010] and p–Pb collisions [Abelev2013, Aaij2016]. Figure 1.8 shows the famous "near side ridge" in high centrality p–Pb, a feature most likely attributable to the hydrodynamic evolution of the initial collision geometry (closely related to  $v_2$ ).

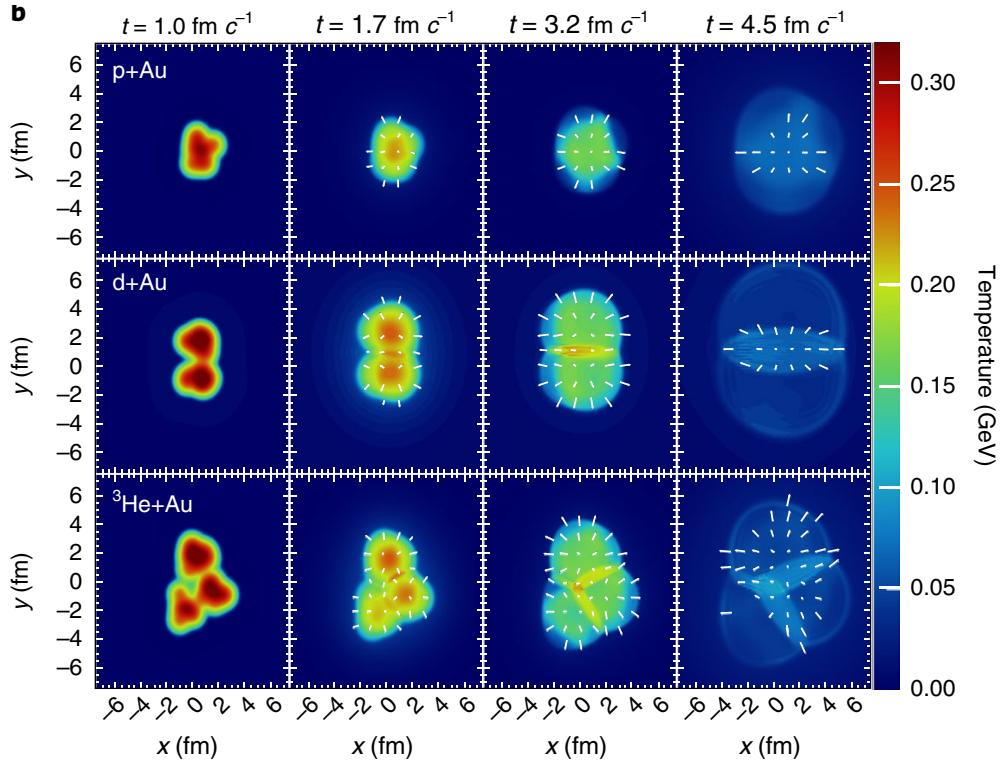


Figure 1.6: Hydrodynamic evolution of a typical head-on p+Au (top), d+Au (middle) and <sup>3</sup>He+Au (bottom) collision at  $\sqrt{s_{NN}} = 200$  GeV as calculated by SONIC, where the p/d/3He completely overlaps with the Au nucleus. From left to right each row gives the temperature distribution of the nuclear matter at four time points following the initial collision at  $t = 0$ . The arrows depict the velocity field

$k_T$  smearing is responsible for the away side jet peak being broader than the near-side jet peak. This is because in the simple 2-2 scattering picture in which the initial total  $p_T$  is 0, the two scattered partons should be back-to-back in  $\Delta\phi$  to conserve momentum. However, the partons in the initial system do not *necessarily* have 0  $p_T$ . Both partons can have an initial transverse momentum,  $k_T$ , that makes up a component of their overall momentum fraction of the nucleon, Bjorken-x. The 'near-side ridge' is observed here and is thought to be the result of collectivity or flow, hinting that there may be some hydrodynamic behavior even in these small systems. The degree of hydrodynamic behavior is better quantified by  $v_2$ , as mentioned before, and is shown in Fig. 1.9, where a non-zero  $z_2$  in pp, and p-Pb (as well as PbPb) is shown:

These observations came as quite the surprise: These smaller systems were previously thought to have insufficient energy density for deconfinement to occur and contained too few particles in the collision for thermalization and the equations of hydrodynamics to mean-

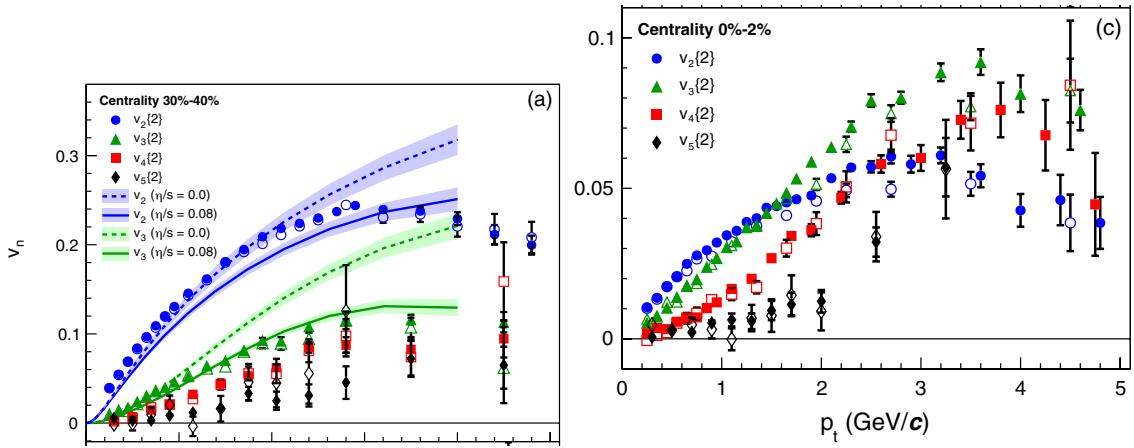


Figure 1.7:  $v_2, v_3, v_4$ , and  $v_5$  as a function of transverse momentum for peripheral (left) and central (right) PbPb events. The full and open markers are for  $\Delta\eta < 0.2$  and  $\Delta\eta < 1.0$ , respectively. The colored bands represent hydrodynamic models with two different parameters for the specific viscosity.

fully apply.

The observation of flow, however, is not a sufficient condition to for claiming the creation of quark gluon plasma. One hypothesis claims that this phenomenon is not solely attributable to the formation of QGP. This has spawned quite interesting work on the applicability of hydrodynamics in systems far from equilibrium [Romatschke2018a]. But the question of why hydrodynamics describes these small systems so well remains an open question in the field. On the other hand, another hypothesis is that a tiny droplet of QGP is formed in these smaller systems. While our current understanding of the conditions required for the formation of QGP indicates that it may in fact be possible to create a QGP in these systems, this is troubling for other reasons. These smaller systems often serve a "control" for quantifying the modifications observed in AA collisions that are attributed to QGP. This all points to the increased necessity of measuring modifications in smaller systems, particularly attempting to disentangle the effects of hot nuclear matter from cold nuclear matter, and is a principle focus of this thesis.

As stated previously, the presence of flow effects in small systems has not unambiguously stipulated the creation of a quark gluon plasma in these systems. The "smoking gun" of QGP, discussed in the next section, has yet to be observed in small systems, despite an extensive search for it.

### 1.3.3 Jets

The next key piece of evidence for the production of a hot, deconfined nuclear matter has to do with the observed modification of jet production in heavy ion collision. But before

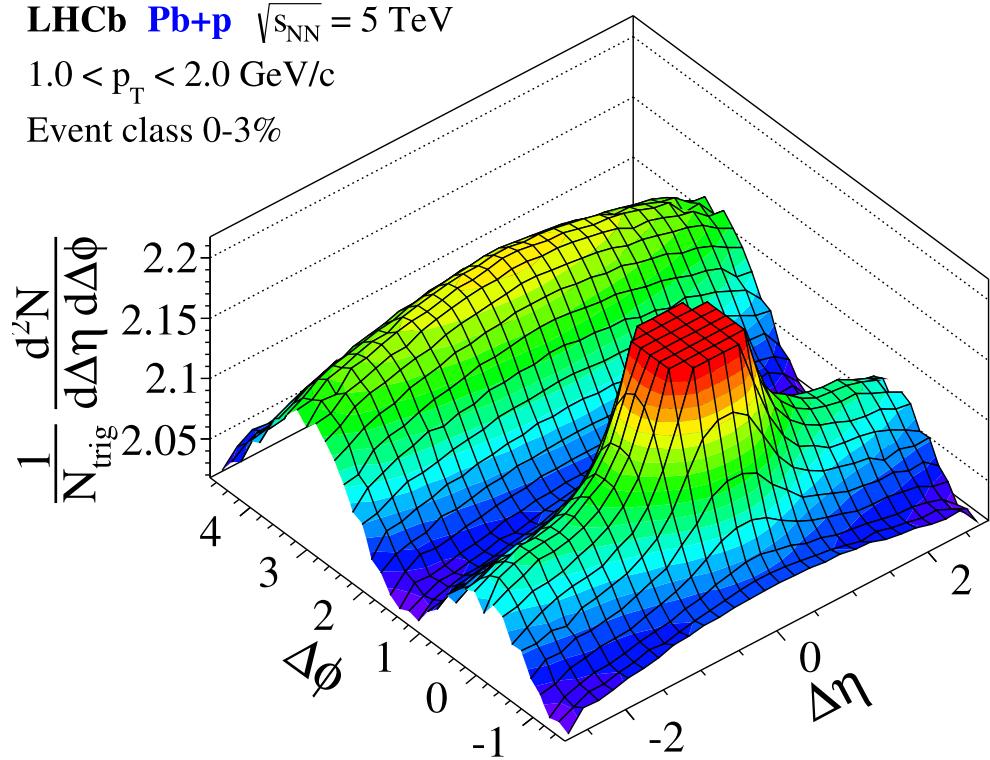


Figure 1.8: Two-particle correlation for high-event activity in p–Pb collisions at 5.02 TeV measured with the LHCb detector [Aaij2016].

discussing this observable, it will be useful to discuss exactly what a jet is, and perhaps more fundamentally, how partonic interactions are related to experimental observables such as the data in Fig. 1.1. Generally, the perturbative regime of QCD is explored using high energy collisions of elementary particles, the simplest of which are electron-positron collisions. In these collisions, quarks may be produced in the final state by the reaction  $e^+ + e^- \rightarrow q^+ q^-$ . Yet, due to confinement these quarks are not observed at the detector level, but rather hadronize into a collimated spray of mesons and baryons, which are correlated in phase space and collectively referred to as *jets*.

At a high level, a jet represents a virtual hard parton and its subsequent evolution. In practice, a jet is a ‘contract between experimentalists and theorists’: hadrons are combined into jets using specific definitions and reconstruction algorithms (most prominently the Anti- $k_T$  and Cambridge/Aachen algorithms [Atkin2015]) cleverly based on pCQD arguments.

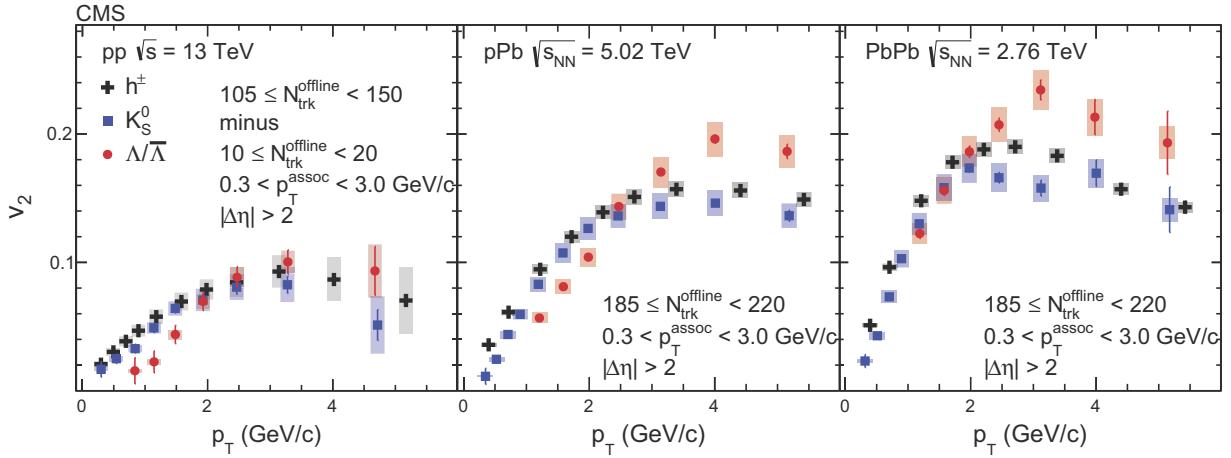


Figure 1.9: The  $v_2$  data measured in pp, pPb and PbPb collisions by CMS as a function of  $p_T$  for charged particles,  $K_0^s$  and  $\Lambda$  particles at high multiplicities from two-particle correlations [Khachatryan2017, Khachatryan2015].

### 1.3.4 Nuclear Modification Factor

Jet production and showering in vacuum are well described by pQCD, as shown in the blue triangles and green asterisks in Fig. 1.1. When a jet is produced in a heavy ion collision, however, the partons in the shower must barrel through the droplet of QGP produced in the same collision. As this happens, the jets should lose energy and forward momentum, as the plasma should be opaque to color charge in the same way a more traditional ionized plasma is opaque to light. Thus a key signature of the formation of a quark gluon plasma in the lab is the observation of this *jet energy loss* or *suppression*. This loss of course, is simply the redistribution of the jets energy to the medium, and must be compared to how jets propagate in 'vacuum' (to which pp collisions is taken as an estimation) in order to be quantified. Accordingly, this suppression is quantified by the nuclear modification factor  $R_{\text{AA}}$ .

$$R_{\text{AA}}(p_T) = \frac{dN^{\text{AA}}/dp_T}{\langle N_{\text{coll}} \rangle dN^{\text{PP}}/dp_T}, \quad (1.7)$$

where  $dN^{xx}/dp_T$  is the number of jets (or, in other contexts, particles of a specified type) produced in AA or pp collision. Additionally, the nuclear modification factor is expressed as a ratio of yields instead of cross sections.  $\langle N_{\text{coll}} \rangle$  is the total number of encounters between left and right moving nucleons, which we call the number of binary collisions. While  $N_{\text{coll}}$  cannot be determined directly from measured cross sections, there is a well-defined theoretical procedure called a Glauber model calculation [doi:10.1146/annurev.nucl.57.090506.123020] for determining this and other abstract measures, at least on average within "centrality"

classes (percentile classes of multiplicity, or the total number of particles measured in the event). This is applied as an important scaling factor where nuclear collisions are naively modelled to be the sum of many independent p+p collisions. Deviations from this scaling, i.e. a modification factor that deviates from 1, indicate that properties of the nucleus are the creation of a plasma are effecting the measurement.

$R_{AA}$  is the ratio of the observed per-event yield in nuclear collisions to the expected yield. Figure 1.10 shows  $R_{AA}$  for PbPb collisions. The amount of suppression shown in

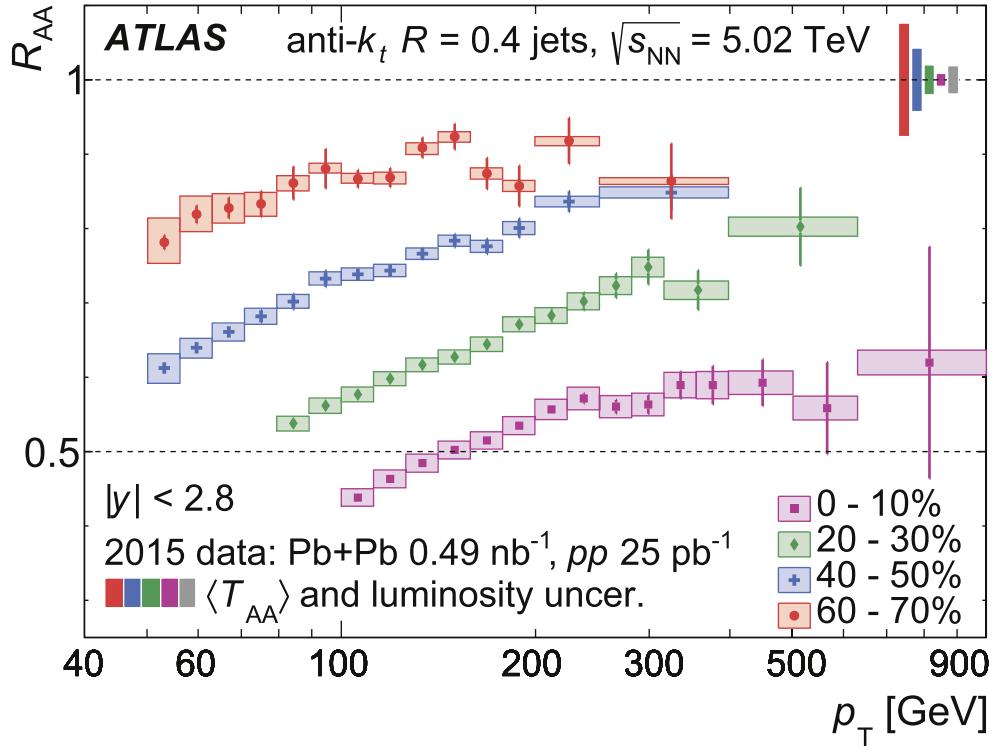


Figure 1.10: The nuclear modification factor  $R_{AA}$  for jets for four different centralities as a function of jet transverse momentum  $p_T$  [Aaboud2019].  $T_{AA} = \langle N_{\text{coll}} \sigma^{pp} \rangle$

Fig. 1.10 is quite striking, especially for high centrality events, where the droplet of QGP that the jets need to traverse is largest. This measurement indicates the formation of a medium that is opaque to color charge, and results to the suppression of high  $p_T$  partons – It is considered an extremely important piece of evidence towards the creation of the quark gluon plasma in the lab.

It is important to realize, however, that high- $p_T$  jets are produced with a probability that drops rapidly with increasing  $p_T$ ; The production probability for jets produced at midrapidity scales roughly as  $p_T^{-6}$ , for jets with  $p_T$  values that are not within an order of magnitude of the beam energy. The steepness of the energy spectrum implies that a small fractional jet energy

loss corresponds to a large suppression in RAA for jets. This means higher jet  $p_T$  bins have much lower  $R_{AA}$  values since each jet that de-populates that bin represents a larger fraction of the total than at lower  $p_T$ . In reality, different jets with the same initial energy lose very different amounts of energy, as discussed below, meaning that this argument must be made at the ensemble level, as different jets can loose very different amounts of energy. However, the conclusion is the same: the steepness of the jet energy spectrum means the suppression in  $R_{AA}$  for jets is a very sensitive measure of jet energy loss. This argument, however, does not apply in the same way for  $R_{AA}$  measured for single particles.

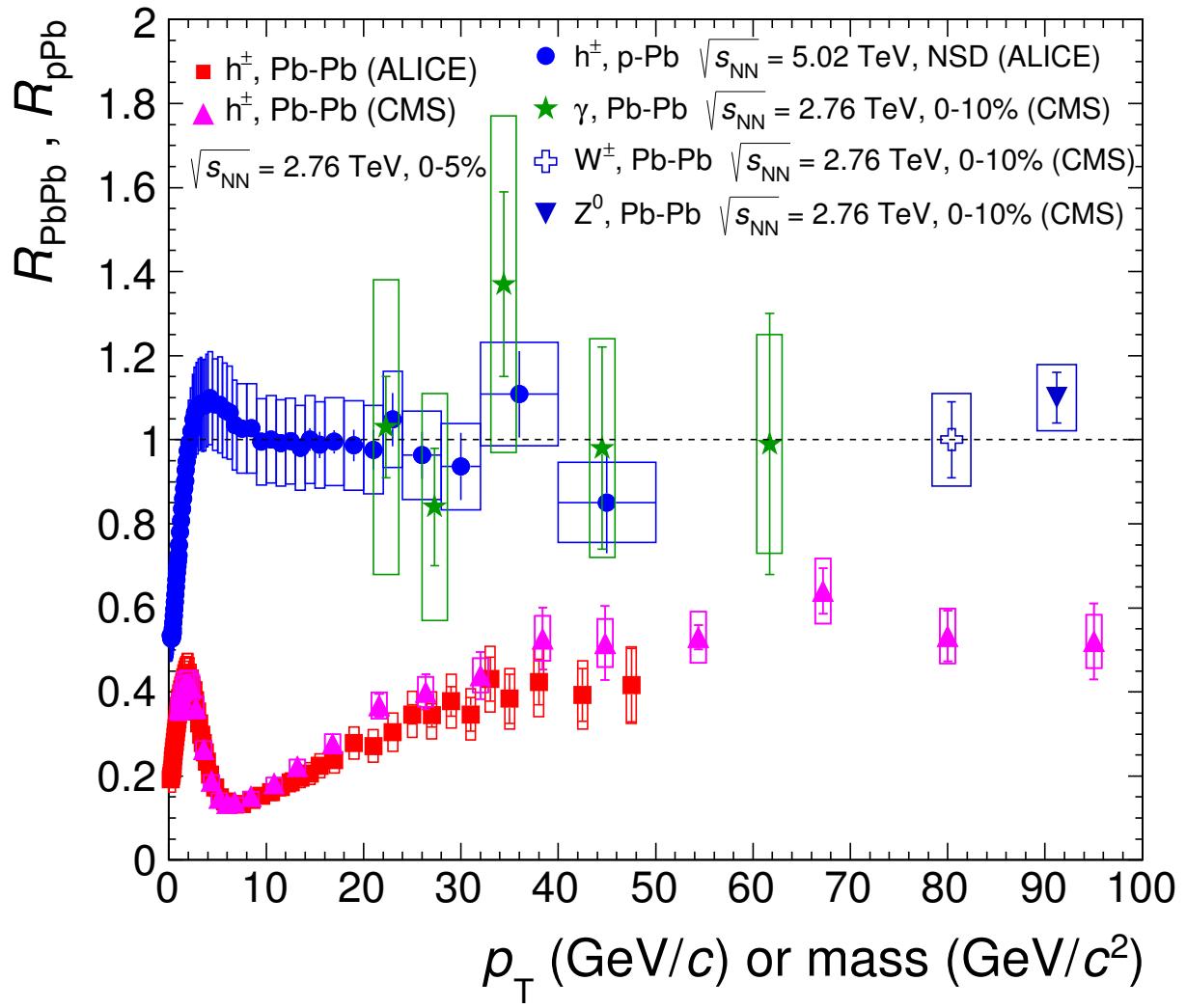
The trend for each centrality class is roughly the same, however, where jets at higher  $p_T$  have modification factors closer to 1. Because of the steepness of the jet spectrum described above, the ensemble of high  $p_T$  jets that comes out of the droplet of QGP will be dominated by those jets that lost relatively little energy. To put it another way, There are fluctuations in how much energy each individual parton will lose in the medium, and selecting jets which look like high  $p_T$  jets in a vacuum may skew measurements towards partons which have lost the least energy in the medium. This is often called the 'survivor bias' [Connors2018].

Fig. ?? shows the  $R_{AA}$  measured for a variety single particles [Bencedi2016] measure by ALICE and CMS.

The single particle  $R_{AA}$  distributions provides several insights. First, for the bosons measured in the figure, the  $R_{AA}$  is consistent with unity. This is a vitally important check, as these particles are not expected to interact with the medium. This will be discussed in more detail in ???. Second, the shape. While an overall suppression is seen, the peak at low  $p_T$  is more likely the result of cold nuclear matter effects, i.e. changes due to the presence of the lead nucleus alone in pPb collisions, instead of hot nuclear matter (QGP). Third, the shape at high  $p_T$ . Note that measuring  $R_{AA}$  for high- $p_T$  hadrons is quite different: In both pp and AA collisions, a high- $p_T$  hadron is statistically likely to come from a specific, unusual type of jet that contains one very hard parton and is very narrow; Wider jets are intuitively predicted to loose less energy in the medium, and some evidence for this has been observed in [Khachatryan2016]. Selecting (i.e., triggering on) hadrons therefore constitutes selecting an unusual sample of jets that lose less energy, and this selection effect becomes stronger at higher  $p_T$ .

## 1.4 Fragmentation Functions

One of the simplest ways to study QCD is measure the hadronic production in  $e^+ + e^-$  collisions, particularly through the process  $e^+ + e^- \rightarrow q\bar{q}$ . The inclusive cross section for hadron production ( $\sigma$ ) may be written as the product of the partonic cross section ( $\hat{\sigma}$ ) and a parametrization of the non-calculable long-range behavior called the fragmentation function (FF), denoted  $D_c^h(z)$ , which is defined as the probability for a parton of flavor  $c$  to fragment into a hadron taking a fraction  $z = p_h/p_c$  of its momentum:



ALI-DER-95222

Figure 1.11: Comparisons of  $R_{\text{AA}}$  and  $R_{\text{pPb}}$  for various single particles measured by ALICE and CMS [Benedicti2016].

$$d\sigma = \sum_c \int dz d\hat{\sigma}(p_a, p_b, p_c) D_c^h(z) \quad (1.8)$$

This property is known as factorization, and is thought to hold for a wide array of observables, though it has only been explicitly proven for small subset of observables. Experimentally, the yield of hadrons associated with a parton (or jet) of known momentum as function of  $z$ , is a measure of the fragmentation function. Since generally  $p_T$ , not  $p$  is measured experimentally, an alternative variable,  $z_T = p_T^h/p_T^{trig}$  is often used, where  $p_T^{trig}$  is the transverse moment of a jet, or other object related to a hard scattered parton.

A more complex but relevant observable is the semi-inclusive cross section in Deep Inelastic Scattering (DIS, or SIDIS). This usually refers to an electron colliding with a proton. Semi-inclusive scattering simply indicates that not all the particles are measured. In e+p collisions, semi-inclusive scattering measures at least one other hadron in coincidence with the scattered electron. *Exclusive* scattering, indicates that *all* particles produced in the collision are measured. These definitions appear counter-intuitive, at least in terms of what's measured in the collision. But the distinction comes from the underlying physical processes that each measurement corresponds to. In semi-inclusive scattering, there are several physical processes that could have resulted in the limited number of particles that are measured. For example, there are a variety of processes that give rise to a  $q\bar{q}$  pair, all of which must be considered in an event where exactly two jets are measured in coincidence with the scattered electron. In exclusive scattering, because all particles are measured in the collision, the underlying physical process producing those particles is much more readily identified, to the exclusion of other potential processes. Unlike e+p collisions, in heavy ion collisions the shear number of particles produced, a substantial fraction of which are neutral particles that are notoriously difficult to measure, exclusive measurements are essentially impossible. The term Deep Inelastic is more straightforward; rather than an elastic collision where momentum is conserved, much of the energy goes towards breaking up the proton(Ion). The semi-inclusive DIS cross section can be written as:

$$d\hat{\sigma} = \sum_{a,c} \int dx_a dz f_a(x_a) d\hat{\sigma}(p_a, p_b, p_c) D_c^h(z). \quad (1.9)$$

where (Bjorken)  $x$  is the fraction of the proton's momentum carried by the parton, and  $f_x(x)$  is the parton distribution function (PDF) that describes the partons in their initial state before the collision. At first glance, this equation is troubling, as it appears that a careful measurement of the hadronic cross section cannot uniquely determine the PDF or fragmentation function. The long-range behavior of these fragmentation functions, however, is thought to be independent of the collision process, a property known as *universality*. Thus, the same fragmentation functions are thought to apply regardless of the particle species being collided. On the other hand, parton distribution function is only a property of the objects being collided and can be factorized from the collision process and subsequent fragmentation.

The cross section for hadro-production in proton-proton collisions can be expressed in a similar way with the addition of a second integral over the additional parton:

$$d\hat{\sigma} = \sum_{a,c} \int dx_a dx_b dz f_a(x_a) f_b(x_b) d\hat{\sigma}(p_a, p_b, p_c) D_c^h(z). \quad (1.10)$$

This gives rise to an interesting picture of progression, albeit a slightly oversimplified one: fragmentation functions are measured in  $e^+ + e^-$  collisions, shown in Fig. 1.12, which are then used in DIS data. Then, the DIS data is used to determine the PDF's, shown in Fig. 1.13 which are applied to  $p + p$  collisions.

In a similar story of progression,  $p+p$  collisions are important baseline data for collisions of heavy nuclei, discussed in Sec. 1.3.4. It turns out, expectations for hadronic observables must be modified in nuclear collisions. Furthermore,  $p+Pb$  collisions are important for disentangling cold and hot nuclear matter effects. Such departures provide a window into physics beyond the vacuum behavior of QCD accessed via elementary particles collisions.

## 1.5 Cold Nuclear Matter Effects

shadowing,  $k_T$  effect, others....

## 1.6 Two Particle Correlations

Another way to study energy loss effects on partons propagating through a medium in heavy ion collisions is by measuring two particle correlations. One of simplest forms of two particle correlations is the di-hadron correlation. Contemporary jet measurements invoke jet reconstruction algorithms to determine the full energy of the jet event-by-event. These methods are difficult to apply in heavy-ion collisions due to the overwhelming background from soft collisions, and may be less sensitive to medium modification depending on the observable being measured. Instead, a very useful approach has been to measure correlations between particles.

$$Y(\Delta\phi) \equiv \frac{1}{n^{\text{triggers}}} \frac{dN}{d\Delta\phi} \quad (1.11)$$

Fig. ?? shows a simplified example of a di-hadron (hadron-hadron, or h-h) correlation in  $p+p$  collisions. The two-peak structure is characteristic of such measurements, and indicates that the event sample is dominated by di-jet events. The particles within the same jet make up the narrow peak centered around  $\Delta\phi = 0$  and the recoil jet appears as the peak around  $\Delta\phi = \pi$ . The away side peak is broadened since kinematically the away side jet can swing along the  $\eta$  direction and  $k_T$ , the initial pair momentum of the colliding partons, can create an imbalance in the jets' energy and cause them to be acoplanar.  $\eta$ , also referred to as pseudorapidity, is defined as  $\eta = \ln(\tan(\theta/2))$ , where  $\theta$  is the polar angle with respect to

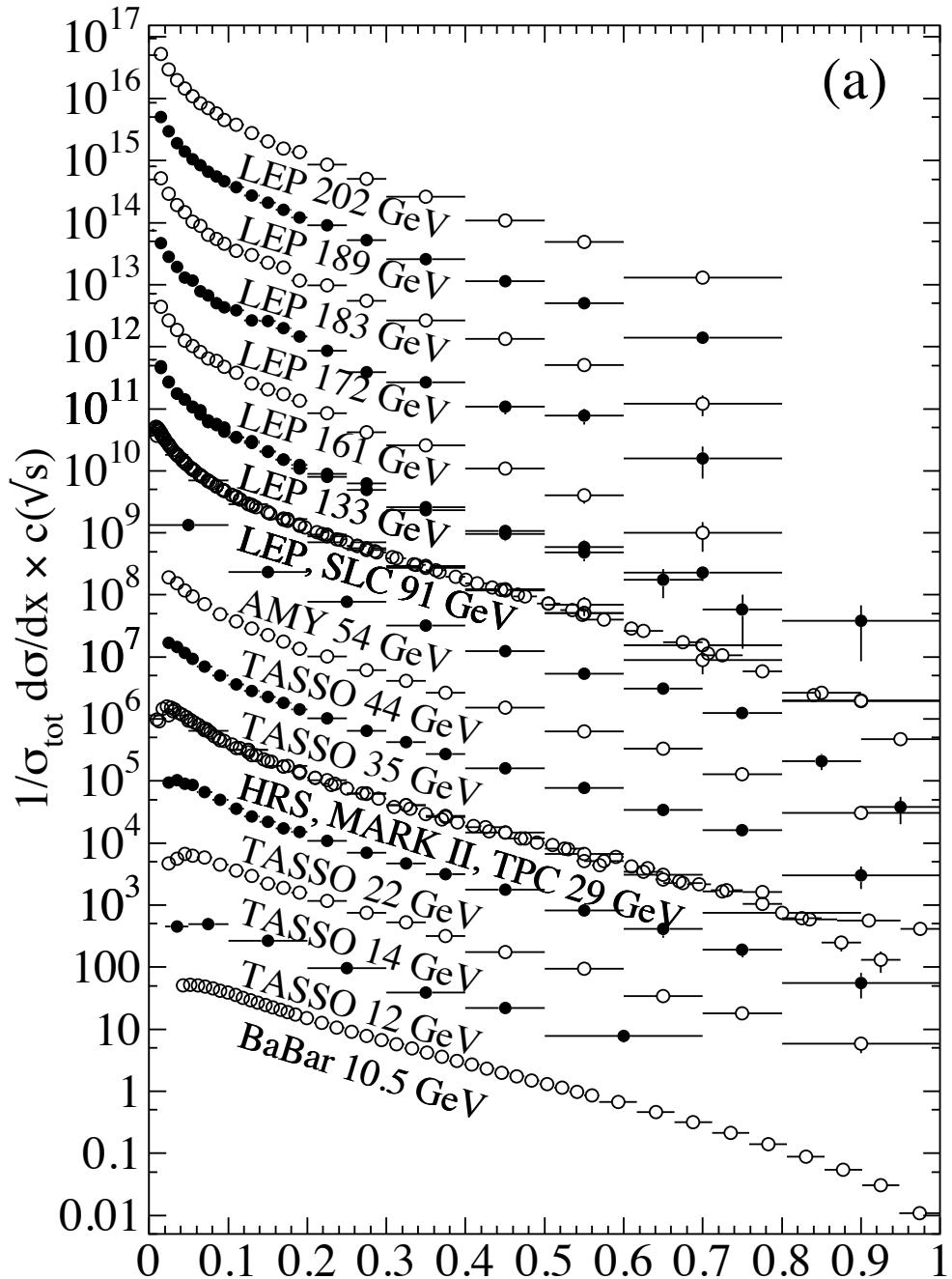


Figure 1.12: The  $e^+e^-$  fragmentation function for all charged particles for different CM energies  $\sqrt{s}$  versus  $x$ . For the purpose of plotting, the distributions were scaled by  $c(\sqrt{s}) = 10^i$ , where  $i$  ranges from  $i=0$  ( $\sqrt{s} = 12$  GeV) to  $i=13$  ( $\sqrt{s} = 202$  GeV) [deFlorian2018]

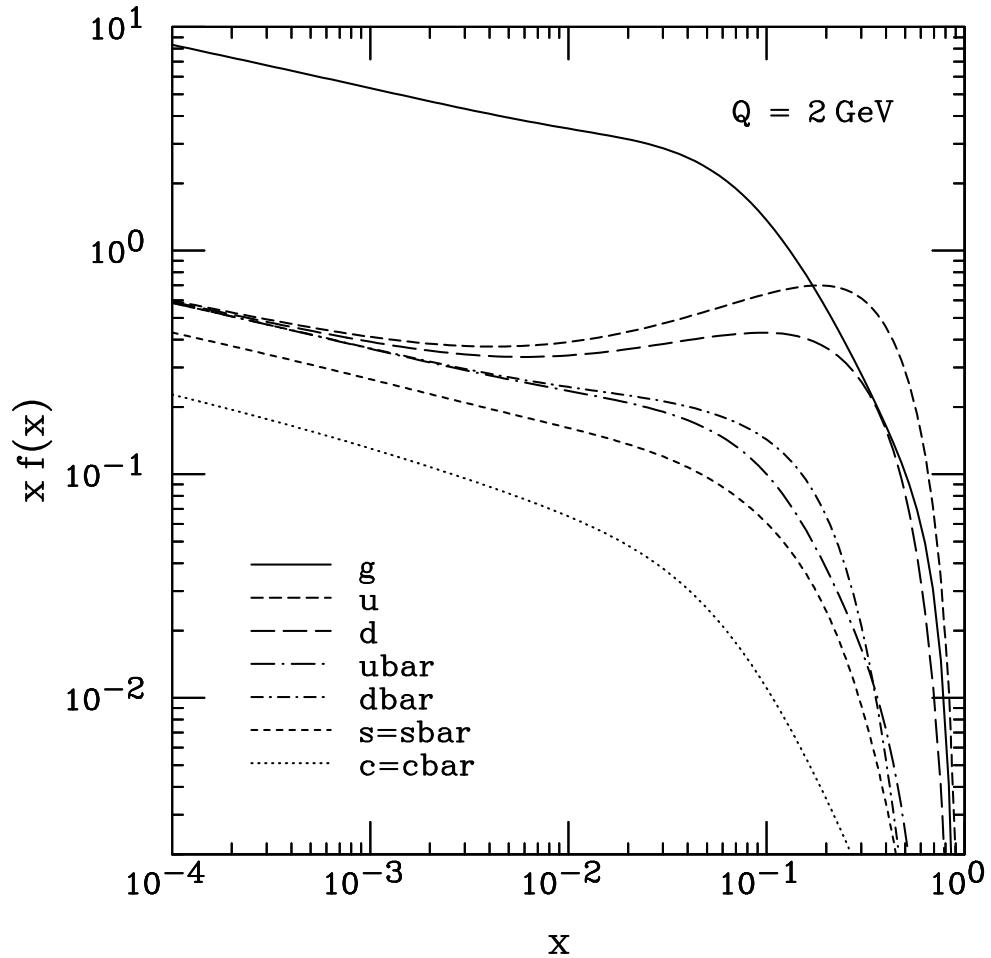


Figure 1.13: Sampling of PDFs from CTEQ [Pumplin2002]

beam direction. A 2D plot ( $\Delta\phi, \Delta\eta$ ) of this correlation in pPb and pp collisions is shown in Fig. 1.8, Sec. 1.3.1. The peaks sit atop a pedestal which is due to initial and final state interactions amongst the beam remnants and the hard-scattered partons. This must be subtracted, and will be discussed in greater detail in Sec. ?? for pp and pPb collisions. For AA collisions, this subtraction is even more complex, where flow is no longer a signal, but a background phenomenon that must be subtracted.

The STAR experiment performed a hadron-hadron correlation measurement with triggers of  $p_{T,t} > 4 \text{ GeV}/c$  and partners of  $2 \text{ GeV}/c < p_{T,a} < p_{T,t}$ . The result, shown in Fig. 1.15 demonstrates that for central Au + Au collisions the near-side jet looks very similar to p+p but the away-side jet completely disappears. This is consistent with a picture in which the near-side jet is usually produced near the surface and the away-side jet is completely absorbed by the medium.

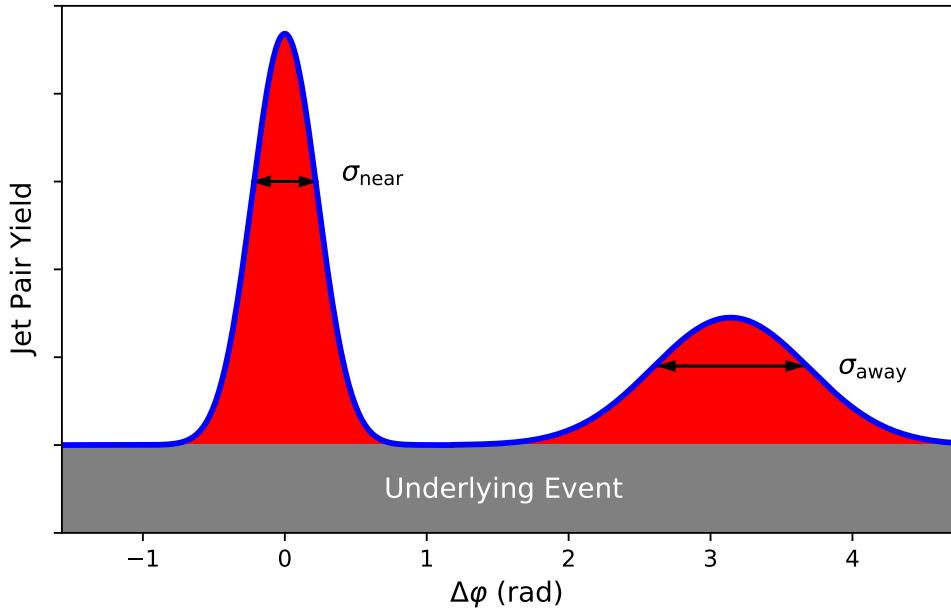


Figure 1.14: Cartoon illustrating a measurement of two-particle correlations from jets

Although hadron-hadron correlations have revealed a great deal about energy loss in the medium, they are limited by the fact that the initial parton momentum is unknown and cannot be used to directly measure the fragmentation function.

## 1.7 Prompt Photons

Prompt photons can be defined simply as the photons produced immediately in the collision, before final state hadrons are produced. At the lowest order in pQCD, prompt photons are produced via two processes: (i) quark-gluon Compton scattering,  $qg \rightarrow q\gamma$ , (ii) quark-antiquark annihilation,  $q\bar{q} \rightarrow g\gamma$ , and, with a much smaller contribution,  $q\bar{q} \rightarrow \gamma\gamma$ . In p+p collisions, the Compton-type process dominates the cross section by roughly an order of magnitude over annihilation as a result of the scarcity of antiquarks. Additionally, prompt photons can be produced in higher-order processes, such as fragmentation or bremsstrahlung [Aurenche1993]. The collinear part of such processes has been shown to contribute effectively also at lowest order. The basic Feynman diagrams for these processes (excluding  $q\bar{q} \rightarrow \gamma\gamma$ ) are shown in Fig. 1.16.

Photons produced during fragmentation, are aptly named fragmentation photons. As a result, fragmentation photons are often produced surrounded by a larger amount of energy and hadronic activity than other prompt photons produced from the initial hard scattering.

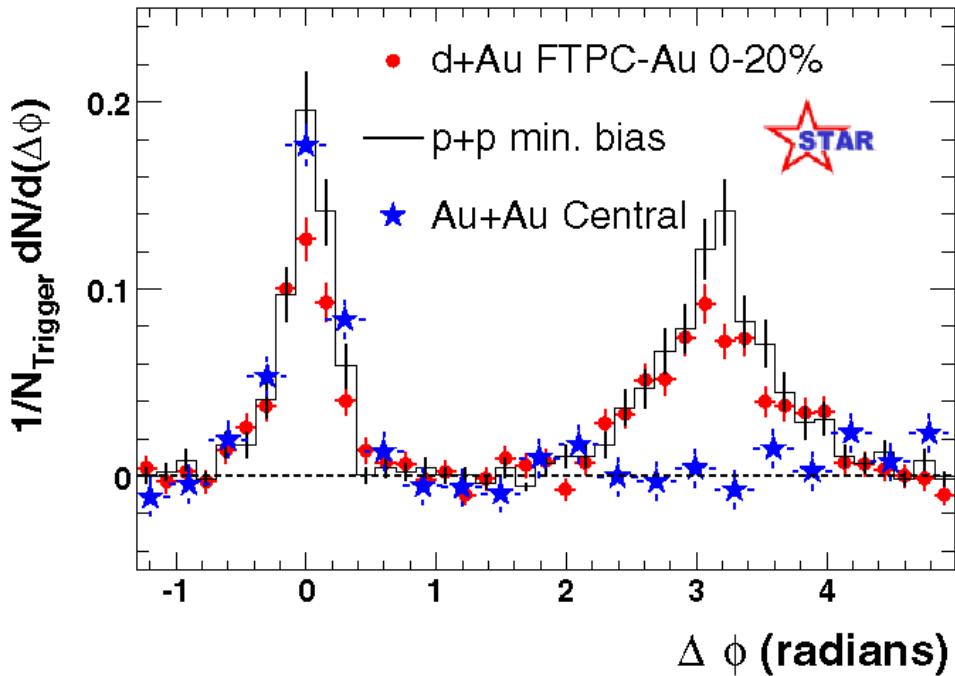


Figure 1.15: Hadron-hadron correlations measured in p+p, d+Au and Au+Au collisions at STAR. The near side jet peaks around  $\Delta\phi=0$  in all three systems but the away-side which peaks around  $\Delta\phi = \pi$  in p+p and d+Au, is suppressed in Au+Au[Adams2005]

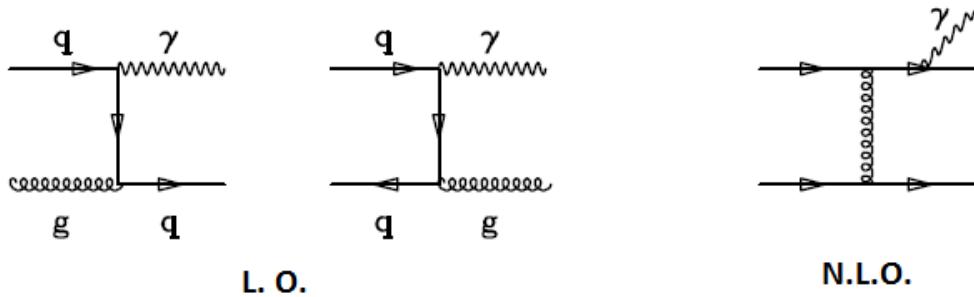


Figure 1.16: On the left are the leading order Feynman diagrams for direct photon production. The diagram on the right is next-to-leading order. Photons resulting from this diagram are referred to as fragmentation photons

Prompt photons, and by extension fragmentation photons, are both included in the definition of direct photons. While the definition of direct photons is not always consistent in the literature, and varies slightly between collaborations, we define direct here to mean any photon not produced from hadronic decays. Those photons originating from the decay of a hadronic bound state are defined as decay photons. The two largest sources of decay photons are the two-photon decay channels of the  $\pi^0$  and  $\eta$  meson. At higher  $p_T$ , decay photons make up the majority of photons produced in the collision.

Together, direct and decay photons make up all the photons observed in a collision, called inclusive photons.

$$\gamma_{\text{inclusive}} = \gamma_{\text{direct}} + \gamma_{\text{decay}}$$

## 1.8 Photons in Heavy Ion Collisions

Direct photons have three very important properties that make them valuable tools in heavy-ion physics. First, There are relatively few leading order diagrams which contribute to direct photon production. Second, The photon-quark coupling is point-like, and not effected by long-range QCD behavior such as fragmentation in the final state. Third, though a property shared with other high- $p_T$  photons, they do not interact with the QGP.

Prompt photons are extremely valuable tools in heavy-ion physics. One property that makes them so useful is that they are not expected to interact with the QGP.

Photons do not carry color charge, and should therefore be unmodified by strong interactions in a medium. While the plasma is made up of quarks that carry (fractional) charge, For example, the mean free path of a 1 GeV photon in a QGP at  $T = 200$  MeV was calculated to be  $\lambda = 480$  fm, much larger than the estimate size of plasma at  $r \approx 10$  fm [David2020].

Thus, in the leading order picture, after prompt photons are produced early in the collision they should propagate through the medium completely unmodified, with no high  $p_T$  suppression. This has been verified, by measuring the  $R_{AA}$  of photons, shown in Fig. 1.17.

A measured  $R_{AA}$  consistent with 1.0 strongly supports the position that photons are unmodified in the quark gluon plasma.

### 1.8.1 $\gamma$ -Jet Correlations

Another important property of leading order prompt photons is that the photon-quark coupling is point-like and therefore not complicated by long-range QCD behavior such as jet fragmentation in the final state, in distinction to hadronic observables. Thus, in the leading order picture, the direct photon exactly balances the away-side parton and resulting jet. Here the photon acts as a reference to the parton from the initial scattering, before any modification in heavy-ion collisions. This means that comparisons between the photon and parton, or jet, as well as large deviations from  $\Delta\phi \approx \pi$  can be used to directly study

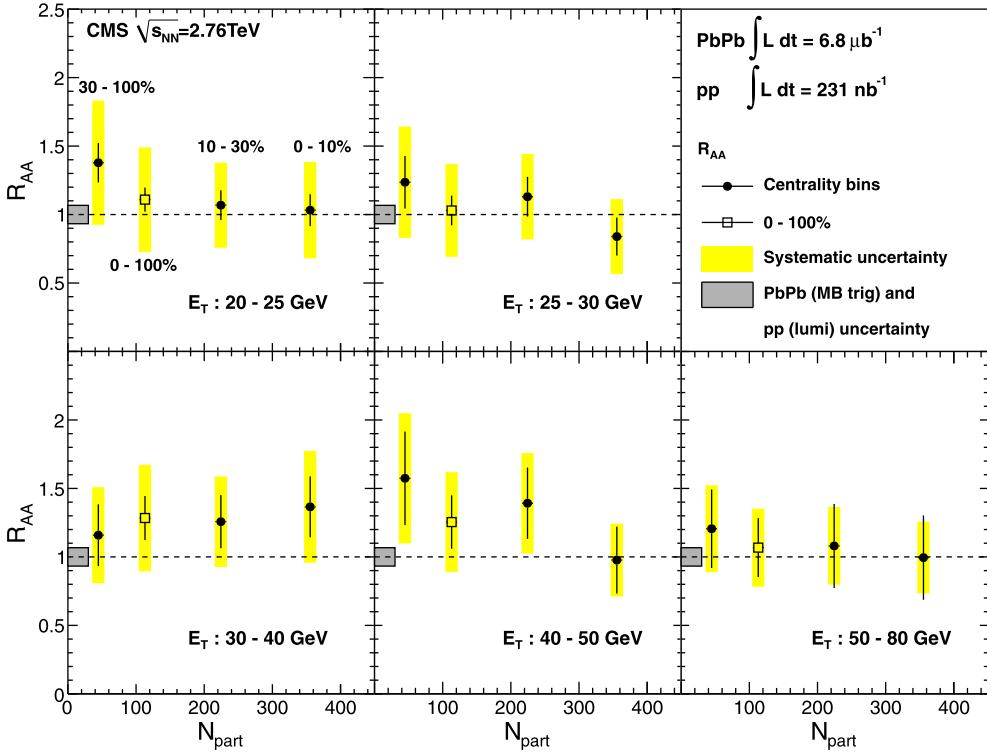


Figure 1.17: The measured nuclear modification factor  $R_{AA}$  as a function of PbPb centrality (given by the number of participating nucleons,  $N_{\text{part}}$ ) for five different photon transverse energy ( $E_T$ ) intervals.

medium-induced effects on the recoiling parton.  $\gamma$ -jet correlations in CMS are shown in Fig. 1.18 [Sirunyan2018].

In conjunction with  $\gamma$ -jet correlations, the  $\gamma$ -jet asymmetry,  $x_{j,\gamma} \equiv p_{\text{T}}^{\text{jet}}/p_{\text{T}}^{\gamma}$  can be measured to quantify in-medium parton energy loss. Derived from Fig. 1.18, the jet asymmetry for jets with  $\Delta\phi > 7\pi/8$  relative to the photon were taken. The asymmetry as a function of photon  $p_{\text{T}}$ , as well as the ratio of yields for large  $\Delta\phi$ ,  $R_{AA}$ , is shown in Fig. 1.19 [Sirunyan2018].

This procedure from CMS also provides a good example of how these correlations are often used to extract other measurements: The angular correlations are measured, from which a region in large  $\Delta\phi$  is taken (corresponding to the photon and parton being back-to-back). The yields in this region are then reported as a function of fractional momentum,  $x_{j,\gamma}$  in order to measure medium effects.

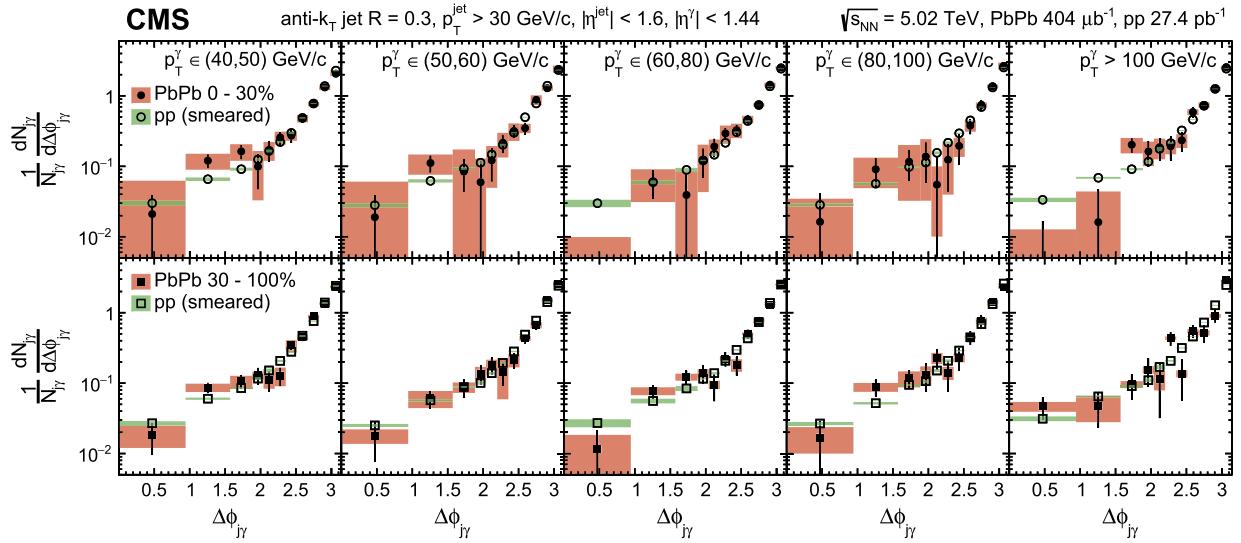


Figure 1.18: The azimuthal correlation of photons and jets in five  $p_T^\gamma$  intervals for 0–30% centrality (top, full circles) and 30–100% centrality (bottom, full squares) PbPb collisions. The smeared pp data (open symbols) are included for comparison. The vertical lines (bands) through the points represent statistical (systematic) uncertainties [Sirunyan2018].

### 1.8.2 $\gamma$ -hadron Correlations

The leading order picture of prompt photons indicates that the photon and recoiling parton have equal and opposite momenta. Therefore, the measurable quantity,  $z_T = p_{T,a}/p_{T,t}$  with the prompt photon as the trigger, is nothing but the fragmentation variable,  $p_{T,hadron}/p_{T,parton}$ . This explains why prompt  $\gamma$ -h correlations are such a powerful measurement. They provide a source of recoil partons of fixed momentum and their conditional yields as a function of  $z_T$  in p+p collisions probe the parton fragmentation. By contrast, dihadron correlations are controlled by the jet cross section which also depend on the PDF’s and the parton scattering cross sections.

When the hadrons roughly opposite the trigger photon are reconstructed as a jet, they are clearly connected to the recoiling parton from the initial scattering in the leading order picture. The hadrons within those jets can then be used to probe the jet fragmentation function.  $\gamma$ -hadron correlations in which a jet is not reconstructed, however, have a distinct advantage over  $\gamma$ -jet correlations: hadrons are more sensitive to in-medium modification. Fig. 1.11 shows a minimum in  $R_{AA}$  for charged hadrons at approximately 6 GeV, and a plateau begins after 20 GeV/c. It is extremely difficult to measure jets below around 20 GeV/c in heavy ion collisions due to the large background [STARCollaboration2017]. Selecting jets at higher  $p_T$  (or jets with kinematics similar to that in ‘vacuum’, or pp collisions), to avoid this background biases the jet population towards jets that will lose the least energy

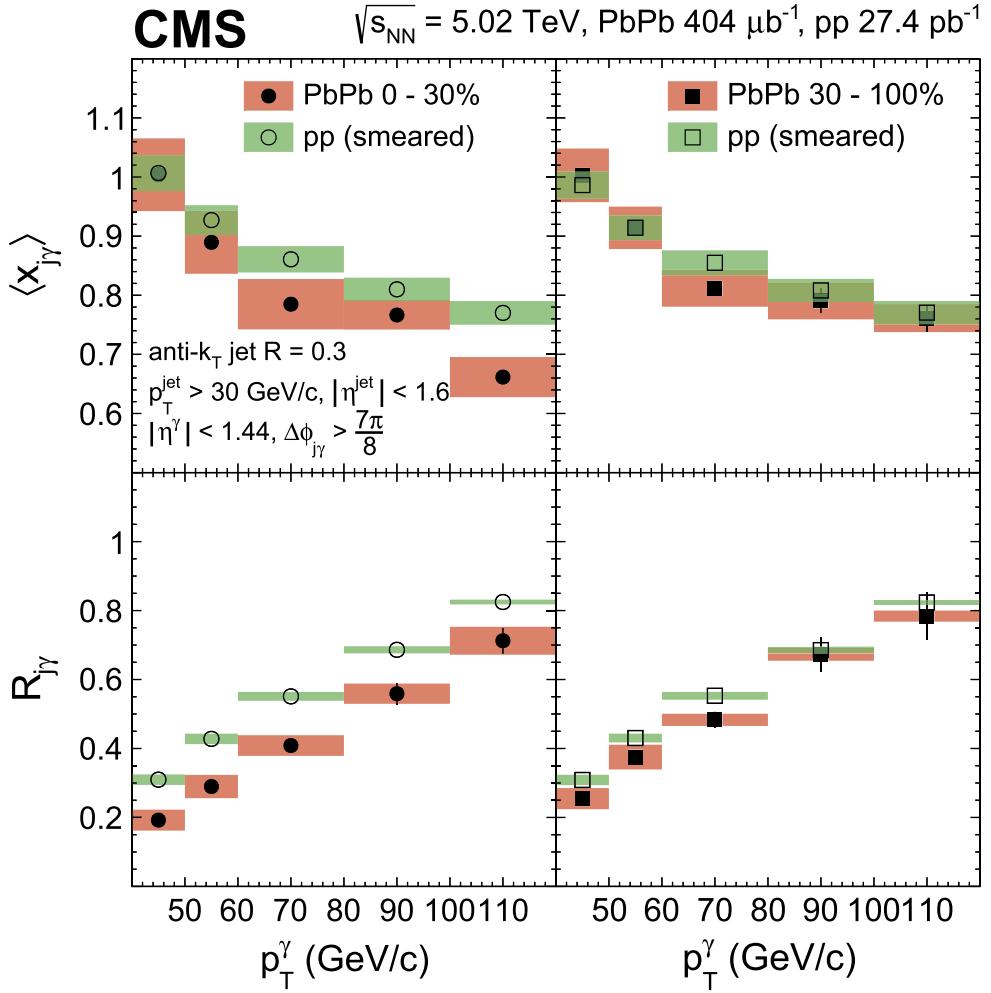


Figure 1.19: The  $\langle x_{j,\gamma} \rangle$  values (top) and  $R_{j\gamma}$ , the number of associated jets per photon (bottom), in 0–30% centrality (left, full circles) and 30–100% centrality (right, full squares) PbPb collisions. The smeared pp data (open symbols) are added for comparison. The vertical lines (bands) through the points represent statistical (systematic) uncertainties.

in vacuum, as discussed in Sec. /refsec:raa. Additionally, triggering on a high  $p_T$  hadron as a proxy for a jet can bias the measurement towards jets towards the surface of the medium [Zhang2007]. Similar to  $R_{AA}$ , the ratio of the conditional yields for  $\gamma$ -h and  $\gamma$ -jet correlations in pp and AA collisions can help quantify medium induced modifications to the parton, in this case, the fragmentation function. Fig. 1.20 shows the direct<sup>1</sup> photon-hadron correlations measured with the PHENIX detector as a function of  $\xi \equiv \ln(1/z_T)$ :

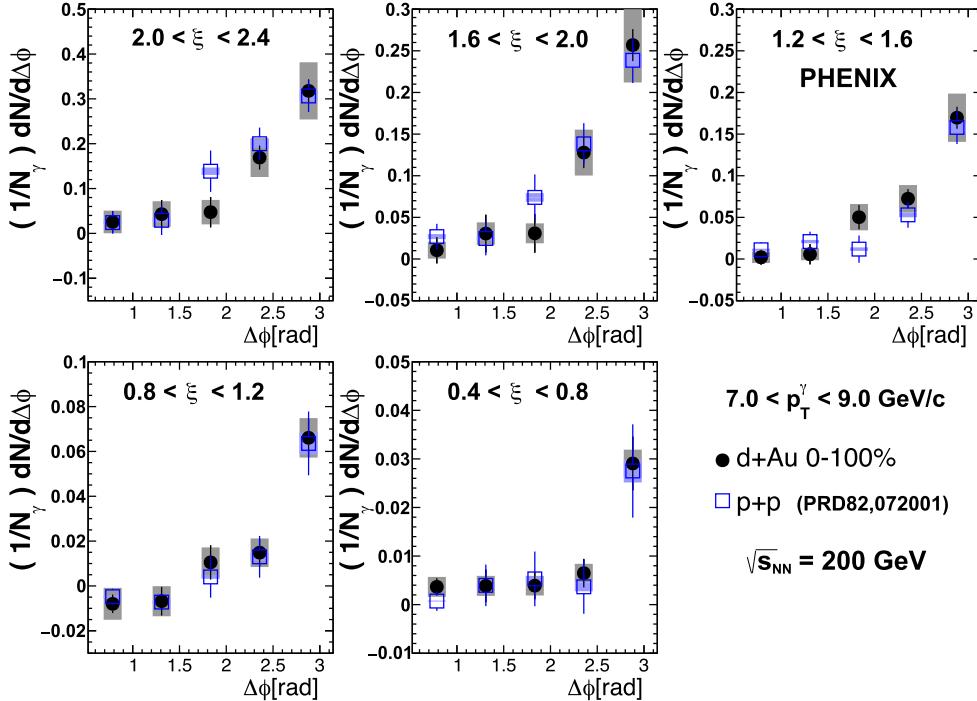


Figure 1.20: Per-trigger yield of hadrons associated with direct photons in Au+Au collisions (closed [black] circles) for direct photon  $p_T$  5–9GeV/ $c$ , compared with p+p baseline (open [blue] squares), in various  $\xi$  bins.

Similar to the  $\gamma$ -jet measurement in the previous section, the conditional yield at higher  $\Delta\phi$  is measured. This conditional yield as a function of  $\xi$  is the experimentally measured fragmentation function. The ratio of this yield in pp and AA collisions,  $I_{AA} \equiv Y_{AA}/Y_{pp}$ , is a nuclear modification factor which quantifies the difference between the fragmentation functions in AA and p+p collisions. This is shown in Fig. 1.21 [PHENIXCollaboration2020].

$\gamma$ -hadron correlations are an incredibly powerful tool: they provide an observable that probes the parton fragmentation function with objects in the collision (low to intermediate  $p_T$  hadrons) that are particularly sensitive to medium modifications. Measuring  $\gamma$ -hadron

<sup>1</sup>Direct is used in lieu of prompt at RHIC due to the smaller contribution of fragmentation photons at the lower center of mass energies compared to LHC.

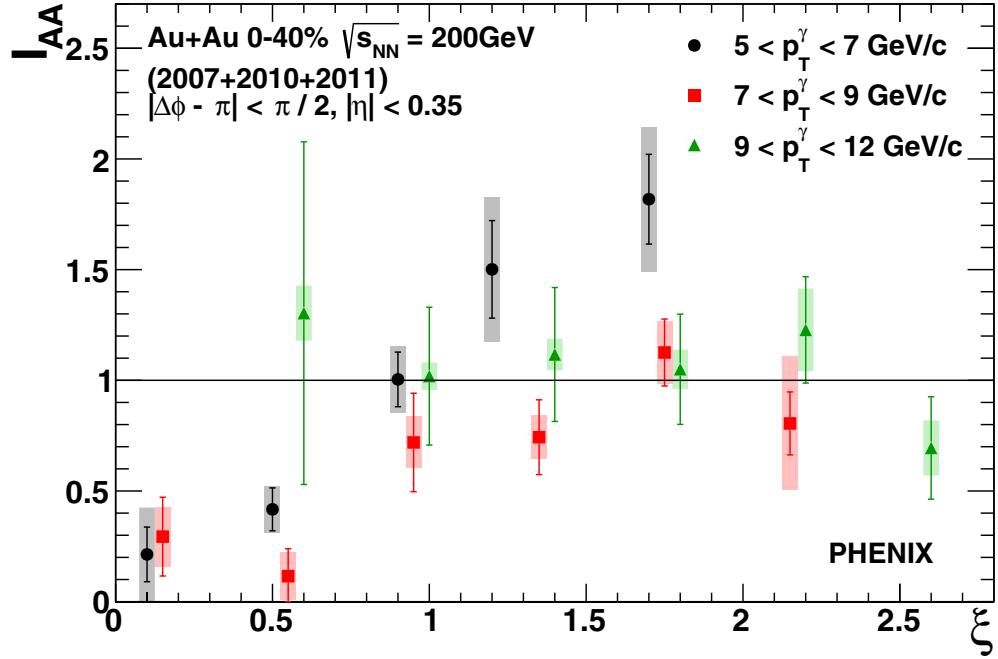


Figure 1.21:  $I_{AA}$  for three direct photon  $p_T^\gamma$  bins.

correlations in both pp and PbPb collisions is an important step in elucidating the modifications to the parton fragmentation function due to QGP. However, a full understanding of these phenomena requires measurements of cold nuclear matter (CNM) effects, which should be present in Au+Au collisions but are difficult to distinguish experimentally from effects due to interactions with the medium.

## 1.9 cold nuclear Matter Effects

In order to quantitatively study the properties of the QGP, it is necessary to separate effects which are due to interactions with the medium from those which are intrinsic to interactions of cold nuclei. The p + p baseline measurements used to calculate the nuclear modification factor RAA can not account for these nuclear effects, since none are present in free protons. For example,  $^{208}\text{Pb}$  contains 126 neutrons, so the majority of nucleon-nucleon collisions in PbPb events will involve neutrons. Any isospin dependant effects would be impossible to model in pp collisions.

### 1.9.1 The Nuclear Parton Distribution Function

The production cross sections for prompt photons and hadrons should be sensitive to the distribution of quarks gluons inside the nucleus, detailed by the nuclear parton distribution function (nPDF). The relation of the bound-proton PDFs with respect to free-proton PDFs  $f_i^p$  is often expressed in terms of a nuclear modification factor in the form of:

$$R_i^A(x, Q^2) = \frac{f_i^p / A(x, Q^2)}{f_i^p(x, Q^2)} \quad (1.12)$$

A typical form of such modifications is shown in Fig. 1.22.

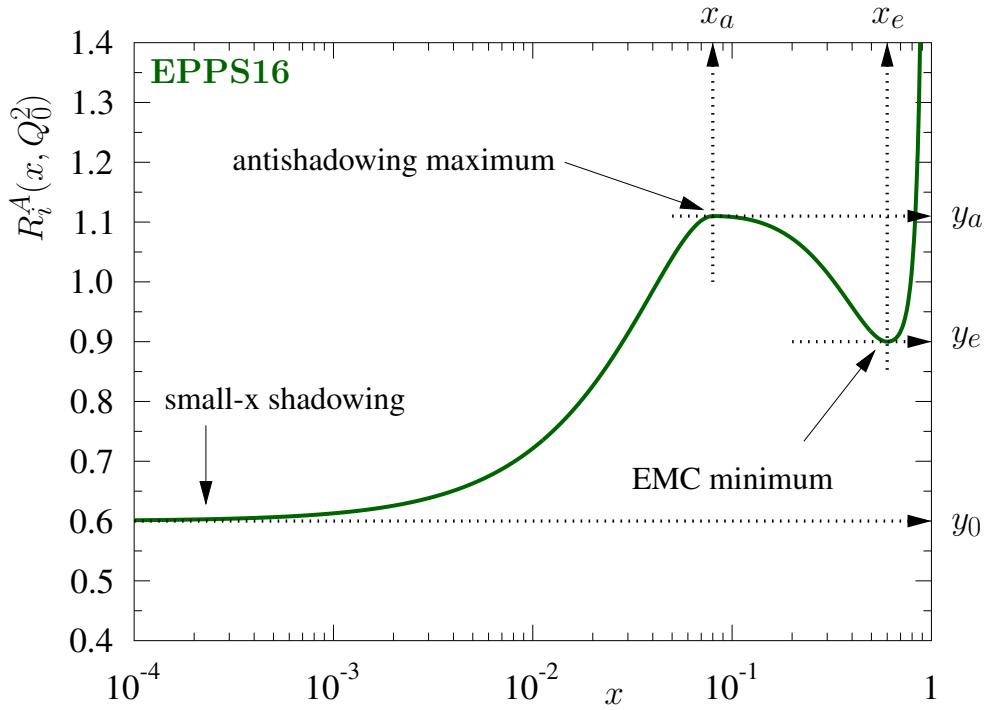


Figure 1.22: Typical form of PDF modifications in a nucleus[Pumplin2001].

The modification in the region where the ratio is less than 1.0 is called "shadowing" at very low  $x$ , and the "EMC" effect at  $0.3 < x < 0.7$ . The region where the modification is larger than 1.0 is called "anti-shadowing". The exact magnitude of these effects on various measurements, as well as an improved understanding of the nPDF's is a currently under investigation in the field. Recently, a major step forward toward this goal has been the inclusion of LHC pPb data, particularly for extracting the gluon-PDF.

### 1.9.2 Nuclei and Fragmentation

Aside from differences between the PDFs in free and bound nucleons, there are potential observable effects of the nucleus as a medium itself. Measurements in p+A collisions showed that particle production at moderate transverse momentum increases faster than the number of binary nucleon-nucleon collisions [Cronin1975]. This was attributed to the Cronin effect where the scattered parton undergoes multiple scatterings in the nucleus, resulting in a transverse momentum boost, *before* the interaction that ultimately produces the final state particle.

Currently, however, an open question in the field is the exact timescale of fragmentation, as both pQCD and lattice calculations are unable to provide estimates. This leaves open the possibility that the parton begins to fragment while still inside the nucleus, potentially modifying the fragmentation process and the final state hadrons of the resulting jet.

The current state on cold nuclear matter effects modifying the parton fragmentation function is ambiguous: In di-hadron and direct photon-hadron correlations, no significant modification of the jet fragmentation was observed in measurements by the PHENIX collaboration in d–Au collisions at a center-of-mass energy of 200 GeV [Adler:2005ad] and the ALICE collaboration in p–Pb collisions at 5.02 TeV [Acharya:2018edi, Adam:2015xea] at mid rapidity. At forward rapidity, probing a lower-x regime, a strong-modification was observed by the PHENIX collaboration in d–Au collisions [Adare:2011sc]. A more recent measurement by the PHENIX collaboration with pp, p–Al, and p–Au data revealed a transverse momentum broadening consistent with a path-length dependent effect [Aidala:2018eqn]. However, a recent ATLAS measurement of the jet fragmentation function in p–Pb collisions showed no evidence for modification of jet fragmentation for jets with  $45 < p_T < 206$  GeV/c [Aaboud:2017tke]. Measurements of the fragmentation of jets with much lower momentum are necessary to limit the Lorentz boost to the timescales of fragmentation, as such a boost may result in fragmentation outside the nucleus. These measurements would test the  $Q^2$  evolution of fragmentation functions in cold nuclear matter, testing factorization theorems that are neither proven nor expected to hold in general for collisions involving nuclei [deFlorian:2011fp].

## 1.10 Statement of Purpose

Parton fragmentation may be modified in the nucleus, offering a way to explore the dynamics of QCD in nuclei including elastic, inelastic, and coherent multiple scattering of partons. Moreover, the known spatial dimensions of nuclei provide a filter possibly shedding light on the timescale of the fragmentation process, which remains unknown [Accardi:2009qv, Accardi:2012qut]. Additionally, because photons produced in hard scatterings do not strongly interact, they constrain the parton kinematics from the same scattering before any modification. Thus, measurements of photon-tagged jet fragmentation in pA collisions serve as a powerful tool to study multiple-scattering effects in cold nuclear matter [Xing:2012ii],

which serve as a control for effects of the quark–gluon plasma (QGP) in nucleus–nucleus collisions, where modifications of the jet spectrum, fragmentation, and substructure have been observed [**Connors:2017ptx**].

In this work, azimuthal correlations of charged hadrons with isolated photons,  $\gamma^{\text{iso}}$ , are analyzed in p–Pb and pp collisions with a center-of-mass energy of  $\sqrt{s_{\text{NN}}} = 5.02$  TeV. Isolated photons are measured at midrapidity,  $|\eta| < 0.67$ , and with transverse momenta in the range  $12 < p_{\text{T}} < 40$  GeV/ $c$ , which yields the scaling variable  $x_{\text{T}} = 2p_{\text{T}}/\sqrt{s_{\text{NN}}} = 0.005\text{--}0.016$ . The kinematic range probed in this analysis offers access to a lower  $Q^2$  than other LHC experiments, which is where the largest nuclear effects can be expected, and to a similar  $x_{\text{T}}$  range as RHIC measurements at forward rapidity [**Adare:2011sc**].

Understanding the dynamics of quarks and gluons in nucleons and nuclei is a key goal of modern nuclear physics. Proton–nucleus (pA) collisions at high energies provide information about the parton structure of nuclei, parton–nucleus interactions, and parton fragmentation in a nuclear medium [**Accardi:2009qv**]. The energy of the Large Hadron Collider (LHC) available for pA collisions is a factor of 25 larger than at the Relativistic Heavy Ion Collider (RHIC), and thus it provides unprecedented reach in longitudinal momentum fraction Bjorken- $x$  and  $Q^2$  [**Salgado:2011wc**].

# Chapter 2

## Experimental Apparatus

°A Large Ion Collider Experiment (ALICE) is the only experiment at the Large Hadron Collider (LHC) dedicated to studying heavy-ion physics. Its detectors measure and identify hadrons, electrons, photons, and muons. The ALICE detectors are optimized for the study of heavy-ion collisions up to the highest energy available. and is designed to be simultaneously capable of measuring bulk properties of the collision involving soft hadronic interactions and large cross section physics, and capable measuring rare probes involving small cross section physics. In particular, ALICE was designed to track and identify particles from very low, 100 MeV, up to fairly high, 100 GeV, transverse momenta in an environment of extreme particle density. This chapter will first briefly discuss the LHC particle accelerator, followed by a description of the subdetectors and triggering system of ALICE.

### 2.1 The Large Hadron Collider

The LHC is the largest particle accelerator ever made. Located near at CERN near Geneva, Switzerland it was optimized to collide protons up to a center-of-mass energy of  $\sqrt{s}=14$  TeV, and heavy ions (including Pb, Ar, and Xe) up to  $\sqrt{s_{NN}}=5.5$  TeV [Evans2008].

The LHC was designed to supply a luminosity of  $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$  for proton-proton collisions, and a luminosity of  $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$ . The proton-proton luminosity was driven by the search for the Higgs boson, with CMS and ATLAS considered as the high luminosity experiments at the LHC. ALICE, on the other hand, is the only dedicated heavy-ion experiment at CERN. The LHC was constructed using the existing tunnel of the Large Electron-Positron Collider (LEP), operational until 2000. The LHC consists of 4 major components:

1. Dipole magnets that bend the beam on its orbit with a maximum magnetic field of 8.33 T.
2. Quadrupole, sextupole, octupole and decapole magnets the focus the beams.
3. Acceleration cavities that increase the beam energy.

#### 4. Two beam pipes with an ultra-high vacuum which contain the two beams.

The magnetic field in the dipoles is provided by superconducting magnets which are filled with liquid helium. The helium is cooled down to 1.9 K in order to reach the super-fluid state. To reduce the number of interactions of the beam with the environment or air, an ultra-high vacuum is kept in the beam pipes reaching a quality of  $10^{-13}$  atm over a total volume of  $150 \text{ m}^3$  [Evans2008].

The LHC has eight possible interactions points, four of them are equipped with large detector systems shown in Figure 3.1. ALICE is the only dedicated heavy-ion experiment, and will be described in the next section. The detectors systems of ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid experiment) were designed as general purpose detectors primarily in pursuit of the Higgs-boson and its properties, as well as precision measurements of the Standard Model particles and searches for physics beyond the Standard Model. Although both detectors have not been optimized for heavy-ion collisions, they contribute extensively to the high  $p_T$  analysis in Pb–Pb collisions profiting from their large pseudorapidity coverage, their excellent high luminosity capabilities, and their high momentum resolution. The LHCb (LHC beauty) experiment dedicated its research program to the search for CP-violation in the B-meson system, as well as precision measurements in the charm and beauty quark sector.

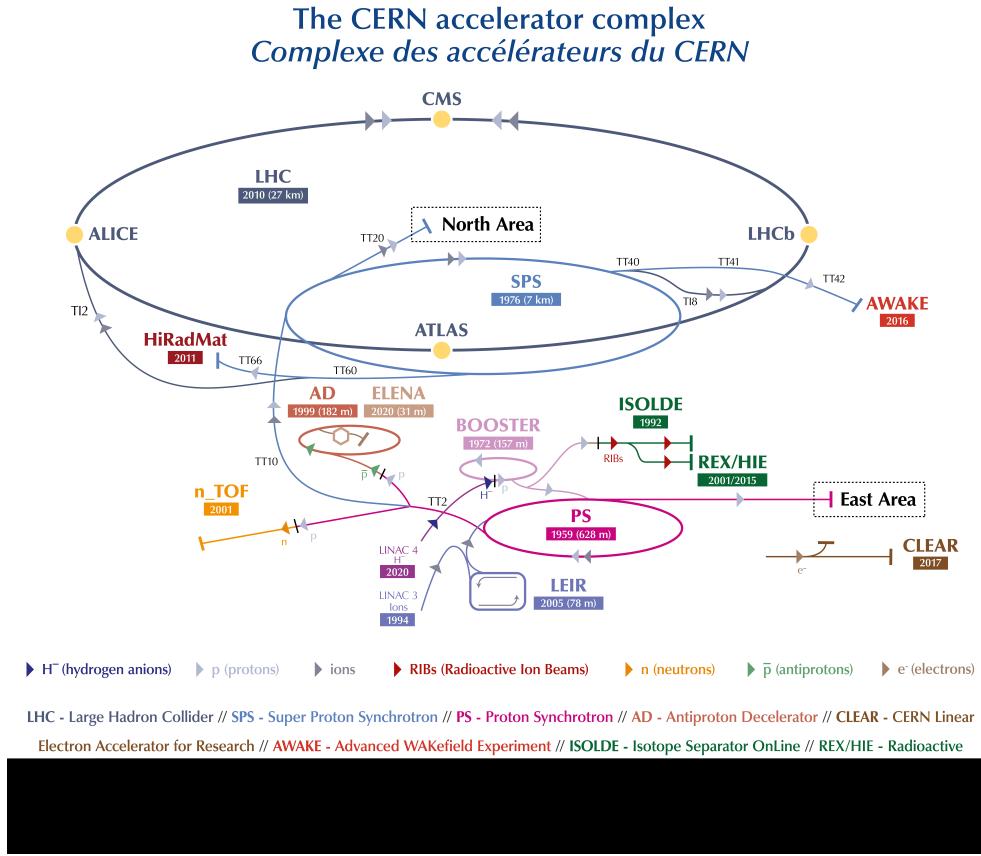


Figure 2.1: Overview of the CERN accelerator complex and the injection chains used for the LHC with their respective top energies for protons and ions after the respective accelerator [Mobs:2197559].

## 2.2 A Large Ion Collider Experiment

ALICE – A Large Ion Collider Experiment – is the only LHC experiment dedicated to study of heavy-ion physics. Therefore, its aim is not to find signatures rare particles in pp collisions as is the primary focus of the other three major LHC experiments, but instead to collect as much data as possible from the large heavy-ion event. This imposes strong demands on the detector. A Pb–Pb collision typically produces approximately 500–1000 charged particles per pseudo-rapidity unit ( $dN_{ch}/d\eta \approx 2000$ ), and can reach as high as  $dN_{ch}/d\eta \approx 2000$  in the most central collisions [PhysRevLett.116.222302]. Fig. ?? shows an overview of the ALICE detector.

As shown in Fig. ??, ALICE can be conceptualized as consisting of three major parts:

1. The central barrel contained inside the magnet with an acceptance in pseudo-rapidity

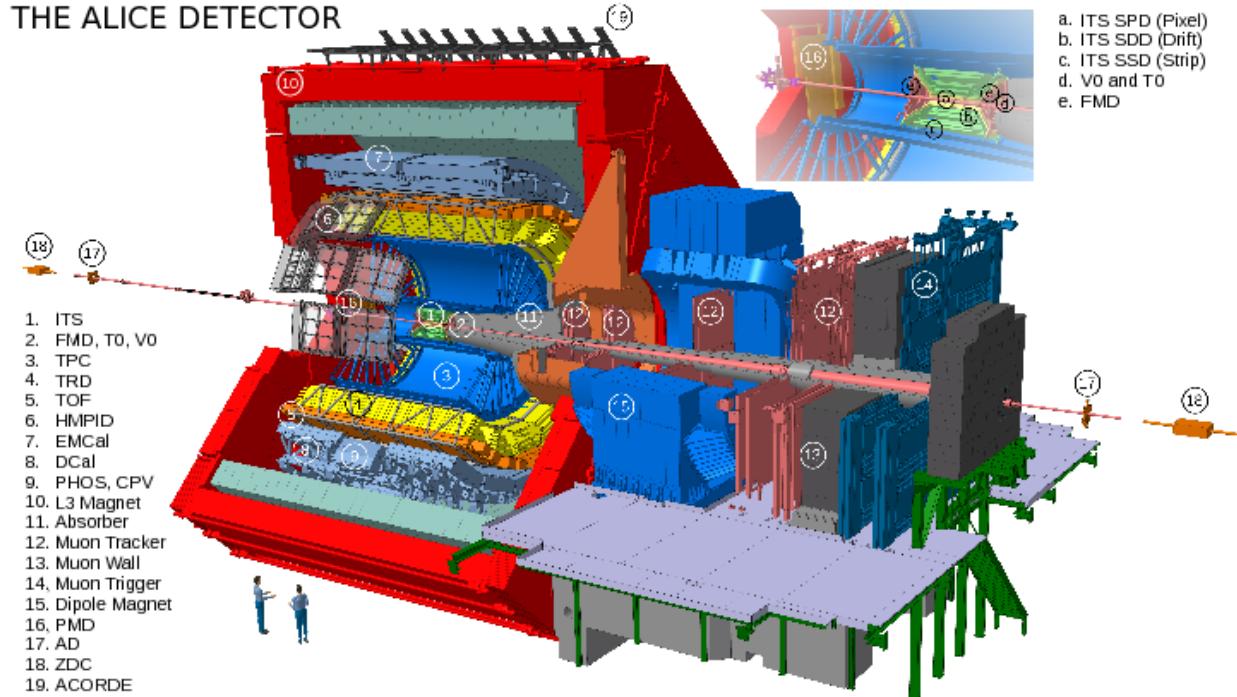


Figure 2.2: Schematics of the ALICE detector [Tauro:2263642].

of  $-0.9 \leq \eta \leq 0.9$  over the full azimuth angle;

2. The muon spectrometer at pseudo-rapidity of  $-4.0 \leq \eta \leq -2.4$ ;
3. Various multiplicity detectors at  $-3.4 \leq \eta \leq 5.1$

The central barrel resides within the L3 magnet, a warm solenoid magnet with a magnetic field of  $B = 0.5$  T. The muon spectrometer extends only in the forward region (C side) and contains a dipole magnet ( $B = 0.67$  T) which bends the muons away from the interaction vertex in the horizontal plane. Lastly, the multiplicity detectors are arrays of scintillator counters, with one in each of the forward regions.

### 2.2.1 Inner Tracking System

Closest to the interaction point, in the middle of the central barrel, is the Inner Tracking System (ITS) that covers the pseudorapidity window  $|\eta| < 0.9$ . The ITS is used for triggering and high-resolution tracking and consists of six concentric layers of silicon detectors surrounding the beam pipe: two layers each of Silicon Pixel Detectors (SPD), Silicon Drift Detectors (SDD), and Silicon Strip Detectors (SSD).

Detector	$r\varphi$ precision ( $\mu\text{m}$ )	$z$ precision ( $\mu\text{m}$ )
SPD	12	100
SDD	35	25
SSD	20	830

A schematic overview is given in Fig. ???. The main purposes of the ITS are to accurately measure the position of the collision vertex and secondary vertices from weakly decaying particles (in particular strange hadrons and B and D mesons, which have typical decay lengths of a couple of cm), measure the  $p_T$  and position of charged particle tracks – often complementing TPC tracking information – and, in the case of the SPD, to provide triggering input for the entire detector. This detector system has a very fine granularity, resulting in a high tracking and vertexing resolution even in extremely high multiplicity environments. The resolution for each subsystem is summarised in Table ???. Moreover, due to its proximity to the beam pipe, a high momentum resolution and large specific energy loss range, particle identification is possible below the lower momentum limit of the TPC (approximately 150 MeV).

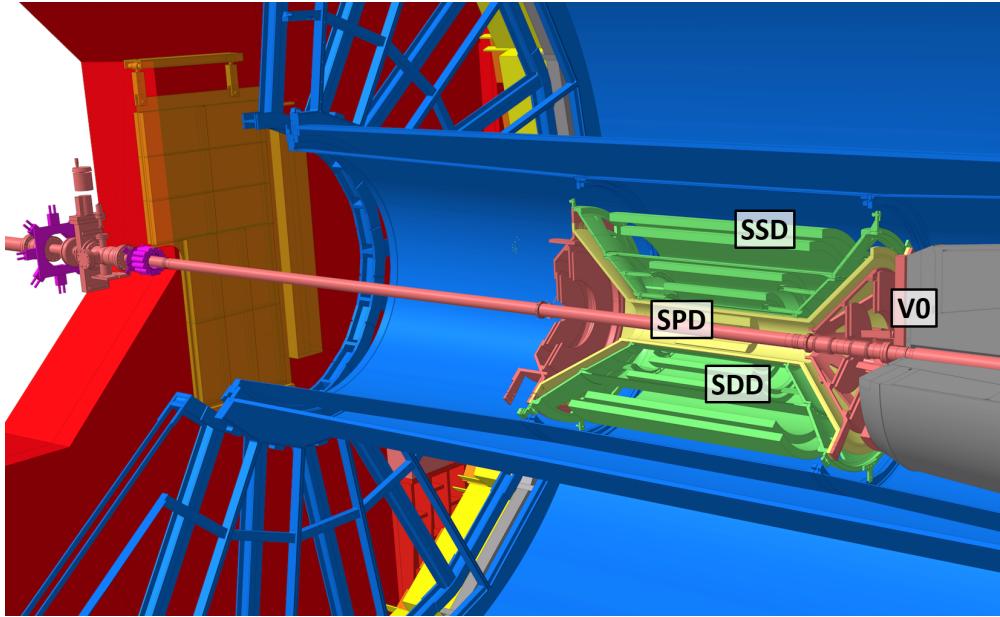


Figure 2.3: Schematics of the Inner Tracking System and nearby detector components, modified from [Tauro:2263642]. The V0 detector is only shown on the C side.

The ITS is a silicon based detector. Silicon is a semiconductor meaning it has a valence band and a conduction band separated by a band gap. The basic operating principle is as follows: a pn junction is used where a reverse bias is applied that depletes the active area

of charge carriers. When a charged particle hits the detector, it will create electron-hole pairs along its trajectory that will in turn generate a current pulse, as shown in Fig. ???. The current will then traverse the electric field and eventually be collected at electrodes connected to a current amplifier. For the purposes of charged particle tracking with the ITS, the only relevant information is whether a particle has hit the detector. This is ensured by triggering if the pulse is above a certain threshold. The advantages of using a semiconductor are that this creates a fast signal, one can achieve a very high tracking resolution, and it has a low material thickness, since the electron diffusion length is only a few  $\mu\text{m}$ .

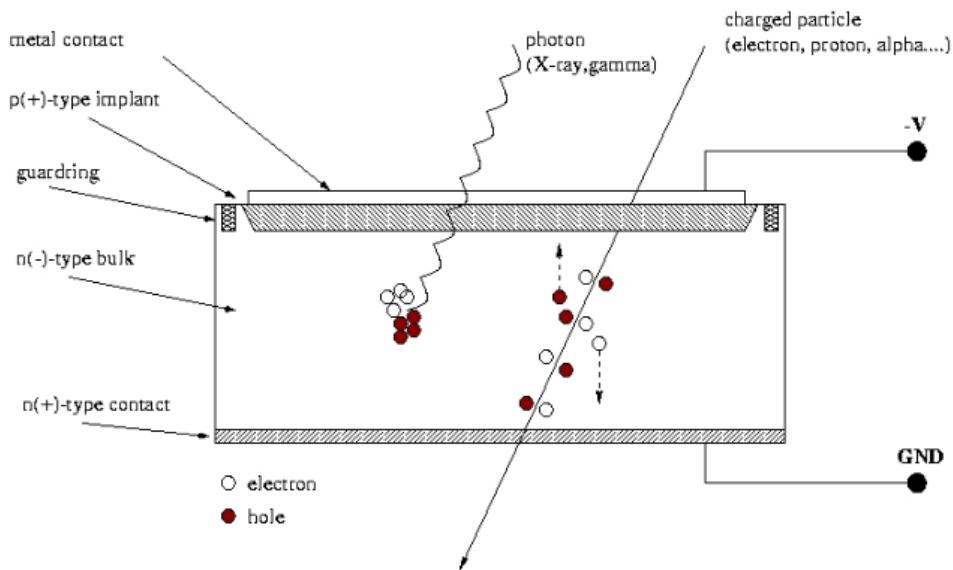


Figure 2.4: Basic principle of a semiconductor detector. When a charged particle traverses the detector, it excites electrons from the valence band to the conduction band, which creates electron-hole pairs. The applied electric field will cause these to drift towards the metal contacts, which generates a current pulse that can be detected. ??,

Each subdetector of the ITS uses a different technology. The SPD is built of pixels of silicon diodes, each measuring  $50 \times 425 \mu\text{m}^2$ . These are read out by 1200 chips per layer, each covering 8192 cells, totalling  $9.8 \times 10^6$  cells per layer. The two SPD layers are located 3.9 and 7.6 cm from the beam axis, respectively, and extend 282 mm in the longitudinal direction.

The SDDs are built of 260 drift cells distributed over the two layers, where the drift time of charge carriers is measured in order to precisely determine where the track has interacted with the cell. Each cell has 256 collection anodes mounted to it along the z axis with a spacing of  $294 \mu\text{m}$ . The drift regions are 35mm long and extend in the  $\varphi$  direction. To determine the  $r\varphi$  coordinate, the drift velocity is monitored by MOS injectors in the substrate that are triggered during gaps in the LHC bunch crossing schedule so they do not interfere with collisions. The position is determined by integrating the velocity over the time measured at

the anode. The effective cell length is  $\leq 202\mu\text{m}$ . These detectors are able to measure the number of collected charge carriers, which is proportional to the energy deposited per unit length,  $dE/dx$ , and therefore can be used for PID. The two SDD layers are located 15.0 and 23.9 cm from the beam axis and have lengths of 443 and 593 mm in the z direction.

Finally, the SSDs each consist of a double layer of silicon strips placed at an angle of 35 mrad ( $\approx 2^\circ$ ) relative to each other. When a particle crossing one of the SSDs yields a signal in both strip layers, the result will be a detection at the crossing point. The number of collected charge carriers is measured, again providing  $dE/dx$  information in the SSD. The strips have a width of  $95\ \mu\text{m}$ . However, due to the arrangement of the strips, the effective resolution is better than what the width alone might predict in  $r\varphi$ , but worse in z. In total,  $1.15 \times 10^6$  strips are used in the inner layer and  $1.46 \times 10^6$  strips in the outer layer. The inner and outer SSD layers are located at 38 and 43 cm from the beam axis, respectively, and extend 86 and 98 cm in the z direction.

### 2.2.2 V0 Detector

The V0 detector is made up of two circular arrays of scintillator counters, with one at each of the forward regions of the ALICE detector. The array on the A side (V0A; opposite to the muon arm) is placed 300 cm from the collision vertex and covers the pseudorapidity region  $2.8 < \eta < 5.1$ . In order to accommodate the muon absorber, however, the V0C array is instead placed only 90 cm from the vertex, covering  $-3.7 < \eta < -1.7$  (this is why V0C is visible in Fig. 4.2, but not V0A). Each array consists of 4 layers with 8 scintillators each, each covering a circular segment spanning  $45^\circ$  in azimuth. The resulting granularity is, while not fantastic, is sufficient for the main purpose of the v0 detector: Triggering and multiplicity measurements.

### 2.2.3 Time Projection Chamber

The Time Projection Chamber (TPC) is the largest tracking detector in ALICE. It is a gaseous detector that relies on the ionisation of a gas for the detection of charged particles. A charged particle travels through the gas, creating a trail of ionisation. The electric field applied through the volume of the TPC causes the electrons (ions) to drift towards the anode (cathode). In the ALICE TPC, thin wires with an applied voltage serve as anodes, resulting in a very strong magnetic field near the anode, which drops off as the distance from the anode increases. Electrons sufficiently accelerated in this field will give rise to secondary ionisations and a subsequent chain reaction known as avalanche multiplication. The detector system can then detect these induced voltage pulses at the cathode readout plane, but the voltage pulse will not stop until the ions with a much slower drift velocity than electrons are collected at the cathode, making the detector relatively slow. This process is illustrated in Fig. ??.

If there is insufficient voltage to cause such an avalanche, instead the energy deposited at the electrodes is proportional to the number of initial ionisations alone, which is pro-

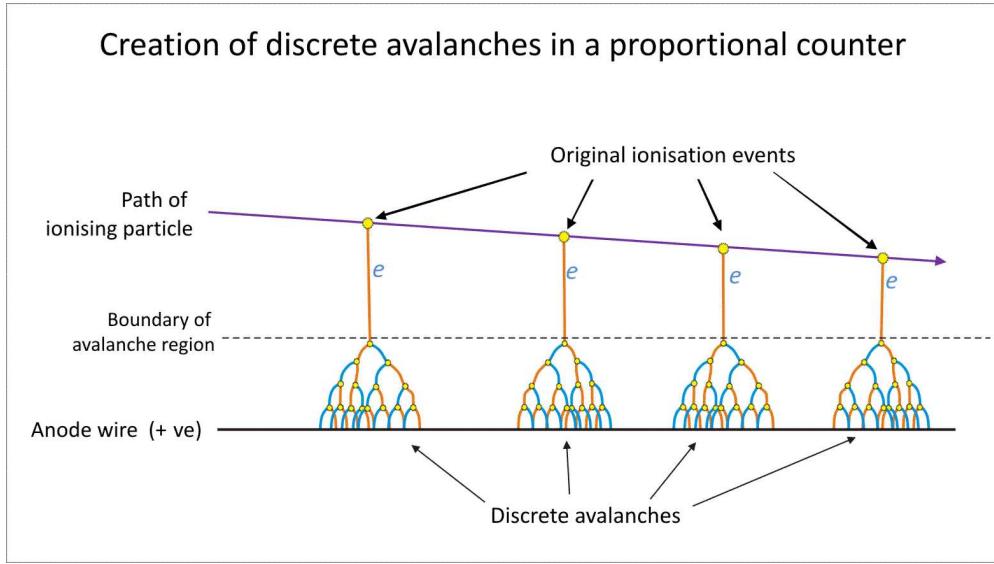


Figure 2.5: Principle of avalanche multiplication in a gas detector [Dousgsm2012].

portional energy deposited per unit length,  $dE/dx$  and provides PID information. For this purpose, the voltage of the anodes are lowered and placed such that they can be in state called "proportional mode".

The ALICE TPC has a cylindrical geometry surrounding the ITS. It has an inner radius of 85 cm, an outer radius of 250 cm, and an overall length of 500 cm. A longitudinal electric field is applied over the cylinder, divided in the center by a cathode plate. A voltage of 100 kV is applied between the center and endcaps, the resulting field strength within the detector volume is 400 V/cm. At the endcaps are Multi-Wire Proportional Chambers (MWPCs), arrays of anode wires operating in proportional mode. When a charged particle from the collision traverses the gas, electron-ion pairs will be created along its trajectory. Ideally, the field is longitudinal with minimal space charge distortions such that the electrons will be projected onto the endcaps, giving precise information about the  $(r, \varphi)$  coordinates of the trajectory. The drift velocity is carefully monitored using a laser system, which also detects inhomogeneities in the electric field. The laser system produces planes of tracks in the detector perpendicular to the electric field that make it easy to measure drift time as well as deviations caused by distortions. With a well determined drift velocity, the z coordinate is obtained by measuring the arrival time of the electrons. For the most frequently used TPC gas mixture – 85.7% Ne, 9.5% CO<sub>2</sub>, and 4.8% N<sub>2</sub> – this results in an average drift velocity of 2.65 cm/ $\mu$ s for the electrons, or an overall drift time of 94  $\mu$ s for the longest drift distances. The endcaps are divided into 18 sectors, with four readout chambers each – two inner readout chambers (IROCs) and two outer readout chambers (OROCs). The wire density is higher for the IROCs as they are closer to the beam pipe, and the voltage applied

is slightly lower. There are no anode wires between the sectors, resulting in gaps in the detector acceptance. The readout chambers have three layers of wires. The innermost wires are thin anode wires spaced 2.5 mm from each other. The middle layer consists of thicker cathode wires, where the ions released during avalanches are collected. The signal is read out from cathode pad planes, with a size in  $(r, r\varphi)$  space ranging from  $4 \times 7.5 \text{ mm}^2$  in the IROCs to  $6 \times 15 \text{ mm}^2$  in the outermost OROC sector. This results in a total number of 159 radial clusters [**Alme2010**].

The operation of the TPC is usually a delicate balance. If the voltage is too high in the normal operational mode, photons released in the de-excitation of gas ions may trigger additional avalanches that skew the initial signal. Another issue has to do with the recombination of ions and electrons at the electrodes, where gas atoms may again enter an excited and prolong the avalanche. To prevent this, a molecular component – usually CO<sub>2</sub> or an organic molecule – is added as a quencher to the gas. Unfortunately, at the high interaction rates delivered in Run 2, unexpected large space-charge distortions were observed at very specific regions of the TPC. They appeared at the boundaries between certain IROCs and deflect the ionization electrons towards these boundaries, leading to a bias in the measured space-point position and effectively decreasing the active readout area. It was suspected that insufficient insulation of some wires in this area could cause strong electric fields leading to amplification and therefore columns of ions drifting back from in between the readout chambers. After the TPC was brought to the surface at the beginning of LS2 indeed tips of anode wires sticking out of the ledges were found on all affected chambers near the expected radii [**Schmidt2020**]. A reduction of the distortions in the hot spots by a factor of about 4 was achieved with the special voltage settings in 2018 compared to 2015. These space charge distortions are shown in Fig. ??.

For the pp and p–Pb data used in this analysis, the TPC was not read out or compromised for several runs. In order to maximize the statistics for the pp and p–Pb, ITS-only tracking was used, a first in ALICE for two-particle correlations. The performance of ITS-only tracking vs. the standard TPC+ITS tracking is discussed in detail in Sec. ??.

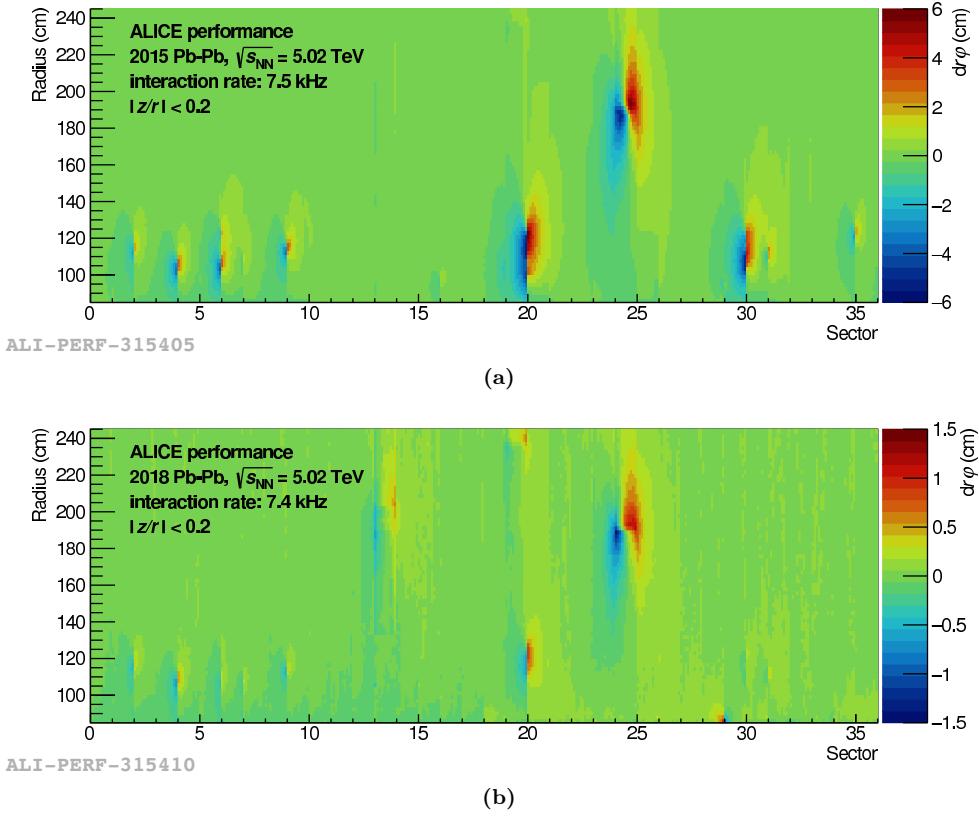


Figure 2.6: Measured space-charge distortions in  $r\varphi$  near the central electrode are shown for Pb–Pb runs with high interaction rates in 2015 (upper plot) and 2018 (lower plot). For both runs the TPC was filled with argon. Note the different z-axis scales. [Hellbar2019]

## 2.2.4 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (EMCal) and Di-Jet Calorimeter (DCal) were mainly designed to measure high  $p_T$  objects. Originally, the calorimeter system in ALICE was proposed as layered lead-scintillator sampling calorimeter attached to wavelength shifting fibers for light collection [Blau2020], covering  $107^\circ$  in azimuth and  $|\eta| < 0.7$  in pseudorapidity. However, in order to enable the study of di-jet-events using full jets (rather than jets reconstructed with charged particle tracking alone), the calorimeter system was extended in 2010 to also include the DCal. The DCal is approximately opposite in  $\varphi$  and uses design as the EMCal. The EMCal and DCal are made up of 12288 and 5377 towers (also called cells), respectively. Each tower has a cross-sectional area approximately twice the Moliere radius of  $\Delta\eta \times \Delta\varphi = 0.0143 \times 0.0143$  with a depth of 24.6cm that corresponds to approximately 20 radiation lengths. The calorimeters are organized into modules that are comprised  $2 \times 2$  cells. These modules are further organized into 10 supermodules that are made up of up to  $12 \times 24$

modules in the EMCAL. Fig. ?? (left) shows the EMCAL made up of these supermodules, while (right) shows the various components that make up a single module.

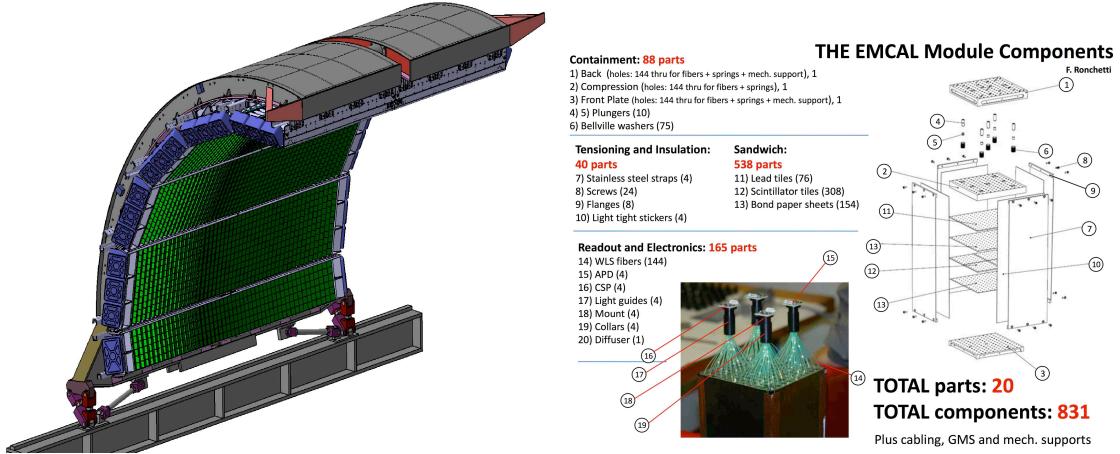


Figure 2.7: Left: The array of Super Modules shown in their installed positions on the support structure. Right: Exploded view drawing of EMCAL module showing all components. Adapted from [Bellwied2010].

The wavelength shifting fibers are wired such that the scintillation light from each cell is read out by a  $5 \times 5 \text{ mm}^2$  avalanche photodiode. The relative energy resolution of the calorimeter is  $\sigma/E = 1.7\% \oplus 11.3\%/\sqrt{E} \oplus 4.8\%/E$  for the energy range of interest for this analysis (roughly 10-50 GeV/c).

## 2.2.5 Time of Flight Detector

The working principle of the TOF detector is to measure the time-of-flight from the interaction point of a particle in order to determine its velocity. Combined with the momentum information, one can extract the mass of the particle. The arrival time measurement is achieved by an array of Multi-gap Resistive-Plate Chambers (MRPCs). MRPCs are thin gaseous detector cells that have a high and uniform voltage applied to them [Aamodt:2008zz]. These are divided into two half-cells, each further divided into five smaller modules blocked by resistive glass plates. The field is sufficiently strong enough such that avalanche multiplication can occur, and thus a particle traversing the cell gives rise to a detectable signal. The glass plates block the avalanches, in order to reduce the time jitter, which scales with the propagation distance. The signals from each gap sum up to the total signal, and using multiple gaps increases the signal strength. The achieved time resolution is about 40 ps.

The TOF detector has full azimuthal coverage and covers the pseudorapidity region of  $|\eta| < 0.9$ . The inner radius of the TOF is 370 cm with an active length of 741 cm. A total of 157,000 cells are arranged into  $122 \times 13 \text{ cm}^2$  strips and put in  $18 \times 5$  modules (in  $(\varphi, z)$ )

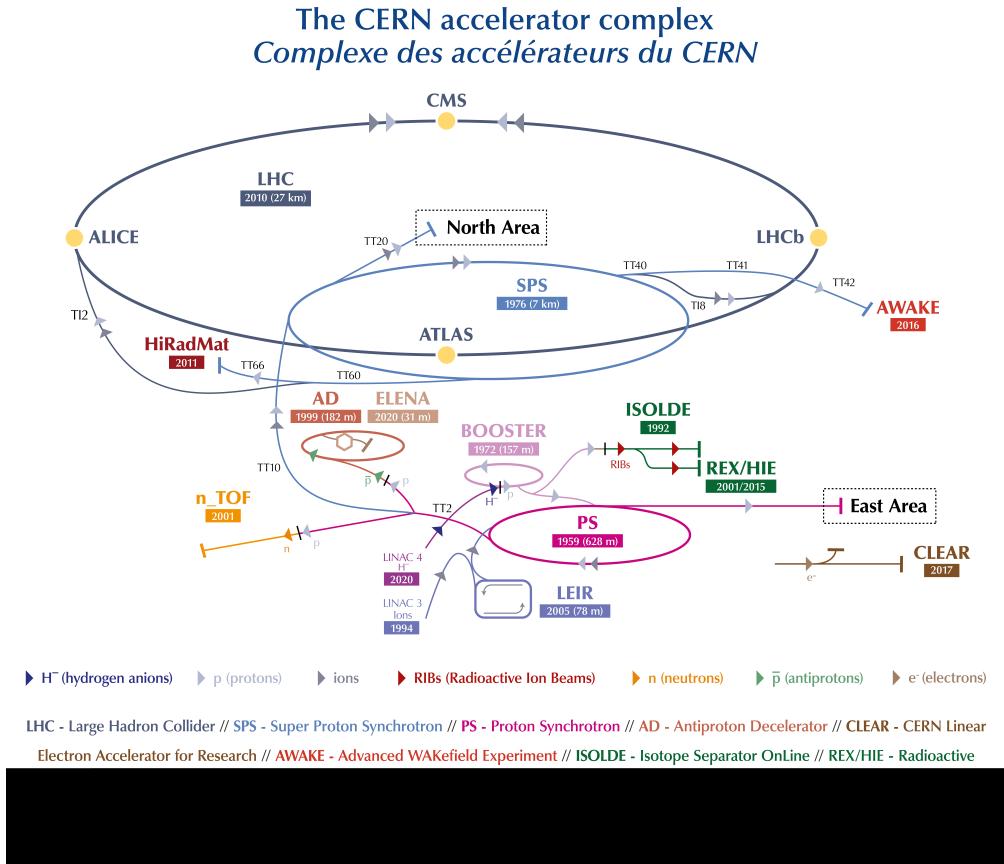


Figure 2.8: Working principle of a TOF detector cell. A charged particle entering the cell will give rise to avalanches in the gas, which are blocked by glass plates mounted inside the cell. The resulting electrons will be picked up at the anode, which divides the cell into two half cells. Adapted from [Adolfsson2020].

space). The strips overlap each other to achieve full coverage within each module. In order to minimize the path length of particles coming from the collision, the modules are tilted so that each strip is facing the interaction point perpendicularly.

## 2.3 Central Barrel Tracking

A tracking algorithm must be used in order to reconstruct the particle tracks from hits in the detector. ALICE uses a Kalman filter algorithm. A Kalman filter uses a parametrisation of the track and optimises trajectories through the addition of additional space points. The algorithm is initiated with a seed, usually consisting of a few clusters in the TPC (the

exception of course being the use of ITS-only tracking). An initial approximation of the track is made with the assumption that the track originates from the collision vertex, and is further extrapolated from hits in the SPD. The process is repeated by removing this constraint to account for the possibility that the track actually originates from a secondary vertex. The track information is continually improved by propagating the track *inwards* and adding additional space points within 4 standard deviations of the best guess of the trajectory while accounting for multiple scattering and energy loss inside detector material. After each iteration, the parameters of the trajectory as well as the covariance matrix of the parametrisation are updated to improve accuracy. If more than one detector hit satisfies the selection criteria, multiple different propagations are tested, and the tracks with the lowest  $\chi^2$  are selected at the end.

As the track is propagated from the outer edge of the TPC inwards, the tracking is further improved by adding hits from the ITS once the inner edge of the TPC is reached. For tracks where the primary vertex constraint has been lifted within the TPC, the same constraint is initially applied to ITS tracks, but repeated without the constraint as well: the higher precision of the ITS makes it possible to find secondary vertices that are much closer to the primary vertex. For tracks where the primary vertex constraint has already been lifted in the TPC, the ITS tracking is done once without imposing the constraint.

Once the track propagation is completed in both the TPC and the ITS, the Kalman filter algorithm is reversed by using track parameters the first procedure as an initial guess, but this time starts from the innermost layer of the ITS and propagates *outwards*. The second iteration is applied in order to remove outliers from the initial track propagation. Once this iteration reaches the outer edge of the TPC, there is an option to incorporate hits from the other detector subsystems such as the Transition Radiation Detector (TRD), and eventually the TOF (as well as others), but this option is not used in this analysis. The main analysis of this thesis uses ITS-only tracking due to the space charge distortions of the time projection chamber discussed in Sec. ???. This results in tracks with fewer space points as input to the Kalman filter track reconstruction, but the performance was found to be comparable to the traditional ITS+TPC Hybrid tracking, Sec. ??.

## 2.4 Triggering System

ALICE is capable of operating at an event rate of a few kHz. The LHC bunch crossing rate, however, is 40MHz. Even though not all bunch crossings result in a collision, an extensive triggering system is required to reduce the event rate and ensure all particles detected in an event are in fact from the same collision. This is also despite that fact that during pp collisions, the accelerator is actually tuned such that the luminosity at ALICE is  $\mathcal{L} \approx 10^{30} - 10^{31} \text{cm}^{-2}\text{s}^{-1}$ , much lower than at ATLAS or CMS with a luminosity of  $\mathcal{L} \approx 10^{30} - 10^{31} \text{cm}^{-2}\text{s}^{-1}$ . This reduces pile-up, but also the effective collision rate [Wenninger:2668326].

The V0 and SPD detectors are the most important detector subsystems for triggering.

The trigger input is handled by the Central Trigger Processor (CTP) that sends a command to all active detectors once it receives a successful trigger. There are three levels of triggers: L0, L1, and L2, and events are only accepted if the highest level, L2, triggers successively. The L0 trigger is simply synchronised with the LHC bunch crossing clock and includes the fastest inputs all processed within  $1.2\ \mu\text{s}$  after the collision. Additional inputs that take longer than  $1.2\mu\text{s}$  and up to  $6.5\mu\text{s}$  are handled by the L1 trigger. The L2 trigger then handles all the slower inputs, with the restriction that it must be completed before the TPC needs to be read out, less than  $88\mu\text{s}$  after the collision.

The signals used by the trigger depend on the trigger configuration. One important configuration is minimum bias (MB), that selects against non-collisions. The MB trigger requires a hit in the both the V0A and V0C detectors. In order to reduce beam-gas events, additional requirements on the V0 timing information and on the correlation between the number of tracklets and clusters in the SPD are enforced. MB events are used in this analysis in the event mixing as well as in the ITS only tracking validation. Another trigger configuration is EG1, as well as EG2. These triggers require a hit in the EMCal up to a certain threshold before tracks can be read out for the event. For EG1, this threshold is  $11\ \text{GeV}/c$ , and  $5\text{GeV}/c$  for EG2.

Once the CTP receives a successful trigger, the CTP will begin a series of read-out operations for each active detector until eventually the trigger input is recorded locally at a Local Data Concentrator (LDC). During this processing, the read-out system will of course be busy, and a signal will be sent to the CTP that blocks triggering until the detectors are ready again. Some of the data from this read-out process is duplicated and sent to the High Level Trigger, where it is processed and decided whether the event should be stored permanently or rejected. This is the final triggering stage. If accepted, the data is compressed by rejecting irrelevant information and the remaining event data will be sent to Global Data Concentrators (GDCs) through the Data Acquisition (DAQ) system. The whole event is built from detector readouts in the GDCs, and is stored in first in Transient Data Storage. Some of the data is also processed online through data quality monitoring and detector algorithms to further reject bad quality events. For run two, the data quality monitoring is manual, and if there is a deterioration in quality, the shift leader may decide to stop a run until the problem is resolved. After all checks are passed, the data will be sent to the GRID for final storage, post-processing, and to be copied to tape. Because the data is so large and often inconvenient to pull for a single analysis, batches of analyses that need access to same dataset are grouped together in *trains*. A train is then sent to the GRID to process a specific set of runs, and is the nominal way most analysis in ALICE are done, often with at least one to two days between analysis iterations. The Berkeley group, however, takes a different approach to this method, discussed in more detail in ??.

# Chapter 3

## Data Analysis

### 3.1 Analysis Summary

This analysis uses the data collected during the  $\sqrt{s_{\text{NN}}} = 5$  TeV p–Pb run in 2013 and during the  $\sqrt{s} = 5$  TeV pp run in 2017. The EMCal trigger was used to select events with a high-momentum calorimeter cluster, a photon  $p_{\text{T}}$ range of 12–40 GeV/ $c$ .

The signal for this analysis, is "prompt" photons, which include "direct photons" and "fragmentation photons". At leading order in perturbative QCD, the direct photons are produced in hard scattering processes such as quark-gluon Compton scattering ( $qg \rightarrow q\gamma$ ) or quark-antiquark annihilation ( $q\bar{q} \rightarrow g\gamma$ ), whereas the fragmentation photons are the product of the collinear fragmentation of a parton ( $q\bar{q}(gg) \rightarrow \gamma + X$ ). At LHC energies, Compton scattering and gluon fusion ( $gg \rightarrow q\bar{q}\gamma$ ) dominate due to the high-gluon density in the proton at small values of Bjorken- $x$ .

Beyond the simplistic leading order picture, the direct and fragmentation components have no physical meaning and cannot be factorized; the sum of their cross sections is the physical observable. For example, the separation between the NLO direct photons and LO fragmentation is arbitrary. However, it is still possible to simplify comparisons with theoretical calculations by applying an isolation criteria. The isolation variable used is the sum of the transverse momentum of the charged particles that are inside an angular cone of radius  $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$  around the photon direction, after subtracting the underlying event.

The main background for this analysis are photons from meson decays, which we will call "decay photons" or  $\gamma^{\text{decay}}$ . The challenge faced in this measurement arises mainly from the small cross-section of the signal compared to that of the decay photon background (about 1% at 10 GeV/ $c$  increasing to about 4% at 30 GeV/ $c$ , according to next-to-leading order calculations [**Arleo:2004gn**]).

This measurement exploits the difference between the electromagnetic shower profiles of prompt photons and of photon pairs from neutral-meson decays. The clusters that pass the isolation and shower shape selections isolated  $\gamma$  candidates or " $\gamma^{\text{iso}}$  candidates".

The main background in within the  $\gamma^{\text{iso}}$  candidate population are from multi-jet events where one jet typically contains a  $\pi^0$  or  $\eta$  that carries most of the jet energy and is misidentified as a photon because it decays into a photon pair that is collinear with respect to the EMCAL cell granularity ( $\Delta\eta \times \Delta\phi \approx 14.3 \times 14.3 \text{ mrad}^2$ ), that is, the two photons are close enough to deposit most of their energy in the same cell and appear as a single shower.

The signal purity of the  $\gamma^{\text{iso}}$  selection is measured by using the "template-fit method", in which the measured shower-shape distribution is fit with the sum of signal and background templates with the relative normalization as the single free parameter<sup>1</sup>. The background template is mostly data-driven, calculated with an anti-isolated sideband requirement, but a MC-based correction is applied to account for estimated biases. The signal template is obtained from photon-jet simulation. The purity of our  $\gamma^{\text{iso}}$  selection is measured to be around 20% at 12 GeV/c and increases to about 55% at 20 GeV/c and above.

Next, the angular correlation of the  $\gamma^{\text{iso}}$  candidates with charged particles is measured. The  $\gamma^{\text{decay}}\text{-hadron}$  correlation function is measured by inverting the shower-shape cut to select merged-clusters from meson decays, and is called the *Background Region* correlation function. Both the signal and background region photon-hadron correlations are corrected for geometrical pair acceptance effects using the mixed-event technique. Afterwards, the  $\gamma^{\text{decay}}\text{-hadron}$  correlation is normalized using the measured purity and subtracted from the normalized the main  $\gamma^{\text{iso}}$  candidate correlations in a process labelled "correlated background subtraction".

After this correlated background subtraction, the uncorrelated background, estimated by the zero-yield-at-minimum (ZYAM) method and checked by using a control region at large  $|\eta^{\text{hadron}} - \eta^\gamma|$  is subtracted from the main  $\gamma^{\text{iso}}$  candidate correlation function. Finally, the away-side of the resulting correlation function is integrated to determine the number of correlated hadrons per  $\gamma^{\text{iso}}$ , i.e. to measure the conditional yield of hadrons. This analysis is performed with photons with  $12 < p_T < 40$  GeV/c, and in intervals of charged particle  $p_T$  and  $z_T \equiv p_{T^h}/p_T^{\gamma^{\text{iso}}}$ .

One of the novel aspects of this analysis is the use of ITS standalone tracking. This approach was developed in order to bypass the serious space-charge distortions that compromised the TPC during the high-luminosity p–Pb data taking in 2013<sup>2</sup>. Furthermore, the ITS-only tracking allowed the 2017 pp run to operate in the CALO mode that yielded a much larger sample than would have been possible otherwise.

The performance of the ITS-standalone tracking (fake rate, efficiency and momentum smearing) was validated by measuring the charged particle spectrum and comparing it with published ALICE measurements at the same center-of-mass energy. These studies showed an agreement between the ITS-standalone measurement and the published data to within

---

<sup>1</sup>Note that this is an standard way to estimate QCD background since at least the Tevatron days. The same exact method is used in the CMS  $\gamma^{\text{iso}}$  and  $\gamma^{\text{iso}}$ -jet measurements in pp and PbPb data, for example in Refs. [Sirunyan:2018gro, Chatrchyan:2012gt].

<sup>2</sup>For details, see <https://alice.its.cern.ch/jira/browse/ATO-351>, <https://alice.its.cern.ch/jira/browse/PWGPP-349>, <https://alice.its.cern.ch/jira/browse/PWGPP-314>

$\approx \pm 5\%$  of the corresponding published data for the range  $0.5 < p_T < 10 \text{ GeV}/c$ , which is the relevant range in this analysis.

One of the main considerations of the larger analysis strategy was to minimize the use of Monte Carlo simulations. By using an isolation variable constructed using only charged particles, the correlations between isolation and shower-shape variables due to the opening angle of neutral-meson decays has been reduced, at the expense of a slightly lower purity. In the template fit analysis, checks were performed that are independent of any input from simulations, suggesting the template fit is not sensitive to the detailed simulation of the exact shower-shape distributions. Moreover, the analysis measures per-trigger quantities such that there is no need to correct for the efficiency of the  $\gamma^{\text{iso}}$  selection.

While an effort was made to collect and use the largest data samples available by pioneering high-rate data taking with ITS+EMCal, the measurement turns out to be dominated by statistical uncertainties. Faced with this reality, efforts to reduce the systematic uncertainties of the measurement were tempered such that they are smaller than the statistical uncertainty.

## 3.2 Datasets

The datasets used in this analysis include the high-luminosity runs of the 2013 p–Pb run (13d,e,f) and the 2017 pp run (17q) that were collected with EMCal triggers, which are listed in Table ??.

Table 3.1: EMCal triggers used in this analysis.

Dataset	Trigger Strings
p–Pb	CEMC7EG1-B-NOPF-CENTNOTRD, CEMC7EG2-B-NOPF-CENTNOTRD,
pp	CEMC7EG2-B-NOPF-CALO, CDMC7DG2-B-NOPF-CALO, CEMC7EG2-B-NOPF-CENT, CDMC7DG2-B-NOPF-CENT

The EMCal gamma triggers (EG1, EG2, DG1, DG2) are based on the summed energy in  $2 \times 2$  adjacent tiles (a tile is composed of an EMCal module,  $2 \times 2$  adjacent cells). The trigger thresholds were 7 and 11  $\text{GeV}/c$  during the 2013 p–Pb run and 5  $\text{GeV}/c$  during the 2017 pp run.

Due to the 2-in-1 magnet design of the LHC, which requires the same magnetic rigidity for both colliding beams, the beams had different energies during the p–Pb run ( $E_p = 4 \text{ TeV}$ ,  $E_{\text{Pb}} = 4 \text{ TeV} \times Z$ , where  $Z = 82$  is the atomic number of lead). In the lead nucleus, the energy per nucleon was therefore  $1.56 \text{ TeV} = (Z/A) \times 4 \text{ TeV}$ , where  $A = 208$  is the nuclear mass number of the lead isotope used. This energy asymmetry results in an average nucleon–nucleon center of mass collision energy of  $\sqrt{s_{\text{NN}}} = 5 \text{ TeV}$  and a rapidity boost of this frame by  $\pm 0.465$  units relative to the ALICE rest frame in the direction of proton beam. Around halfway through the 2013 p–Pb run, the beam directions were flipped, yielding similar integrated luminosities in both beam configurations.

During the 2013 p–Pb run period, the TPC suffered from space-charge distortions<sup>3</sup> that affect tracking, leading to a very drastic drop in efficiency for tracks with  $p_T > 4 \text{ GeV}/c$ . We bypass this issue by using ITS-only tracking as detailed in Section ???. For the 2017 pp data, the TPC was also inactive due to the high luminosity of the runs considered in this analysis. We use the 17q period, during which all six layers of the ITS were active.

The average number of inelastic collisions per bunch crossing,  $\mu$ , is 0.020–0.060 for the 2013 p–Pb data set and in the range 0.015–0.045 for the 2017 pp dataset<sup>4</sup>.

### 3.3 Monte Carlo simulations

We use Monte Carlo (MC) simulations to obtain the signal shower-shape distributions for the template fits (section ??) and to study tracking performance (section ??).

The simulations of hard processes are based on the PYTHIA event generator. In PYTHIA, the signal events are included via  $2 \rightarrow 2$  matrix elements with  $gq \rightarrow \gamma q$  and  $q\bar{q} \rightarrow \gamma g$  hard scatterings, defined at the leading order, followed by the leading-logarithm approximation of the partonic shower. The soft underlying events in pp collisions as well as fragmentation are included with the default PYTHIA models.

For the simulation of p–Pb events, the pp samples are embedded into p–Pb inelastic events generated with DPMJET. The boost of  $\Delta y = +0.465$  in the direction of the proton beam is reproduced.

### 3.4 Event Selection

The following event selection criteria is used to ensure good event quality and uniform acceptance:

- Run passes QA for EMCAL and ITS (the selected runs are listed in Table ??).
- At least one EMCAL cluster with  $p_T > 12 \text{ GeV}/c$ .
- Selected at least one of the EMCAL triggers (logical OR of the trigger strings listed in Table ??).
- Valid vertex ( $|z| \neq 0.0$ ) and  $|z| < 10 \text{ cm}$

Where the EMCAL is mentioned, we use the same criteria on the DCAL (used in the 17q dataset). We use EMCAL in text for brevity. The number of events that pass our selection in

---

<sup>3</sup>For more information on the problems with space-charge distortions due to high-luminosity in p–Pb run, see: <https://alice.its.cern.ch/jira/browse/PWGPP-314>.

<sup>4</sup>This information can be found in [http://aliquevs.web.cern.ch/aliquevs/data/2013/LHC13d/pass4/global\\_properties.pdf](http://aliquevs.web.cern.ch/aliquevs/data/2013/LHC13d/pass4/global_properties.pdf) and [http://aliquevs.web.cern.ch/aliquevs/data/2017/LHC17q/cpass1\\_pass1/global\\_properties.pdf](http://aliquevs.web.cern.ch/aliquevs/data/2017/LHC17q/cpass1_pass1/global_properties.pdf)

each sample is shown in Table ???. We report the events selected for each trigger separately, as well as the logical OR combination. In p–Pb events, the number of events is dominated by the EG1 trigger (11 GeV/c threshold), and by the EG2 trigger (5 GeV/c) in pp collisions.

Table 3.2: Number of events that passed our full event selection for each of data taking period used in this analysis. The numbers are also shown separately for EG1 (DG1) and EG2 (DG2) triggers.

Dataset	$N^{EG1  EG2}$	$N^{EG1}$	$N^{EG2}$
13d	134024	133326	12528
13e	198108	196745	22409
13f	340607	338198	38353
13f_new	241870	240074	30310
Dataset	$N^{EG2  DG2}$	$N^{EG2}$	$N^{DG2}$
17q	406934	301086	119498

## 3.5 Calorimeter Cluster Reconstruction

### 3.5.1 Definition

EMCal clusters are formed by a clustering algorithm that combines signals from adjacent towers. We use calorimeter clusters defined with the “V1” algorithm. This algorithm starts from a “seed” cell, found from a local-maximum scan, and adds “neighbor” cells to the cluster if they are above a given threshold. The cluster definition is exclusive, i.e. once a cell is assigned to a cluster, it is not considered for other clusters. The minimum energy for the seed and neighbor were set to 500 and 100 MeV respectively; these values are several times larger than the standard deviation of the electronic noise<sup>5</sup>.

### 3.5.2 Corrections

We apply several corrections at the cell level, implemented within the “EMCal Correction Framework,”<sup>6</sup> before the clustering algorithm is run over the data and simulations. The following corrections are applied:

- “CELLENERGY”

This performs an energy calibration of cells, with coefficients obtained with  $\pi^0 \rightarrow \gamma\gamma$  mass measurements.

---

<sup>5</sup>Some photon analysis use a 50 MeV threshold, but 100 MeV has been found to improve cell time measurements. The 100 MeV threshold has been used for example in Ref [Acharya:2017tlv].

<sup>6</sup>[http://alidoc.cern.ch/AliPhysics/master/\\_r\\_e\\_a\\_d\\_m\\_eemc\\_corrections.html](http://alidoc.cern.ch/AliPhysics/master/_r_e_a_d_m_eemc_corrections.html)

- “CELLBADCHANNEL”

This removes cells that declared hot or dead for a given run period.

- “CELLTIMECALIB”

This correction applies constant offsets, which are arbitrary, to the cell time measurements to minimize the spread among cells.

- “CELLEMULATECROSSTALK”.

This correction, described in detail in Ref [CrossTalk], modifies the simulated cell energies to emulate the cell cross-talk that has been observed in data. This is applied to all the simulations described in Table ??.

### 3.5.3 Selection

The following selection is applied on the resulting clusters<sup>7</sup>:

- Cluster  $p_T$  cut:  $12 < p_T < 40 \text{ GeV}/c$ .

- Cluster pseudorapidity:  $|\eta| < 0.67$

The cluster pseudorapidity is corrected for the position of the primary interaction vertex.

- Number of cells cut:  $N_{\text{cell}} \geq 2$

This requirement removes clusters that are likely dominated by noise.

- Exoticity cut:  $E_{\text{cross}}/E_{\text{cluster}} > 5\%$

We remove “exotic” or “spiky” clusters likely coming from slow neutrons or highly-ionizing particles hitting the avalanche photo-diode of a cell by a requirement on the ratio of the summed energy around the leading cell to the total cluster energy.

- Cluster time cut:  $|t| < 20 \text{ [ns]}$

We require a cluster time measurement of  $|t| < 20 \text{ ns}$  to remove out-of-bunch pileup.

- Number of local maxima cut:  $N_{LM} < 3$

This cuts suppresses background and improves the MC simulation description of the background [Acharya:2019jx].

- Distance seed-cell to bad-channel  $\geq 1$  cells.

Figures ?? and ?? show the distribution of the variables used in the cluster selection and the effect of sequential selection (“cut flow”) for the p–Pb and pp data respectively. Table ?? shows a summary of the effect of sequential selection on the number of selected clusters in both pp and p–Pb data.

---

<sup>7</sup>This event selection also closely follows previous and concurrent isolated-photon spectra analyses in pp and p–Pb data.

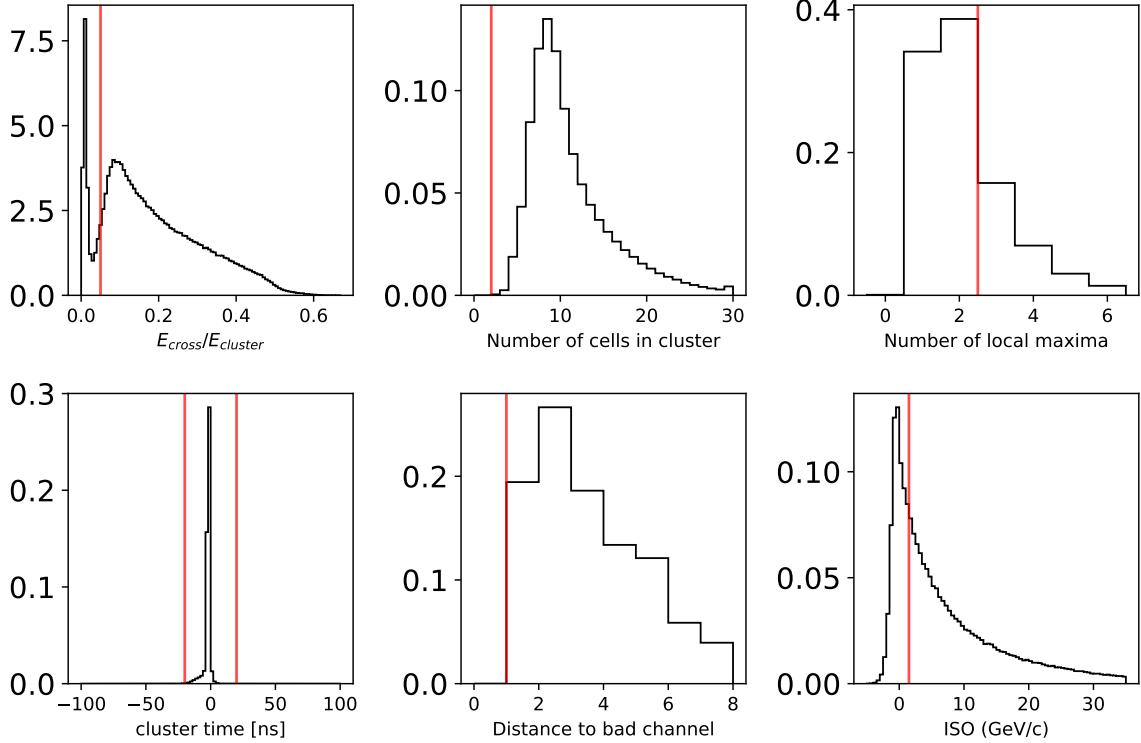


Figure 3.1: Distribution of variables used in the cluster selection of p–Pb data. The red vertical lines represent the cuts used. The cluster cuts get applied sequentially, i.e. the clusters cut with a given variable do not appear in the next.

Table 3.3: Number of clusters, with  $12 < p_T < 40 \text{ GeV}/c$ , that pass our selection in 2013 p–Pb and 2017 pp data.

Selection	p–Pb data	pp data
$ \eta  < 0.67$	714834	385220
$E_{\text{cross}}/E_{\text{cluster}} > 5\%$	613560	323750
$N_{\text{cell}} \geq 2$	613560	323750
$N_{LM} < 3$	443102	231490
$ t  < 20 \text{ [ns]}$	441639	171470
Distance-to-bad channel $\geq 1$	441639	171470
$\text{ISO} < 1.5 \text{ GeV}/c$	137895	58638
$0.1 < \sigma_{\text{long}}^2 < 0.3$	40027	16628

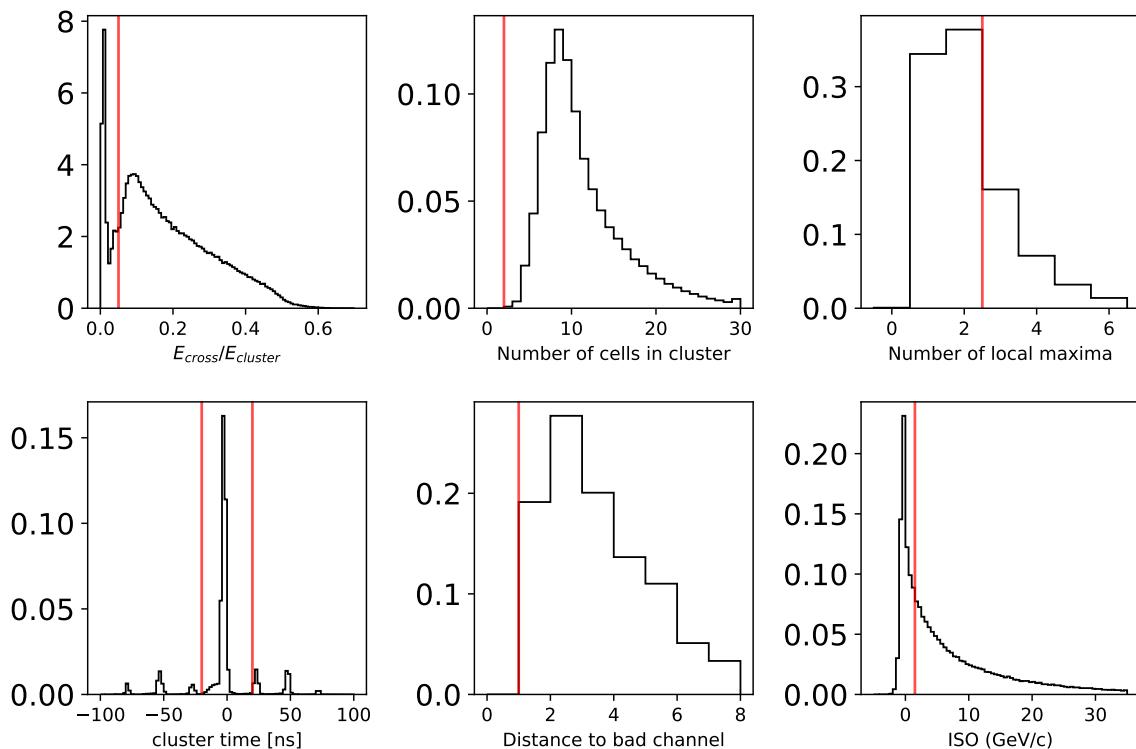


Figure 3.2: Distribution of variables used in the cluster selection in pp data. The red vertical lines represent the cuts used. The cluster cuts get applied sequentially, i.e. the clusters cut with a given variable do not appear in the next.

## 3.6 Isolated Prompt Photon Selection

A key challenge of this measurement is the relatively small cross section of prompt photons compared to decay photons. In order to reduce the background of decay photons, and to identify prompt photons, a combination of isolation and electromagnetic shower profile selections are made. Photons are observed in their final state as an reconstructed cluster in the calorimeter. Clusters are obtained by grouping all adjacent cells with common sides whose energy is above 100 MeV, starting from a seed cell with at least 500 MeV. Furthermore, a cluster must contain at least two cells to remove single-cell electronic noise fluctuations. Clusters are required to have a minimum  $p_T$  of  $p_T^{\gamma \text{iso}} \geq 12 \text{ GeV}/c$ . The time of the highest-energy cell in the clusters relative to the main bunch crossing must satisfy  $\Delta t < 20 \text{ ns}$  to reduce out-of-bunch pileup. In order to limit spurious signals caused by particles hitting the EMCAL APDs, clusters are required to have  $E_{\text{cross}}/E_{\text{cluster}} > 0.05$ , where  $E_{\text{cross}}$  is the sum of the energy in the cells adjacent to, but not including, the leading cell, and  $E_{\text{cluster}}$  is the total energy of the entire cluster. The number of local maxima in the cluster is required to be less than three to reduce hadronic background. To select against photons from neutral-meson decays or fragmentation processes, a shower profile selection and isolation requirement are applied to these clusters.

### 3.6.1 Shower Profile Selection

The primary background for this measurement are photons from the 2-body decay channel of neutral mesons  $\pi^0$ 's. A  $\pi^0$  with a higher  $p_T$  will decay with a smaller opening angle between the two decay photons. As the opening angle becomes smaller, the electromagnetic showers from the decay photons get closer together in the EMCAL until they begin to overlap. For this reason, photons from  $\pi^0$  decays begin to merge into a single cluster in the EMCAL above approximately 6  $\text{GeV}/c$ . This cluster, made up of two showers from decay photons, will therefore tend to have a more elongated shower profile than a cluster resulting from a single, ideally prompt, photon. Thus, in order to reject clusters produced by two photons from a meson decay, and select clusters from single photons, we select clusters using variables that encode the shape of the calorimeter shower.

A variable that encodes the shape of a cluster's electromagnetic shower profile for this purpose is  $\sigma_{\text{long}}^2$ . The  $\sigma_{\text{long}}^2$  variable is defined as the square of the larger eigenvalue of the energy distribution in the  $\eta$ - $\varphi$  plane:

$$\sigma_{\text{long}}^2 = (\sigma_{\varphi\varphi}^2 + \sigma_{\eta\eta}^2)/2 + \sqrt{(\sigma_{\varphi\varphi}^2 - \sigma_{\eta\eta}^2)/4 + \sigma_{\varphi\eta}^2}, \quad (3.1)$$

where  $\sigma_{ij}^2 = \langle ij \rangle - \langle i \rangle \langle j \rangle$  are the covariance matrix elements; the integers  $i, j$  are cell indices in  $\eta$  and  $\varphi$  axes;  $\langle ij \rangle$  and  $\langle i \rangle, \langle j \rangle$  are the second and the first moments of the cluster position cell. The position is weighted by  $\max(\log(E_{\text{cell}}/E_{\text{cluster}}) - w_0, 0)$ . Following previous work [Acharya:2018dqe], the cutoff in the log-weighting is chosen to be  $w_0 = -4.5$ . Cells that contain less than  $e^{-4.5} = 1.1\%$  of the total cluster energy are not considered in the  $\sigma_{\text{long}}^2$

calculation. Thus,  $\sigma_{\text{long}}^2$  discriminates between clusters belonging to single photons, having a  $\sigma_{\text{long}}^2$  distribution which is narrow and symmetric, and merged photons from neutral meson decays, which are asymmetric and have a distribution dominated by a long tail towards higher values.

Most single-photon clusters yield  $\sigma_{\text{long}}^2 \approx 0.25$ , as shown in Figure ?? where the signal is displayed in blue. Consequently, a cluster selection of  $\sigma_{\text{long}}^2 < 0.30$  is applied irrespective of  $p_{\text{T}}$ . Simulations indicate this results in a signal efficiency of about 90% with no significant  $p_{\text{T}}$  dependence.

Fig. ?? shows an example of a cluster with an elongated shower profile with a large  $\sigma_{\text{long}}^2$ , and a narrow, symmetric cluster with a smaller  $\sigma_{\text{long}}^2$ .

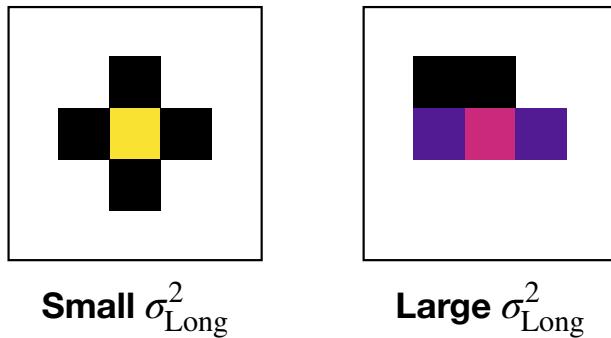


Figure 3.3: Cartoon of a narrow EM shower profile with a small  $\sigma_{\text{long}}^2$ (left), and an elongated shower profile with a larger  $\sigma_{\text{long}}^2$ .

The shower shape profile is an important tool to help discriminate between clusters belonging to single photons, for which the  $\sigma_{\text{long}}^2$  distribution is narrow and symmetric, and merged photons from neutral-meson decays, for which the distribution is dominated by a long tail towards higher values. This analysis uses a cutoff of  $\sigma_{\text{long}}^2 < 0.3$  to identify single photon candidates, which will be discussed in Sec. ??.

### 3.6.2 Photon Isolation Requirement

At leading order in pQCD, prompt photons are produced in  $2 \rightarrow 2$  processes surrounded by very little hadronic activity, in contrast to fragmentation photons and high  $p_T\pi^0$ 's found within a jet. Beyond leading order, the direct and fragmentation components cannot be factorized. As a result, the sum of their cross sections becomes the physical observable.

Despite this, the contribution from fragmentation photons can be suppressed by enforcing an isolation criteria, where the energy surrounding a photon must be less than a certain threshold. Theoretical calculations can also be simplified through the use of an isolation

requirement. [PhysRevD.82.014015]. This also has the benefit of suppressing the background from decays of neutral mesons often found within jets.

The simplest definition of isolation is defined as the scalar sum of the transverse momentum of charged particles within an angular radius,  $R = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2}$ , around the cluster direction. This measurement uses  $R = 0.4$ , which is a common value used in various jet measurements.

$$p_T^{\text{iso,raw}} = \sum_{\text{track } \in \Delta R < 0.4} p_T^{\text{track}} \quad (3.2)$$

This does not, however, take into account the energy arising from the underlying event, described in the following section.

### 3.6.3 Underlying Event Estimation for Photon Isolation

The underlying event (UE) is defined as the sum of all processes that make up the final hadronic state in a collision, excluding the leading order hard scattering. When the two nuclei, lorentz contracted into discs tiny fraction of a femtometer thick, overlap or collide, a small fraction of the incident partons suffer hard perturbative interactions as the discs overlap initially. Most of the incident partons, however, lose some energy but are not deflected by any large angle. Most of these interactions are soft and involve little transverse momentum transfer. In the language of fields and particles, as the two discs of strongly interacting transverse color fields and associated color charges collide, some color charge exchange occurs between the discs, and longitudinal color fields are produced, which fill the space between the two receding discs, reducing the energy in the discs themselves, and then gradually decay into  $q\bar{q}$  pairs and gluons. These processes can be labelled as multi-parton interactions, as well as both initial and final state radiation, but the term underlying event also includes measured beam fragments. Essentially, the underlying event is made up of all the particles not directly associated with the initial hard scattering of the collision.

Here we describe the method used to estimate the underlying event for the purposes of correcting the isolation requirement (not to be confused with the later section ??, where the contribution from the underlying event to the azimuthal correlation measurement is described).

We use the jet area/median method<sup>8</sup> which estimates the underlying event energy density,  $\rho$ , from the median of the distribution of the transverse momentum densities of the jets in the event [Cacciari:2009dp]. Jets are reconstructed by running the  $k_T$  reconstruction algorithm over all charged particles in the event, using a resolution parameter of  $R = 0.3$ . The  $k_T$  algorithm is used here in place of the more standard anti- $k_T$  as it groups particles with the lowest momentum first to construct the jet. This makes the  $k_T$  algorithm more sensitive to the softer objects in the event, and therefore more suitable for studying the underlying

---

<sup>8</sup>From the FASTJET software package::VoronoiAreaSpec [http://www.fastjet.fr/repo/doxygen-2.4.5/classfastjet\\_1\\_1VoronoiAreaSpec.html](http://www.fastjet.fr/repo/doxygen-2.4.5/classfastjet_1_1VoronoiAreaSpec.html)

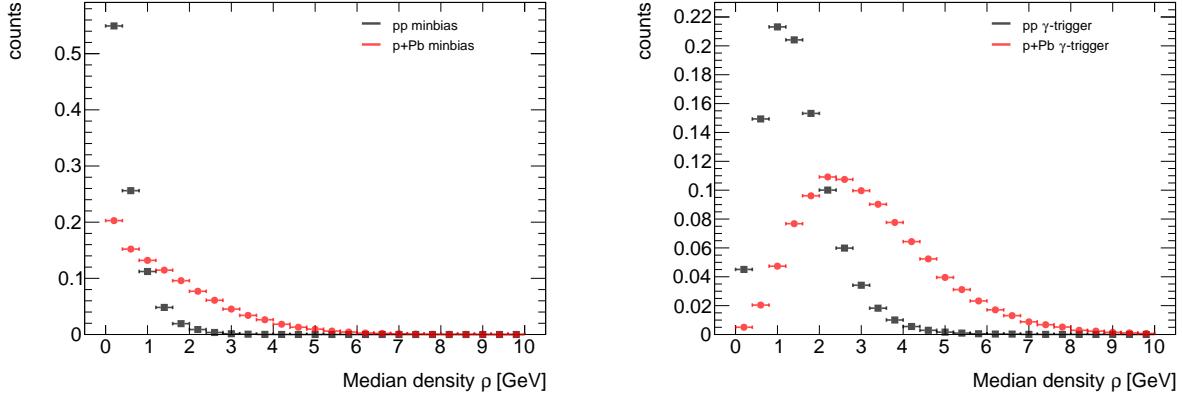


Figure 3.4: Distribution of the median charged-particle transverse momentum density,  $\rho$ , in pp and p-Pb data, for a minimum-bias selection (left panel) and in photon-triggered events (right panel).

event. The transverse momentum density of each jet is simply the momentum of the jet divided by its area, determined by the sum of the voronoi cells of each particle within the jet<sup>9</sup>. This median calculation is described in Equation ??:

$$\rho = \text{med} \left\{ \frac{\sum_{i \in J'_k} p_{T,i}}{\sum_{i \in J'_k} A_i} \right\} \quad (3.3)$$

where  $p_{T,i}$  is the transverse momentum, and  $A_i$  the Voronoi area of the particle  $i$  within the jet,  $J'_k$ , reconstructed for UE estimation purpose. The median is determined from all jets in the event with the important exception of the two leading (highest moment) jets in the event, as those are most often associated with the hard scattering of the collision. This therefore assumes that most of the charged particles in the event is made up of soft particles, and that the charged particles originating from the hard scattering of the collisions are reasonably contained within the leading jets of the event [Cacciari:2009dp].

The charged-particle density,  $\rho$ , is calculated for each event. Figure ?? shows the distribution of  $\rho$  for minimum bias and gamma-triggered events in pp and p-Pb. Average values are 3.2 GeV/c in photon-triggered events in p-Pb and 1.6 GeV/c in pp collisions, demonstrating the a larger underlying event activity in p-Pb compared to pp.

The mean and standard deviation for each distribution is shown in Table ?? . The difference in UE-density in p-Pb is expected due to the increased number of nucleon-nucleon collisions. The UE-densities shown here are still about a factor of 50 lower than in central Pb-Pb collisions.

<sup>9</sup>The voronoi cell is the region for each "seed", or particle, that consists of all points in the same plane that are closer to that seed than to any other.

Table 3.4: Median transverse momentum density mean and standard deviation in minimum-bias and photon-triggered events in pp and p–Pb data, calculated with negligible statistical uncertainties.

	pp minbias	pp $\gamma$ -trigger	p–Pb minbias	p–Pb $\gamma$ -trigger
$\langle \rho \rangle$	0.49 GeV/c	1.51 GeV/c	1.56 GeV/c	3.19 GeV/c
$\sigma_\rho$	0.47 GeV/c	0.85 GeV/c	1.32 GeV/c	1.60 GeV/c

### 3.6.4 UE Correction to Isolation Variable

For each cluster in the event, the underlying event is subtracted using the measured charged-particle density  $\rho$  that is calculated event-by-event as described in Section ??:

The result is an average subtraction for the isolation cone of  $R = 0.4$  is about 1.6 GeV/c and 0.8 GeV/c for p–Pb and pp collisions, with a standard deviation of 0.9 GeV/c and 0.4 GeV/c, respectively.

For photons near the edge of the detector, the isolation energy requirement is scaled to account for any missing area in the isolation cone<sup>10</sup>. A check on this scaling procedure was also done in Section ??.

$$p_T^{\text{iso}} = p_T^{\text{iso,raw}} - \rho \times \pi(0.4)^2. \quad (3.4)$$

Figure ?? shows the isolation distribution before and after underlying event subtraction for p–Pb and pp collisions. The distributions have a positive tail that decreases exponentially. This is likely due to the sensitivity of the isolation variable on multi-jet production cross section. The difference between the p–Pb and pp distribution at low  $p_T^{\text{iso}}$  values can be attributed to the effect of enhanced soft-particle production in p–Pb collisions, i.e. a larger underlying event due to the presence of the pPb nucleus. The underlying event subtraction modifies the isolation distribution only slightly at high  $p_T$ . At low and negative  $p_T$ , however, the distributions show a negative tail after subtraction, which arises from an over-subtraction of the underlying event. This occurs due to region-to-region fluctuations in the underlying event, where a cluster contains an energy density that is smaller than the median calculated according to Section ???. In both cases, this tail falls by more than three orders of magnitude by  $p_T^{\text{iso}} = -3$  GeV/c, indicating that over-subtraction is a small effect.

The left panel of Figure ?? shows the distribution of cluster isolation after UE subtraction for photon-jet and dijet simulations of p–Pb data (see Table ??). The distributions exhibit different behavior: whereas the dijet simulation shows a prominent exponential tail at large  $p_T^{\text{iso}}$  values, the photon-jet simulation shows a more Gaussian-like shape that is mostly symmetric with the exception of a very small fraction of events that have large  $p_T^{\text{iso}}$  values. In both cases, however, the negative tail falls rather sharply, as it arises from region-to-region fluctuations of the UE that are independent of the hard-process involved.

---

<sup>10</sup>The final isolated photon-hadron correlations are normalized to the number of reconstructed photons. As a result, the  $\gamma^{\text{iso}}$  efficiency was not studied in detail.

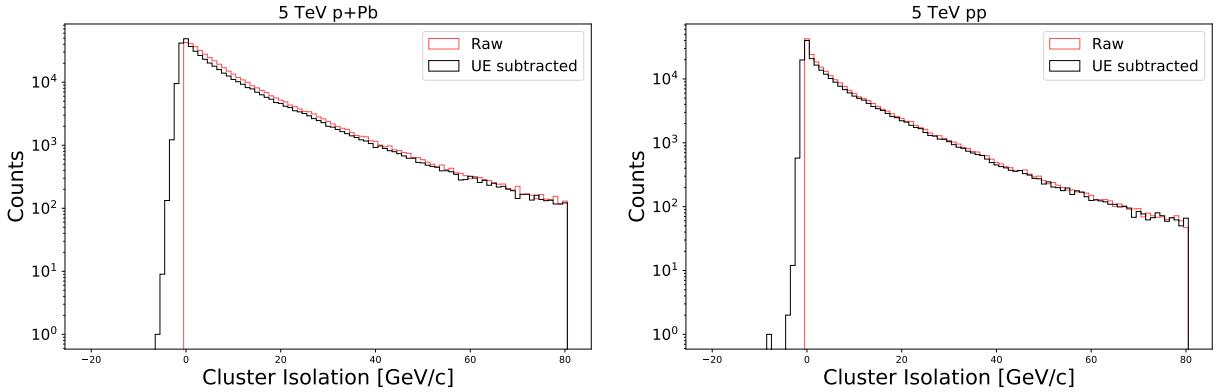


Figure 3.5: Cluster isolation before and after underlying event subtraction in p–Pb (left panel) and pp (right panel) collisions.

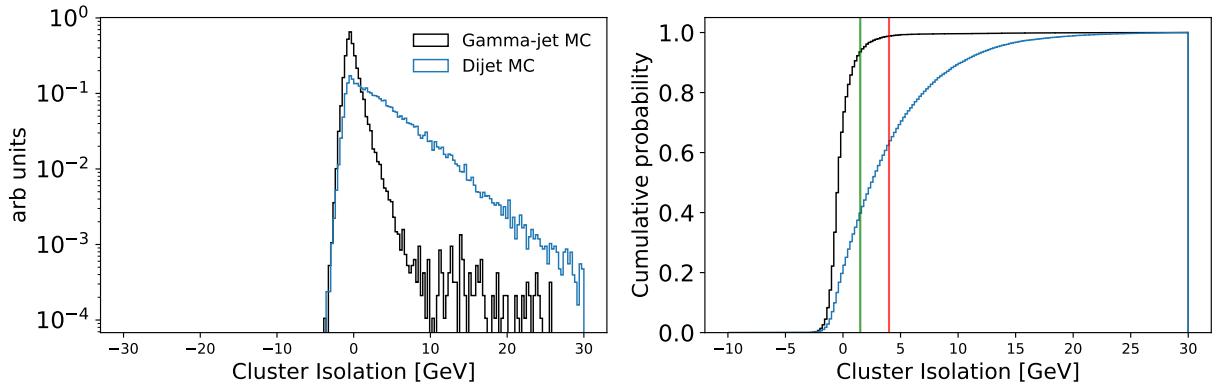
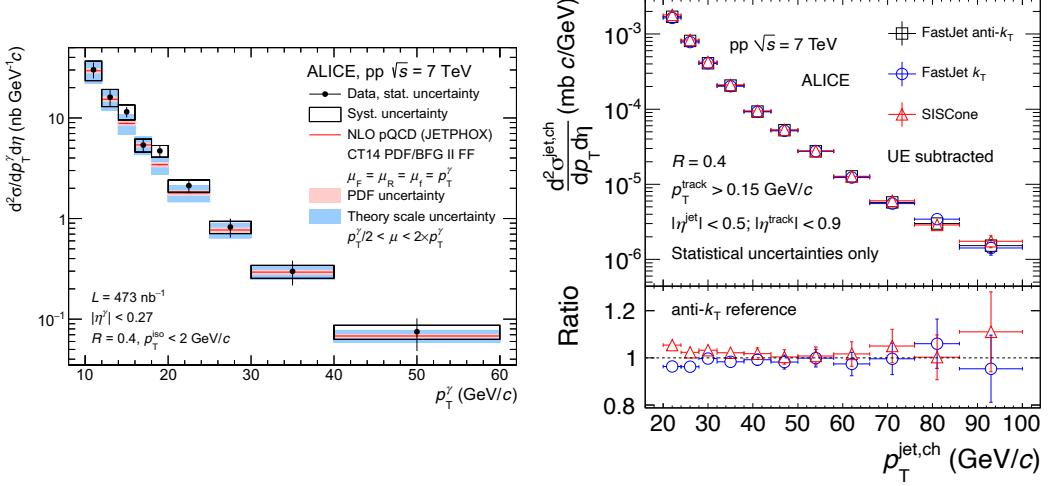


Figure 3.6: Isolation distribution of clusters that pass our selection in p–Pb photon-jet and dijet simulations, and corresponding cumulative distribution. Two vertical lines at  $p_T^{\text{iso}} = 1.5$  GeV/ $c$  (green) and  $p_T^{\text{iso}} = 5.0$  GeV/ $c$  are shown in the right panel for reference.

The cumulative distributions (Figure ??, right panel) show that a  $p_T^{\text{iso}} < 1.5$  GeV/ $c$  selection keeps about 90% of the signal and rejects about 60% of the background. This relatively loose photon isolation criteria is used in order to reduce the dependence of the results on the details of the simulation of the detector noise, tracking resolution, and the underlying event.



### 3.6.5 Remaining Background after Photon Selection

This isolation cut of  $p_T^{\text{iso}} < 1.5 \text{ GeV}/c$  is used in conjunction with the shower-shape cut of  $0 < \sigma_{\text{long}}^2 < 0.3$  to complete the isolated-photon selection or “ $\gamma^{\text{iso}}$  selection”. The population of clusters that pass this selection are labelled “ $\gamma^{\text{iso}}$ -candidates” (rather than simply ”prompt photons”) because there is still a significant fraction of remaining background. Some other sources of background not yet mentioned arise from charged-to-neutral fluctuations of jet fragmentation that leads to low observable  $p_T^{\text{iso}}$  (that considers only charged-particles). However, the main background present in the  $\gamma^{\text{iso}}$  selection arise from multi-jet events where one jet typically contains a  $\pi^0$  or  $\eta$  which carries most of the jet energy, and is therefore surrounded by relatively less energy within the jet. The pair is also frequently misidentified as a single photon because it decays into a pair of photons that are collinear with respect to the EMCAL cell granularity. The first indication of this is shown in Figure ??, where approximately 40% of the dijet cross section (expressed as a cumulative probability as a function of cluster isolation, shown in blue) is within  $p_T^{\text{iso}} < 1.5 \text{ GeV}/c$ .

Distributions of background and signal as function of  $\sigma_{\text{long}}^2$  are shown later in ??

Figure ?? expands on this. The left panel shows the isolated photon differential cross section as a function of  $p_T^{\gamma^{\text{iso}}}$  in proton-proton collisions at  $\sqrt{s} = 5.02 \text{ TeV}$  measured by the ALICE detector. The right panel of Figure shows the differential cross section of charged jets as a function of  $p_T^{\text{jet}}$ . The cross section for an isolated photon at  $p_T = 20 \text{ GeV}/c$  is roughly  $2 \text{ nb GeV}/c^{-1}$ . In contrast, the cross section for charged jets at  $p_T = 20 \text{ GeV}/c$  is roughly  $2 \times 10^{-3} \text{ mb GeV}/c^{-1}$ , approximately three orders of magnitude larger.

Of course, jets containing a neutral meson with a large fraction of its total momentum make up only a fraction of the total jet cross section, and recoil partons produced from the same hard scattering as prompt photons will contribute to the total charged jet cross section. But the stark difference between the isolated photon and charged jet cross sections speaks to the rarity of isolated photons in these collisions and the abundance of background and illustrates the need to measure the purity of our  $\gamma^{\text{iso}}$ -candidate selection.

## 3.7 Purity

The isolation and shower shape selections remove the bulk of the neutral meson decay background, but a substantial fraction of the  $\gamma^{\text{iso}}$  candidates are still background photons. It is therefore necessary to quantify the ratio of true signal photons in our candidate sample in order properly subtract it. The estimate of the ratio of true signal photons in our  $\gamma^{\text{iso}}$  sample is called the *purity*.<sup>11</sup>

### 3.7.1 The Template Fit Method

The purity of the isolated photon sample is determined with a two-component template fit, a method used by the CMS collaboration in Ref. [Sirunyan:2017qhf]. The distribution of the shower shape variable for the isolated cluster sample is fit to a linear combination of a signal distribution and the background distribution. The shape of the signal distribution is determined by a photon-jet simulation (see Table ??) and the shape of the background distribution is determined from data using an anti-isolated sideband<sup>12</sup> with an additional correction computed from a dijet simulation. This is described in more detail in the following sections.

### 3.7.2 Signal Template and Background Templates

The shape of the background distribution of the shower-shape for isolated clusters is estimated with a *sideband* technique: the shower shape distribution of clusters from isolated decay photons is estimated with clusters that are anti-isolated but pass all other selection criteria. This method assumes that the correlation between the isolation variable and shower shape variable can be corrected for; the procedure for doing so is described below. The signal and sideband regions defined using the isolation variable are illustrated in Figure ??.

For simplicity, the same definitions are used for pp and p–Pb data. The lower bound of the sideband region is defined as ISO = 5 GeV/c; according to photon-jet simulations, less than 1% of prompt photons are beyond this range. The upper bound is chosen such that the sideband is as narrow as possible, to minimize a possible bias to the shower-shape distribution due to a positive correlation with ISO, while still containing a number of clusters comparable to the signal region. A more rigorous study on the sensitivity of our purity estimate on the choice of sideband region is shown in Section ??.

Figure ?? summarizes the signal and background templates used in the template fit. The distributions are quite different, which is key for the stability of the template fit. The background shape in the  $\sigma_{\text{long}}^2$  variable shows a peak in the single-shower region but a “bump” that reflects a  $\pi^0$  peak. In both cases, the peaks in the single-shower region that are observed in the background templates come mostly from collinear  $\pi^0 \rightarrow \gamma\gamma$  decays.

---

<sup>11</sup>For a more rigorous discussion of the Purity calculation, please see Alwina Liu’s UC Berkeley Thesis.

<sup>12</sup>The inversion of an isolation cut to estimate QCD background is a standard technique in several measurements at the LHC and previous hadron colliders.

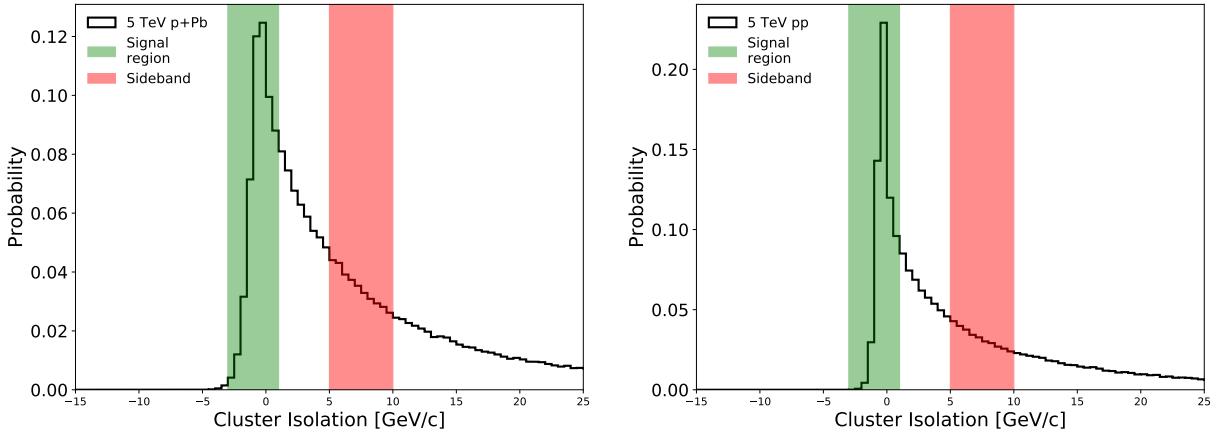


Figure 3.7: Isolation variable distribution of clusters with  $p_T$  between 12 and 16 GeV/ $c$  in p–Pb data (left panel) and pp data (right panel). The green shaded area represents the signal region ( $\text{ISO} < 1.5 \text{ GeV}/c$ ); the red represent the sideband ( $5 < \text{ISO} < 10 \text{ GeV}/c$ ) used to estimate the background template.

The background template is corrected for a bias due to correlations between the shower-shape and isolation variables [**Khachatryan:2010fm**]. This correlation leads to clusters in the isolation sideband having a somewhat higher hadronic activity than the true isolated background. Consequently, a background template constructed from this sideband region has an increased number of background-like clusters and purity values obtained using this systematically overestimate the true purity. A correction for this bias,  $R(\sigma_{\text{long}}^2)$ , is determined using dijet simulated events which also contain the correlation between trigger photon shower-shape and isolation cut. The ratio of the shower-shape distributions of clusters in the signal (Iso,  $p_T^{\text{iso}} < 1.5 \text{ GeV}/c$ ) region and sideband (Anti-iso,  $5.0 < p_T^{\text{iso}} < 10.0 \text{ GeV}/c$ ) region is constructed via

$$R(\sigma_{\text{long}}^2) = \frac{\text{Iso}_{\text{MC}}(\sigma_{\text{long}}^2)}{\text{Anti-iso}_{\text{MC}}(\sigma_{\text{long}}^2)}. \quad (3.5)$$

This ratio of shower shape distributions is applied as a multiplicative correction to the background template:

$$B^{\text{corr.}}(\sigma_{\text{long}}^2) = \text{Anti-iso}_{\text{data}}(\sigma_{\text{long}}^2) \times R(\sigma_{\text{long}}^2). \quad (3.6)$$

This background template correction results in an absolute correction on the purity of 8%–14% depending on the cluster  $p_T$ . An example of a fit with and without the correction is shown in Figure ???. The correction greatly improves the fit.

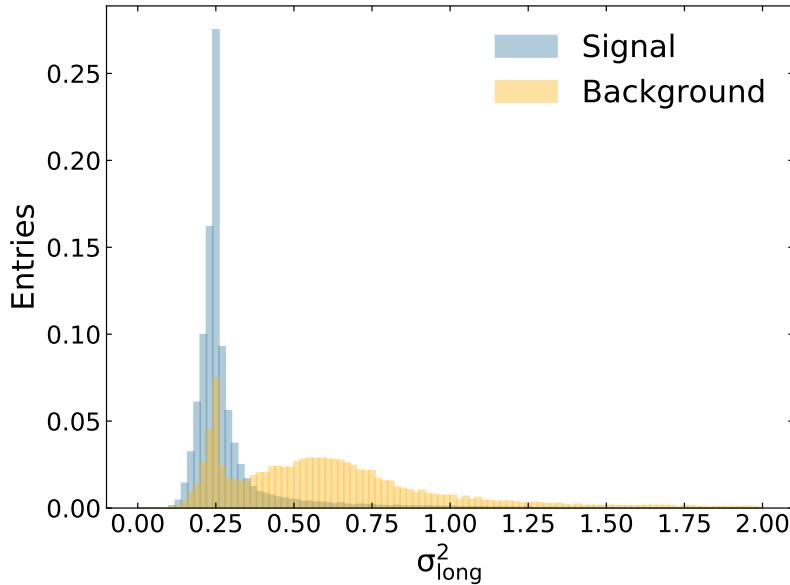


Figure 3.8: Normalized signal (blue) and background (yellow) distributions used as input for the template fit. These distributions correspond to clusters with  $p_T$  in the 15–20 GeV/ $c$  range.

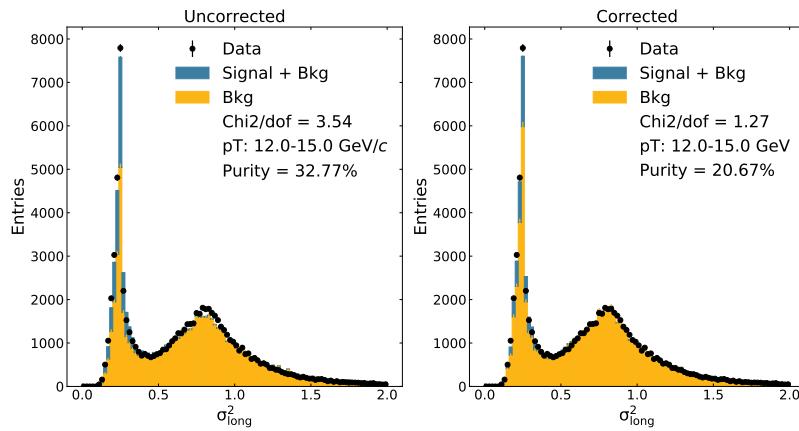


Figure 3.9: An example of the template fit with and without the background template correction in p–Pb for clusters with  $12 < p_T < 15$  GeV/ $c$ . The goodness of fit is better after the correction and the purity is significantly lower.

### 3.7.3 Fit Results

The distribution of isolated clusters is fit with a linear combination of the signal and background templates. The MINUIT [James:1975dr] package is used for  $\chi^2$  minimization and the MIGRAD package for uncertainty estimation. The only free parameter in the fit is the number of signal clusters,  $N_{\text{sig}}$ , because the overall normalization,  $N$ , is fixed to the total number of isolated clusters:

$$N^{\text{observed}}(\sigma_{\text{long}}^2) = N_{\text{sig}} \times S(\sigma_{\text{long}}^2) + (N - N_{\text{sig}}) \times B(\sigma_{\text{long}}^2), \quad (3.7)$$

where  $S(\sigma_{\text{long}}^2)$  and  $B(\sigma_{\text{long}}^2)$  are the normalized signal and background templates. Examples of template fits are shown in Figure ???. The peaks observed in the background templates originate mostly from collinear or very asymmetric  $\pi^0 \rightarrow \gamma\gamma$  decays. Photons from  $\eta$  decays also contribute to the peaks in the background template.

The purities in the pp and p–Pb datasets are compatible within the uncertainties. The  $\text{Weights}(\sigma_{\text{long}}^2)$  function for different  $p_{\text{T}}$  ranges is shown in Appendix ?? and the evaluation of the systematic uncertainty associated with this correction is described in Section ??.

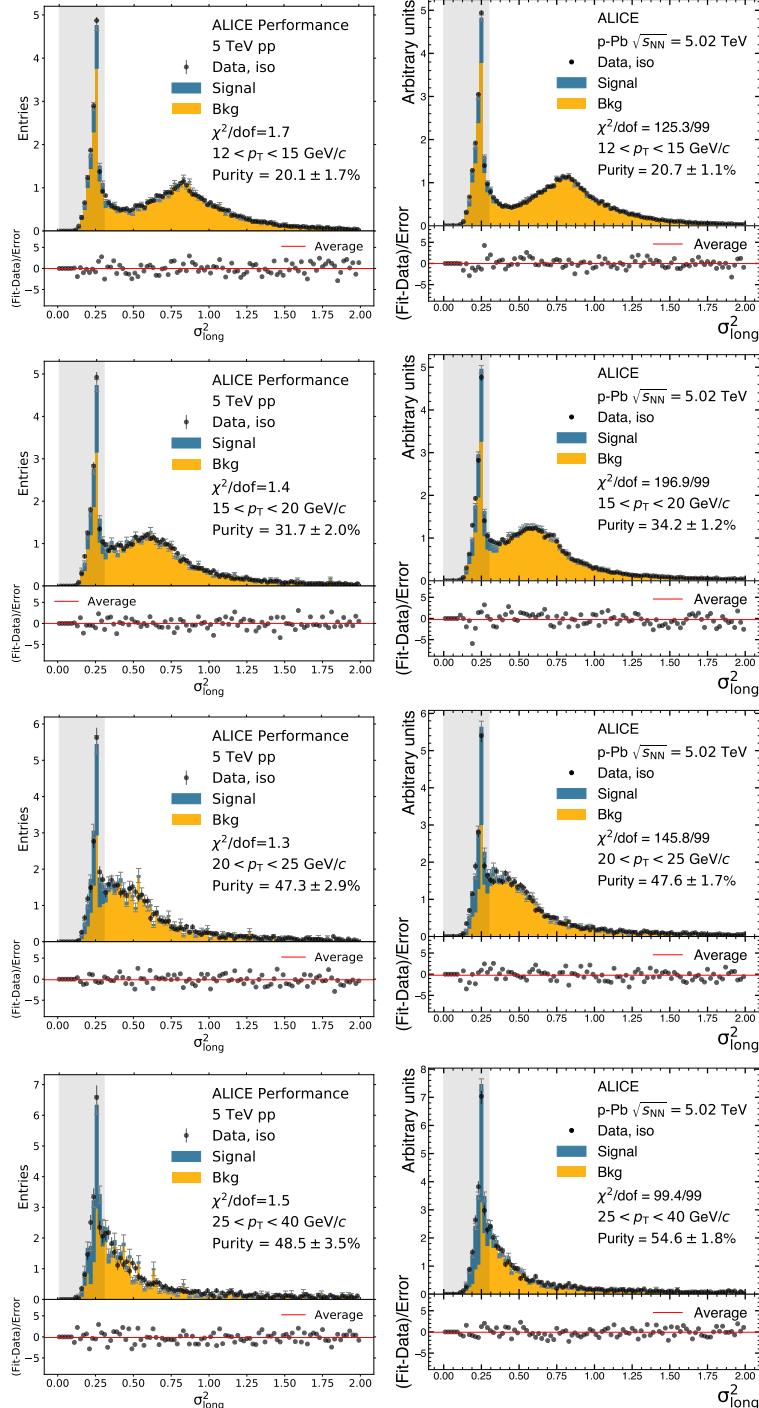


Figure 3.10:  $\sigma_{\text{long}}^2$  distribution of isolated clusters (black) and template fit results for p–Pb data in various  $p_T$  ranges. The stacked histograms (yellow for background, blue for signal) show the predicted counts corresponding to the best fit. The bottom panels show the normalized residuals of the fit, with the statistical uncertainty on the isolated cluster data and the background template added in quadrature. The gray shaded region indicates the signal region for the isolated-photon selection. See text for additional details.

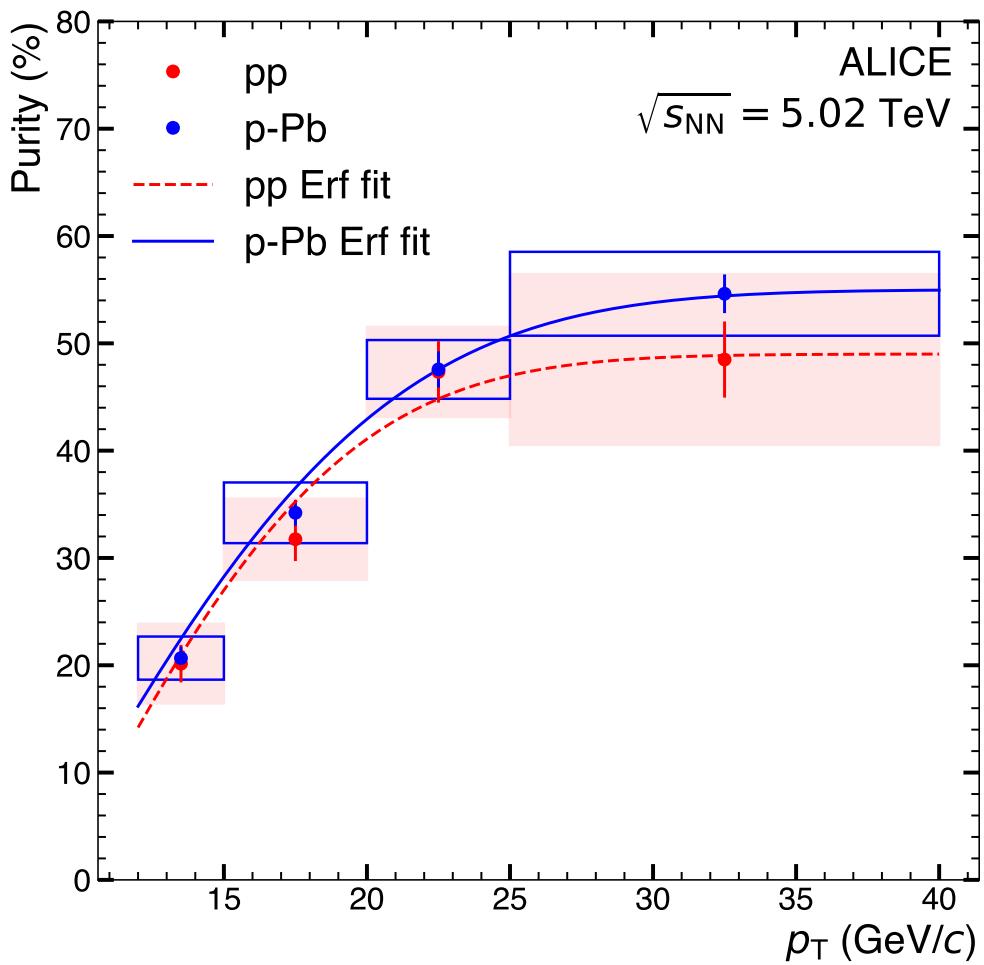


Figure 3.11: Purity of the  $\gamma^{\text{iso}}$  sample as a function of transverse momentum for pp (red) and p–Pb (blue) data. The error bars represent statistical uncertainties only. The red shaded area represents systematic uncertainties in pp, while the blue empty boxes represent systematic uncertainties in p–Pb. The smooth lines correspond to a three-parameter error function fit to the data.

## 3.8 Charged Particle Tracking

The detectors usually responsible for measuring the charged hadrons in this analysis are the ALICE Inner Tracking System (ITS) and Time Projeciton Chamber (TPC). During the 13def and 17q periods, however, the TPC was either not read out or compromised due to space-charge distortions, mentioned in Sec. ???. As a result, this analysis relies on only the ITS for track reconstruction.

ITS-only tracking in ALICE is new for this flavor of analysis. As a result, the charged-particle  $p_T$ -spectrum using ITS-only tracking is compared to the normal TPC+ITS tracking in the data-taking period when the TPC was active and free of space-charge distortions (the low-luminosity 13b data-taking period at 5 TeV p–Pb minimum-bias data) as well as with published ALICE measurements [Acharya:2018qsh] that use the same dataset (a more thorough comparison to published ALICE data is done in Sec. ??).

The combined effect of tracking efficiency, fake rate, and track momentum smearing corrections are calculated using MC simulations, and validated by the comparison to the established hybrid-tracking method in ALICE. The comparison to hybrid tracking also serves as an estimate of the systematic uncertainties due to mis-modeling of the tracking performance.

To study the performance of ITS-only tracking, minimum bias p–Pb events for both data and Monte Carlo. The Monte Carlo simulations used for this section are LHC13b2\_efix\_p1, a DPMJET simulation anchored to LHC13b,c and LHC17l3b, a pp PYTHIA8 simulation anchored to LHC17p. The events are sampled from the LHC13b period and the datasets from [Acharya:2018qsh], as mentioned previously. Only events with the minimum bias trigger (CINT7) that also the vertex and pileup selections described in Section ???. The tracks reconstructed from the ITS (“ITS-only tracks”), are compared with tracks reconstructed from information obtained from both the TPC and ITS (“TPC+ITS tracks”, or “hybrid” tracking). Here, ITS-only tracks are reconstructed in a *stand-alone* way and are not simply the ITS-segment of a ITS+TPC track.

In order to select good-quality tracks emerging from the primary vertex while maintaining a high efficiency, each track is required to satisfy the cuts summarized in Table ???. A set of standard PWG-JE cuts are applied to all tracks, and additional track cuts are applied depending on whether the track is a TPC+ITS track or an ITS-only track.

Table 3.5: Summary of the cuts used in Track Selection.

<b>Common Cuts</b>	
track $\eta$	$ \eta  < 0.8$
track $p_T$	$p_T \geq 0.150 \text{ GeV}/c$
SetMaxDCAToVertexXY	2.4 cm
SetMaxDCAToVertexZ	3.2 cm
SetDCAToVertex2D	TRUE
<b>TPC+ITS Cuts</b>	
SetMinNClustersTPCPtDep	70.+30./20.*x, 20.0
SetMinNClustersTPC	70
SetMaxChi2PerClusterTPC	4
SetMaxChi2PerClusterITS	36
SetMaxFractionSharedTPCCClusters	0.4
SetMaxChi2TPCCConstrainedGlobal	36
SetRequireTPCStandAlone	TRUE
SetRequireTPCRefit	TRUE
SetRequireITSRefit	TRUE
SetRequireSigmaToVertex	FALSE
SetAcceptKinkDaughters	FALSE
<b>ITS-Only Cuts</b>	
SetRequireITSPureStandAlone	TRUE
SetMinNClustersITS	4
SetMaxChi2PerClusterITS	36

### 3.8.1 Efficiency and Fake Rate

The tracking efficiency is calculated as the ratio of the number of reconstructed primary particles<sup>13</sup>,  $N_{\text{prim,rec}}(p_T)$ , to the number of generated primary particles,  $N_{\text{prim,gen}}$ . The truth-to-reconstructed matching is done following the standard ALICE method.

$$\epsilon(p_T^{\text{true}}) = \frac{N_{\text{prim,rec}}(p_T^{\text{true}})}{N_{\text{prim,gen}}(p_T^{\text{true}})}. \quad (3.8)$$

Shown in Eq. ??, the simulated or "truth" transverse momentum,  $p_T^{\text{true}}$ , is used in both the numerator and denominator in order cancel effects of efficiency from bin-migration due to momentum smearing.

The numerator of Equation ?? is restricted for charged particles with generated pseudo-rapidity in the range  $|\eta^{\text{true}}| < 0.8$  and azimuth  $0 < \varphi^{\text{true}} < 2\pi$ . Therefore, the correction factor accounts for both geometrical acceptance, detector inefficiencies, and dead channels.

---

<sup>13</sup>No special tuning of the particle type composition is performed. This typically only matters at low  $p_T$  and enforces a small (percent level) correction to the out-of-the-box results.

Fake tracks are defined as reconstructed tracks that do not match to a truth particle. The fake rate is calculated by taking the ratio of the number of fake tracks to the total number reconstructed tracks. It is parametrized as a function of the reconstructed transverse momentum of the track,  $p_T^{\text{reco}}$ :

$$\text{fakerate}(p_T^{\text{reco}}) = \frac{N_{\text{unmatched}}(p_T^{\text{reco}})}{N_{\text{allreco}}(p_T^{\text{reco}})}. \quad (3.9)$$

Figure ?? shows the efficiency and the fake rates for the TPC+ITS and ITS only tracks. In both cases the efficiency grows with  $p_T$  up to about 1 GeV/c where it dips and it reaches a plateau value with no significant  $p_T$  dependence. The efficiency starts at about 57% for the TPC+ITS tracks and at 70% for ITS-only tracks at 150 MeV and plateaus at 84% and 88% respectively.

The lower efficiency for the TPC+ITS tracks compared to ITS-only tracks is expected since the former requires a matching between ITS and TPC track segments, which has some inefficiency. This study shows that the matching efficiency is high at large  $p_T$  but leads to substantial differences at low  $p_T$ .

The fake rate for the TPC+ITS tracks is less than a percent over the entire range shown. In comparison, the fake rate is larger in the ITS-only tracks. It is below 5% up to 5 GeV/c, and it grows roughly linearly and reaches 15% at 10 GeV/c. The higher fake rate is due to the much lower number of clusters associated with ITS-only tracks (maximum of 6) than to TPC+ITS tracks (minimum of 70).

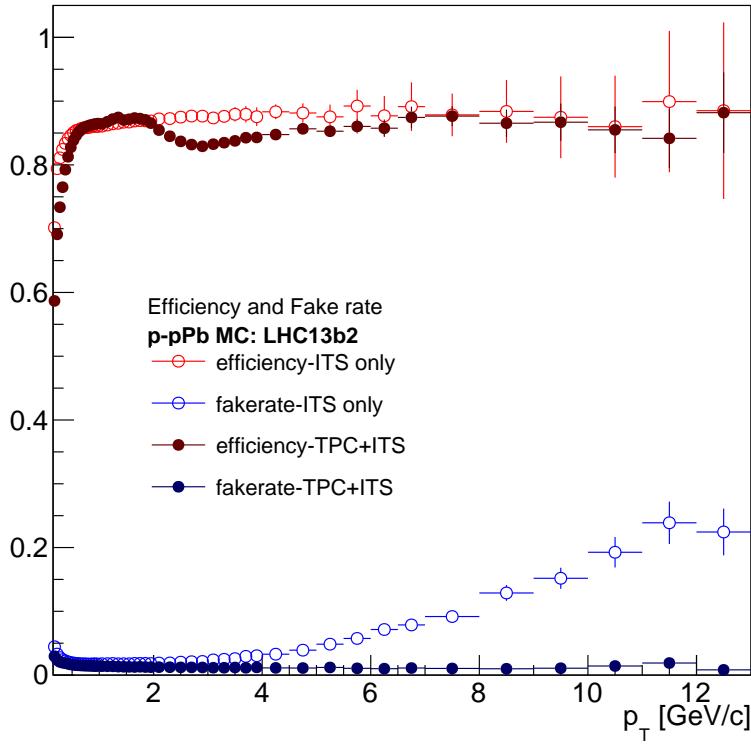


Figure 3.12: Efficiency and fake rate for combined TPC+ITS tracking (filled circles) and ITS-only tracking (open circles) obtained with  $p$ -Pb simulation. The error bars represent statistical uncertainty only.

### 3.8.2 Resolution, Response Matrix and Bin Migration

The main difference between TPC+ITS and ITS-only tracking lies in the much worse momentum resolution of the latter. This is driven by geometry and  $\int B dl$  as the TPC covers up to  $z = 258$  cm but the ITS only to  $z = 48$  cm.

Figure ?? shows the momentum resolution as a function of  $p_T^{\text{true}}$  for TPC+ITS and ITS-only tracks. The momentum resolution of both increases with  $p_T$ ; however, the resolution for TPC+ITS never exceeds a relative 2% below 20 GeV/c, while the ITS-only tracking resolution is about a factor of 7 worse and reaches  $\sim 15\%$  by 10 GeV/c. In both cases the resolution curves have the expected shape: the growth at low momentum is due to multiple-scattering and the linear growth at higher  $p_T$  arises from the number and depth of position measurements, and the track bend at the measurement planes.

The tracking response matrix is defined as the correlation between the reconstructed and generated transverse momentum. This matrix is filled only for reconstructed tracks with a

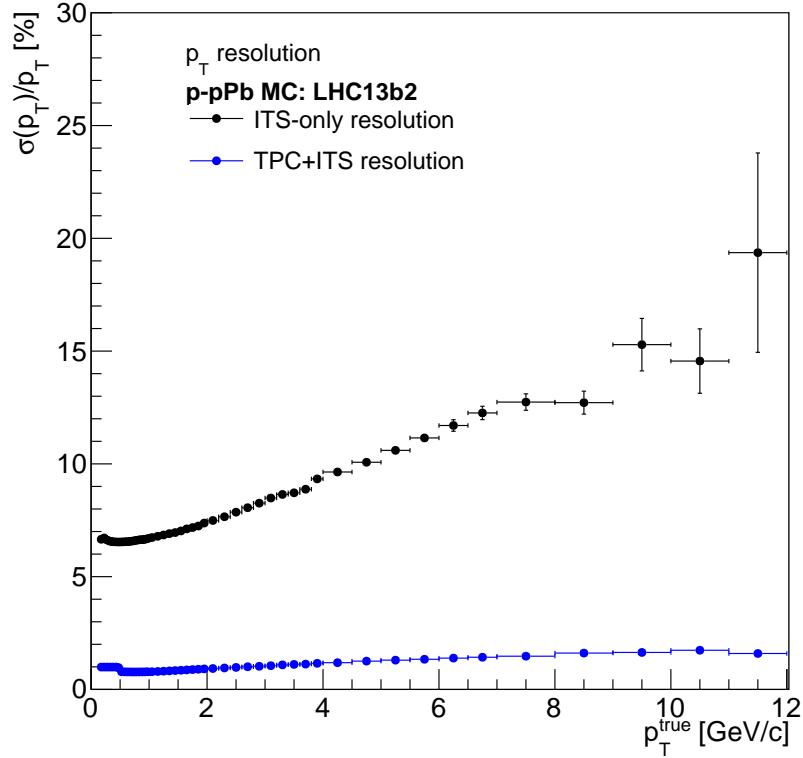


Figure 3.13: The relative  $p_T$  resolution for TPC+ITS tracking and ITS only tracking. The error bar represents statistical uncertainty only.

true match; fake tracks are explicitly excluded. Figure ?? shows the response matrix, its one-dimensional projections, and the ratio of true to reconstructed spectra. The ratio of the true to reconstructed spectra is used to correct for bin migration effects, where a track is placed in the wrong  $p_T$  bin due to the finite momentum resolution of the detector:

$$\text{bin migration}(p_T^{\text{reco}}) = \frac{N_{\text{prim,reco}}(p_T^{\text{true}})}{N_{\text{prim,reco}}(p_T^{\text{reco}})}. \quad (3.10)$$

The effect of momentum smearing on tracks is clearly visible in the projection plots, where the reconstructed spectrum is significantly harder at high  $p_T$ . The ratio of truth to reconstructed  $p_T$  is very close to unity in the TPC+ITS case, as expected. On the other hand, the ratio deviates significantly from unity in the ITS-only case; it reaches 0.9 at 6  $\text{GeV}/c$  and drops quickly, reaching 0.5 at about 13  $\text{GeV}/c$ . The quick drop at high  $p_T$  comes mainly from the linear degradation of the relative momentum resolution combined with the fast drop of the true spectrum to produce large effects in the tails.

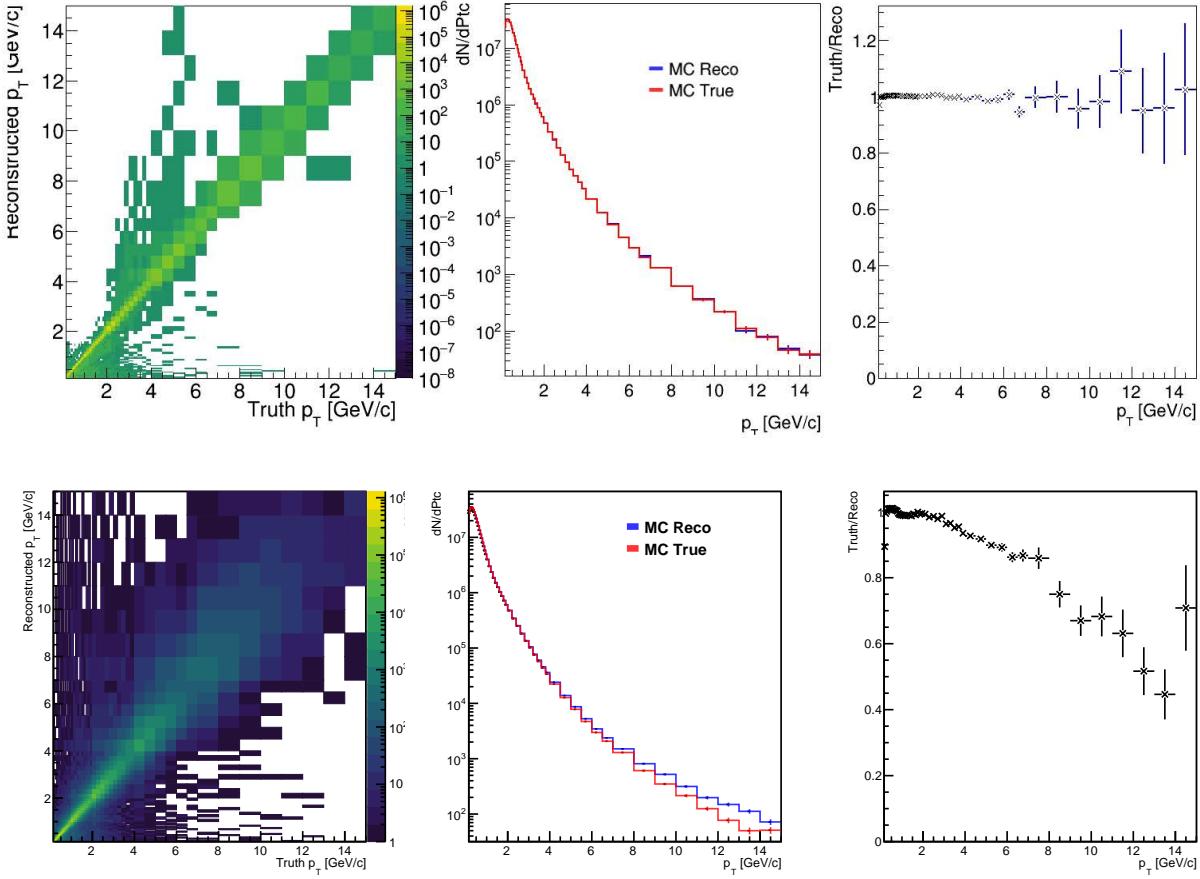


Figure 3.14: Left panel: correlation matrix between true  $p_T$  and reconstructed  $p_T$  of tracks reconstructed with TPC+ITS (upper row) and ITS-only (bottom row). Middle panel: projections of the response matrix into the true and reconstructed  $p_T$ . Right panel: ratio of true to reconstructed spectra. This ratio used as part of the bin-by-bin correction factors.

The bin migration ( $b$ ), along with the tracking efficiency ( $\epsilon$ ) and the fake rate ( $f$ ) are used as the correction factor equation ?? for the charged hadron tracks; all of them are shown together in Figure ???. Based on overlapping plot of pp and p–Pb, we can see that correction factors are similar for tracks from a pp collision or a p–Pb collision. We conclude that the multiplicity in p–Pb is low enough such that it does not affect tracking performance.

$$w = \frac{1}{\epsilon}(1 - f)b \quad (3.11)$$

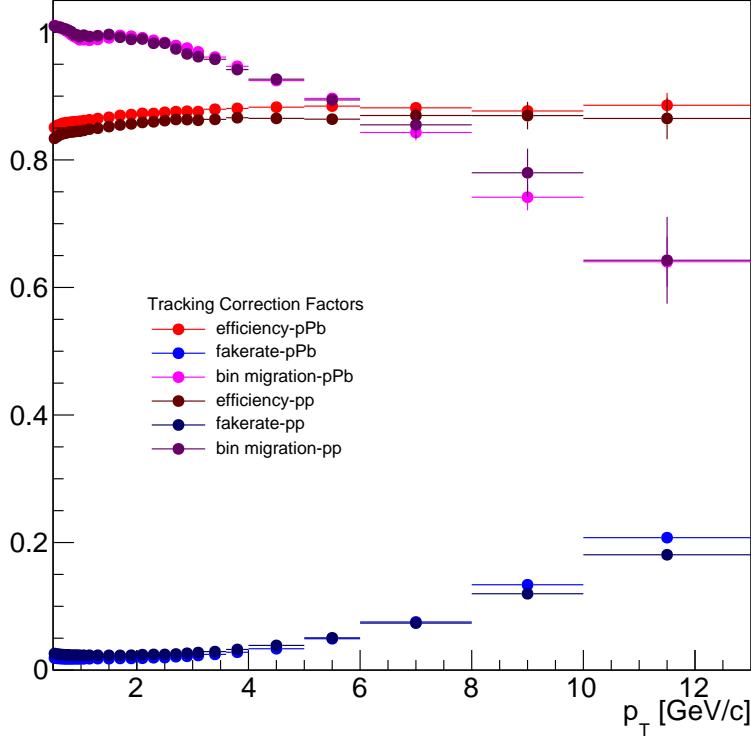


Figure 3.15: The efficiency, fake rate, and momentum smearing correction factors for pp and p–Pb data.

### 3.8.3 Angular Dependence of Tracking Efficiency

The 2D  $\varphi$ - $\eta$  efficiency is calculated in a similar way to the  $p_T$  efficiency described in Equation ??, but instead of being functions of  $p_T^{\text{true}}$ , the efficiency is a function of the true azimuthal angle and pseudorapidity,  $\varphi^{\text{true}}$  and  $\eta^{\text{true}}$  respectively. Only tracks with  $p_T^{\text{true}} > 1 \text{ GeV}/c$  are considered to avoid tracks with sharp bends in the magnetic field that would obscure the impact of dead regions.

Figure ?? shows the resulting efficiency for TPC+ITS and ITS-only tracks. While the TPC+ITS 2D  $\varphi$ - $\eta$  distribution looks uniform, this is not the case for the ITS-only distribution, which has visible dips in the efficiency at various  $\varphi$ . The efficiency is close to unity for most of the phase space covered. There are no big  $\eta$  variations, but there are large  $\varphi$  variations. The efficiency holes at  $\varphi = -0.8$  and  $-0.2$  are very visible and reach values close to zero. These are attributed to ITS-staves that are completely dead. Any variations in  $\varphi$  in the  $\gamma^{\text{iso}}$ -hadron analysis are corrected for using the event mixing technique described in Section ??

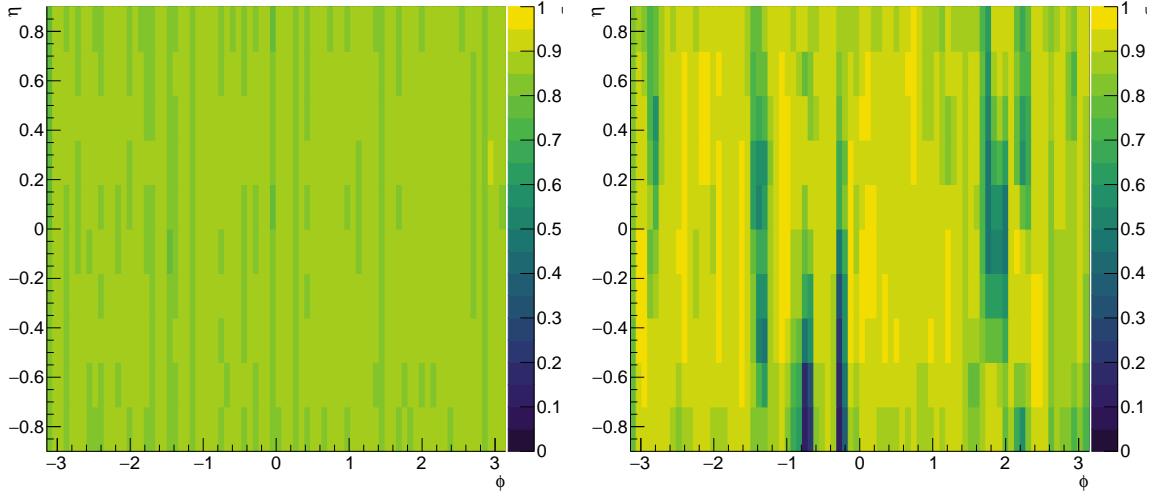


Figure 3.16: Tracking efficiency as a function of  $\varphi^{\text{true}}$  and  $\eta^{\text{true}}$  for TPC+ITS (left) and ITS-only (right) tracks.

### 3.9 Photon Hadron Correlations

In this study, the exact cross section of isolated photons is not the observable, nor is the cross section of charged hadron production in a given event. The quantity of interest is the number partner particles associated with those trigger photons, relating to the leading order picture of back-to-back prompt photon and parton. Therefore, instead of quoting the absolute yield of pairs, it is useful to divide by the number of triggers to obtain the conditional yield (also called per-trigger yield) of associated hadrons. This quantity is typically plotted as a function of the azimuthal angle between the trigger and partner particle,  $\Delta\phi$ :

$$\Delta\phi \equiv \phi^{\text{cluster}} - \phi^{\text{track}} \quad (3.12)$$

The  $\Delta\phi$  distribution of cluster-track pairs, divided by the number of triggers is defined as the trigger normalized correlation function,  $C$ :

$$C \equiv \frac{P}{T} \quad (3.13)$$

Where  $T$  is simply the number of triggers, and  $P$  is the  $\Delta\phi$  distribution of cluster-track pairs. Such distributions often highlight the structure of jets. They typically contain a large, very narrow peak at  $\Delta\phi = 0$ , arising from the autocorrelation of the particles within an jet, and a broader peak centered at  $\Delta\phi = \pi$  arising from the correlation between two jets in the event, shown in Fig. ??.

Enforcing an isolation requirement will heavily suppress the near side peak, however. The near side peak is further removed after the decay photon hadron correlation is subtracted, likely removing the remaining contribution of background jets to the isolated photon triggers ???. This becomes particularly important for the underlying event estimate detailed in ??.

At lowest order in QCD, prompt photons exactly balance the hard scattered parton in the event. For this reason, studies of photon-tagged jets and jet fragments have been dubbed the "golden channel" for studying parton.

Eq. ?? is closely related to Eq. 1.11. The former can be thought of as an ingredient to the latter; Eq. ?? refers to the correlation function over the full range of  $\Delta\phi$ , while Eq. 1.11 refers to the yields, often the result of integrating  $C$  at large  $\Delta\phi$ .

### 3.9.1 Signal Correlation Function

Section ?? introduced isolated prompt photons as the primary photon signal for this measurement. However, as detailed in ??, there is a substantial amount of background contributing to  $\gamma^{\text{iso}}$  candidates and a subsequent contribution to the  $\gamma^{\text{iso}}$ -hadron correlations. In this section, we disentangle the signal of this measurement from what is initially measured.

We directly measure the trigger-normalized correlation of shower signal region isolated clusters and associated hadrons,  $C_{\text{SR}}$ . This quantity will be made up of the true signal correlation function,  $C_S$  – the correlation of isolated prompt photons and associated hadrons from the recoiling parton– as well as the true background correlation,  $C_B$ , predominantly arising from the correlation between decay photons that pass the cluster selection and hadrons. We call this background the correlated background, as it arises from the correlation of background photons and the hadrons in the event. To separate what is initially measured,  $C_{\text{SR}}$ , and the true signal,  $C_S$ , we begin by taking a look at the ingredients of  $C_{\text{SR}}$ . First, we denote the number of trigger clusters in the shower signal region, also called the number of  $\gamma^{\text{iso}}$  candidates,  $T_{\text{SR}}$ . We can write  $T_{\text{SR}}$  as:

$$T_{\text{SR}} = T_S + T_B. \quad (3.14)$$

Here we define  $T_S$  as the number of "true" signal triggers (i.e. the number isolated prompt photon triggers) and  $T_B$  as the background, most prominently decay photons that pass our  $\gamma^{\text{iso}}$  selection. Similarly, the  $\Delta\phi$  distribution of signal region cluster-track pairs,  $P_{\text{SR}}$  can be written as:

$$P_{\text{SR}} = P_S + P_B. \quad (3.15)$$

Now, following the notation of Equation ??, we can write the trigger-normalized correlation functions for shower signal region clusters as:

$$C_{\text{SR}} = \frac{1}{T_{\text{SR}}} P_{\text{SR}}. \quad (3.16)$$

These quantities are directly measured. Similarly, we can write the true signal and true background correlation functions:

$$C_S = \frac{1}{T_S} P_S \quad (3.17)$$

$$C_B = \frac{1}{T_B} P_B \quad (3.18)$$

The goal of this formalism is to write the true signal correlation function ( $C_S$ ) in terms of quantities that are measured –  $C_{SR}$  and the measured purity,  $p$ . To this end, the next step is to write the measured quantity,  $C_{SR}$ , in terms of true signal and true background correlation functions:

$$C_{SR} = \frac{1}{T_{SR}} P_{SR} = \frac{1}{T_{SR}} (P_S + P_B) \quad (3.19)$$

$$C_{SR} = \frac{1}{T_{SR}} (T_S C_S + T_B C_B) \quad (3.20)$$

Where we have substituted equation ?? into Equation ??, followed by using Equations ?? and ?? to substitute out  $P_S$  and  $P_B$ , respectively.  $T_S$  and  $T_B$  are not directly measured, however they can be expressed in terms of measurable quantities. The purity is defined as the fraction of true signal in our  $\gamma^{iso}$ clusters, or in other words,  $p \equiv T_S/T_{SR}$  (see Section ??). Substituting this into equation ?? and solving for  $T_B/T_{SR}$ , one obtains:  $T_B/T_{SR} = 1 - p$ . This a natural result of the definition of purity: If  $p$  is the fraction of the true signal making up the  $\gamma^{iso}$ candidates,  $1 - p$  must be everything else, i.e. background. To summarize:

$$p \equiv T_S/T_{SR} \quad (3.21)$$

$$1 - p = T_B/T_{SR} \quad (3.22)$$

Substituting  $p$  into Equation ??:

$$C_{SR} = p C_S + (1 - p) C_B \quad (3.23)$$

Finally, we can solve for  $C_S$  in terms of  $C_{SR}$  and  $p$ :

$$C_S = \frac{C_{SR} - (1 - p) C_B}{p} \quad (3.24)$$

Equation ?? shows  $C_S$  in terms of the measured quantities,  $C_{SR}$  and the purity. It also includes, however, a term for the background correlation function  $C_B$ , that cannot be determined from  $C_{SR}$  and the purity alone. The correlation function  $C_B$  is described in more detail in section ??.

### 3.9.2 Decay Photon Hadron Correlations

There is a large fraction of  $\gamma^{\text{decay}}$ background within the isolated photon sample, and a subsequent  $\gamma^{\text{decay}}$ -hadron correlation present in the measured correlation function. The correlation between hadrons and background photons within the  $\gamma^{\text{iso}}$ sample, most prominently photons from neutral meson decays, correspond to the  $C_B$  term in in Equation ???. While the equation gives us the scale of this background,  $1 - p$ , it does not offer information on the shape of this  $\Delta\phi$ distribution. To understand the shape, we take advantage of the fact that the most prominent source background photons within the  $\gamma^{\text{iso}}$ population are photons from neutral meson decays. Outside of the  $\gamma^{\text{iso}}$ selection, these photons tend to have more assymmetric shower profiles, and thus larger values of  $\sigma_{\text{long}}^2$ . Therefore, in order to select on clusters arising from decay photons, an inverse shower shape selection is applied which we define as the shower background region, BR:

$$\sigma_{\text{long}}^2(\text{BR}) > 0.4 \quad (3.25)$$

Thus, in order to approximate  $C_B$ , a  $\gamma^{\text{decay}}$ hadron correlation function is measured by taking the correlation of clusters in the shower background regions with associated hadrons in the event. This shower background region correlation function, much like Eq. ?? is defined as  $C_{\text{BR}}$ :

$$C_{\text{BR}} = \frac{1}{T_{\text{BR}}} P_{\text{BR}}. \quad (3.26)$$

With  $T_{\text{BR}}$ as the number of clusters in the shower background region, and  $P_{\text{BR}}$  as the  $\Delta\phi$  distribution of shower background region clusters and hadrons.

Notably, because the underlying physics process that dictates the number and distribution of correlated pairs is independent of the opening angle of the neutral-meson decay, which is what drives the shower-shape, the approximation  $C_{\text{BR}} \approx C_B$  can be made.<sup>14</sup> In other words, the shower background region correlation function is a good approximation of the correlated background that contributes to the shower signal region correlation function,  $C_{\text{SR}}$ . Therefore, Eq. ?? can be re-written:

$$C_S = \frac{C_{\text{SR}} - (1 - p)C_{\text{BR}}}{p} \quad (3.27)$$

As a result, the true signal correlation function,  $C_S$ , is finally written in terms of measurable quantities: The shower signal region correlation function,  $C_{\text{SR}}$ , the purity,  $p$ , and the newly defined shower background region correlation function (or  $\gamma^{\text{decay}}$ correlation),  $C_{\text{BR}}$ .

Thus, hadrons are correlated with clusters in the shower background region to directly measure the  $\gamma^{\text{decay}}$ -hadron correlation functions. This  $\gamma^{\text{decay}}$ -hadron correlation function is

---

<sup>14</sup>Here again we take advantage of trigger normalized quantities. The number of isolated photons in the shower background region vastly outnumber the number of isolated photons within the shower signal region. By focusing on the associated yield of hadrons per each photon, which is not correlated with  $\sigma_{\text{long}}^2$ , this very useful approximation can be made.

then subtracted from the the shower signal region correlation function to obtain the signal correlation. This scaling, however, must be done carefully due to the  $p_T$  dependance of the purity. The next few sections describe corrections to the correlation functions. Unless explicitly stated, these corrections (purity weighting, acceptance corrections, and charged particle tracking corrections) are applied to both the  $C_{SR}$  and  $C_{BR}$  before the two are finally subtracted in Sec. ??

### 3.9.3 Photon Purity Weighting

Equation ?? shows the background correlation function,  $C_B$ , approximated by the  $\gamma^{\text{decay}}$  correlation function, and scaled by  $1 - p$  according to its relative contribution measured correlation function. The purity is a  $p_T$ -dependent quantity, rising quickly with  $p_T$ , correlations using a low- $p_T$  cluster have a higher fraction of background than high- $p_T$  clusters (see Figure ??). As a result, measuring the  $\gamma^{\text{decay}}$ -hadron correlation function for all clusters and scaling by the mean of the purity would lead to an underestimation of the background at low  $p_T^{\text{cluster}}$ , and an overestimation at high  $p_T^{\text{cluster}}$ . This will have a non-trivial effect on the corresponding  $z_T$  bins, which include clusters with a wide range of  $p_T^{\text{cluster}}$ . Additionally, Equation ?? includes an overall scale of  $1/p$  in order to obtain the correct conditional yield of hadrons after the subtraction in the numerator, and will yield similar complications if applied correlations using clusters over the full  $p_T^{\text{cluster}}$  range.

In order to avoid these complications, clusters are weighted by the purity corresponding to their exact  $p_T^{\text{cluster}}$  when constructing the correlation functions. In order to capture the quickly rising behavior of the purity at low  $p_T^{\text{cluster}}$ , the purity is fit to a 3-parameter error function. This fit to the purity is shown in Figure ?? The  $p_T^{\text{cluster}}$  becomes an input to this function, and precise purity weighting is applied precisely to each cluster.

According to Equation ??, the overall purity weights will be  $1/p$  and  $(1 - p)/p$  for shower signal and shower background region clusters, respectively.

### 3.9.4 Track Efficiency, Fake Rate, and Bin Migration Weights

In order to correct for the tracking efficiency, fake rate, and bin-migration we apply a track-by-track weighting according to:

$$w_{\text{tracking}}(p_T^{\text{track}}) = \frac{1}{\epsilon} \times (1 - f) \times b. \quad (3.28)$$

Here  $\epsilon$  is the track efficiency,  $f$  is the fake rate, and  $b$  is the bin-to-bin migration correction. These are described in Section ???. The corrections are estimated independently for pp and p-Pb data although the performance is very similar. The weights are applied to the measured charged hadrons as the correlation functions are being constructed, analogous to how the purity weighting is applied to the photons in the correlations functions.

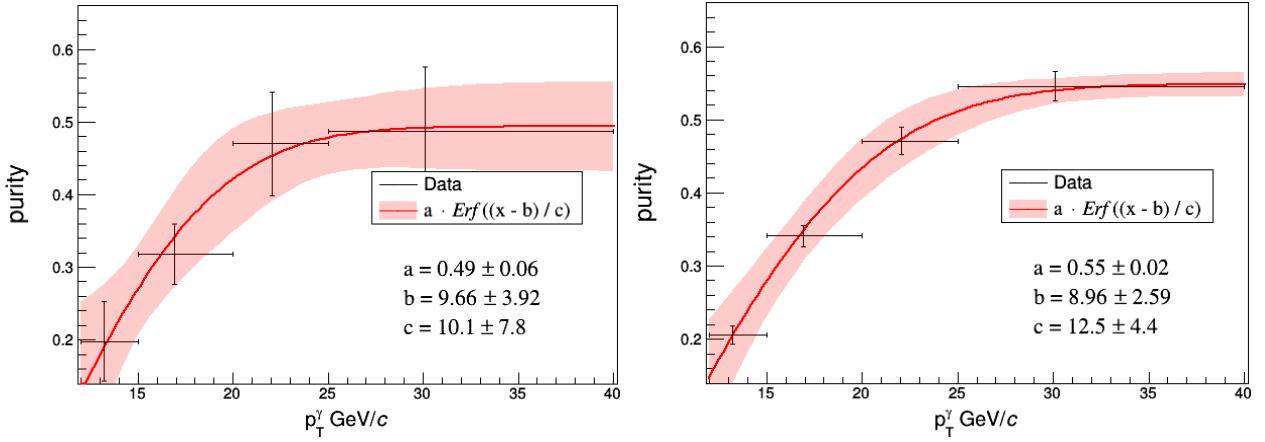


Figure 3.17: A 3-parameter error function is fit to the purity values measured in pp (left) and p-Pb (right) data. The width of the band represents the uncertainty on the fit.

### 3.9.5 Pair-Acceptance Correction with Event Mixing

Initially, two particle correlations consist of a combination of true physical correlations and detector effects. The detector effects result from inefficiencies and limited acceptance in the detectors. The study of trigger normalized yields of associated hadrons eliminates the need to correct for isolated cluster efficiency, and Section ?? outlines the charged tracking efficiency correction.

The remaining detector effect on the correlations are pair-acceptance effects: correlations are constructed through pairs of clusters and charged tracks. This results in a convolution of acceptance effects from the limited acceptance of EMCAL and ITS, which we call pair acceptance effects. These effects are corrected for by using the event mixing technique. Event mixing is a data driven approach to correcting for detector acceptance effects <sup>15</sup>. By constructing observables with particles from different events, we remove true physics correlations from the correlation functions, isolating detector effects from limited acceptance in  $\eta$  and detector inhomogeneity in  $\eta$  and  $\varphi$ .

Cluster-track pairs in same event correlation functions obviously share the properties of the event, and such properties often effect detector response. In order to make the mixed-event correlations as analogous as possible to the same-event correlation functions, events that are as similar as possible with respect to these event properties are used for event mixing. The two most important event properties for this measurement are multiplicity and the z-coordinate of the reconstructed primary vertex (i.e. the position of the primary interaction vertex along the beam direction)<sup>16</sup>.

<sup>15</sup>Event mixing is also used for estimating combinatorial background.

<sup>16</sup>In Lead-Lead collisions, the event-plane angle which determines the anisotropic distribution of final state particles, or  $v_2$ , is also one of the most important event properties to match in Event mixing. However,

The goal of event mixing is to isolate detector effects by completely removing true physics correlations. To this end,  $\gamma$ -triggered events are not mixed with other  $\gamma$ -triggered events. Triggering on a high  $p_T$  photon will result in an enhancement of away-side hadrons due to the recoiling jet. Due to the limited acceptance of the EMCAL, this enhancement will be concentrated in a small area of the ITS, approximately  $180^\circ$  opposite the EMCAL trigger. Mixing only with triggered events will result in a "pseudo" recoiling jet signal that would be suppress the true signal in the same-event correlation function when the event mixing correction is applied. Instead,  $\gamma$ -triggered events are mixed with minimum bias events to avoid this bias, and to sample the full acceptance of the ITS properly.

For this analysis, depending on the  $z_T$  bin, each  $\gamma$ -triggered event was mixed with up to 300 minimum-bias events.

Traditionally, events are often placed into bins of multiplicity (V0 amplitude, sum of V0A and V0C) and primary vertex  $z$ -position, and then mixed within these bins. This has the advantage of conceptual simplicity, but is not very efficient and requires a large amount of cpu time. Instead of bins, the mixing in this analysis is carried out by using a stable matching algorithm [**GaleShapley:1962amm**]. Generally, this algorithm is used to pair two sets of populations, where members from both populations have a well defined and ordered preference list. Here, the two populations are  $\gamma$ -triggered and minimum bias events. The use of this algorithm avoids the need for binning in multiplicity and primary vertex, and is much faster than the standard binning method.

The stable matching algorithm first creates a preference list made up of all other events based on how close events are in multiplicity and  $z$ -vertex. After each event has a preference list, the algorithm loops over all events, with a nested loop that iterates over each event's preference list. The algorithm then pairs pairs the current event to the first unpaired event on that list. As the loop iterates, if an event towards the end of the main loop has an already-paired event high on its preference list, the algorithm loops through the already-paired event's preference list and decides if the paired event should stay paired to its current match, or switch to the new event. If the latter is chosen, the previously matched event is "unpaired" and added back into the loop. A stable state is met when all paired events have a match that is higher on their preference list than any remaining unpaired events in the loop. Such a stable state is guaranteed to eventually be met according to Ref. [**GaleShapley:1962amm**].

The pseudo code below follows this description, using  $\gamma$  to denote a  $\gamma$ -triggered event, and  $M_B$  to denote a minimum-bias event. The *unrequested* state refers to a  $M_B$  event on a  $\gamma$ -event's preference list that has not yet been requested for pairing.

```
procedure GALESHAPLEYPAIRING
  while  $\exists$  free  $\gamma$  with an unrequested  $M_B$  on  $\gamma$ 's list do
     $M_B$  = first unrequested MinBias Event on  $\gamma$ 's list.
```

---

because this measurement focuses on smaller systems, pp and p–Pb, this effect can be neglected

```

if  $M_B$  is free then
     $(\gamma, M_B)$  become paired
else some pair  $(\gamma', M_B)$  exists
    if  $M_B$  prefers  $\gamma$  to  $\gamma'$  then
         $\gamma'$  becomes free
         $(\gamma, M_B)$  become paired
    else
         $(\gamma', M_B)$  remain paired
    end if
end if
end while
end procedure

```

The difference distributions for Z-vertex and multiplicity between a  $\gamma$ -triggered and minimum bias events in p–Pb data are shown in Figure ???. The resulting distributions show a sharp peak that is below  $\Delta z < 0.5$  cm and also a long tail. Less than 6 % of the distribution lies beyond  $\Delta z > 2$  cm. The multiplicity difference, however, does not have as sharp a peak near  $\Delta \text{Multiplicity} = 0$ . About 20% of pairs have a multiplicity difference above 40, and cuts at  $\Delta V_z > 2$  cm and  $\Delta \text{Multiplicity} > 40$  were applied to pairs before calculating correlation functions as a precaution. Skimming p–Pb events with particularly high multiplicities before pairing had a similar effect on the tail of the distribution. However, both skimming events before pairing and applying the previously mentioned cuts after the pairing process had no observable effect on the mixed-event correlation.

Figure ?? shows the V0 multiplicity distributions for pp and p–Pb data in  $\gamma$ -triggered events. This shows that a multiplicity matching requirement of  $\Delta \text{Multiplicity} < 40$  is indeed very tight.

Ideally, the mixed event distribution should be flat in  $\Delta\varphi$  and have a trapezoidal shape in  $\Delta\eta$ , because the limited acceptance in  $\eta$  increases the likelihood to reconstruct pairs with a small  $\Delta\eta$  (i.e., due to the convolution of two uniformly distributed functions). However, the use of ITS-only tracks and holes in the ITS acceptance result in deviations from a flat distribution in  $\Delta\varphi$ .

The correlation function corrected by pair-acceptance effects is then given by:

$$C(\Delta\varphi, \Delta\eta) = \frac{S(\Delta\varphi, \Delta\eta)}{M(\Delta\varphi, \Delta\eta)}, \quad (3.29)$$

where  $S(\Delta\varphi, \Delta\eta)$  is the same-event correlation, and  $M(\Delta\varphi, \Delta\eta)$  is the mixed-event correlation.  $S(\Delta\varphi, \Delta\eta)$  is given by:

$$S(\Delta\varphi, \Delta\eta) = \frac{1}{N_{\text{trig}}} \frac{d^2 N_{\text{same}}}{d\Delta\varphi d\Delta\eta} \quad (3.30)$$

With  $N_{\text{trig}}$  as the number of trigger particles and  $N_{\text{same}}$  as the number of same event cluster-track pairs and  $d^2 N_{\text{same}} / d\Delta\varphi d\Delta\eta$  is found by pairing trigger particles with tracks

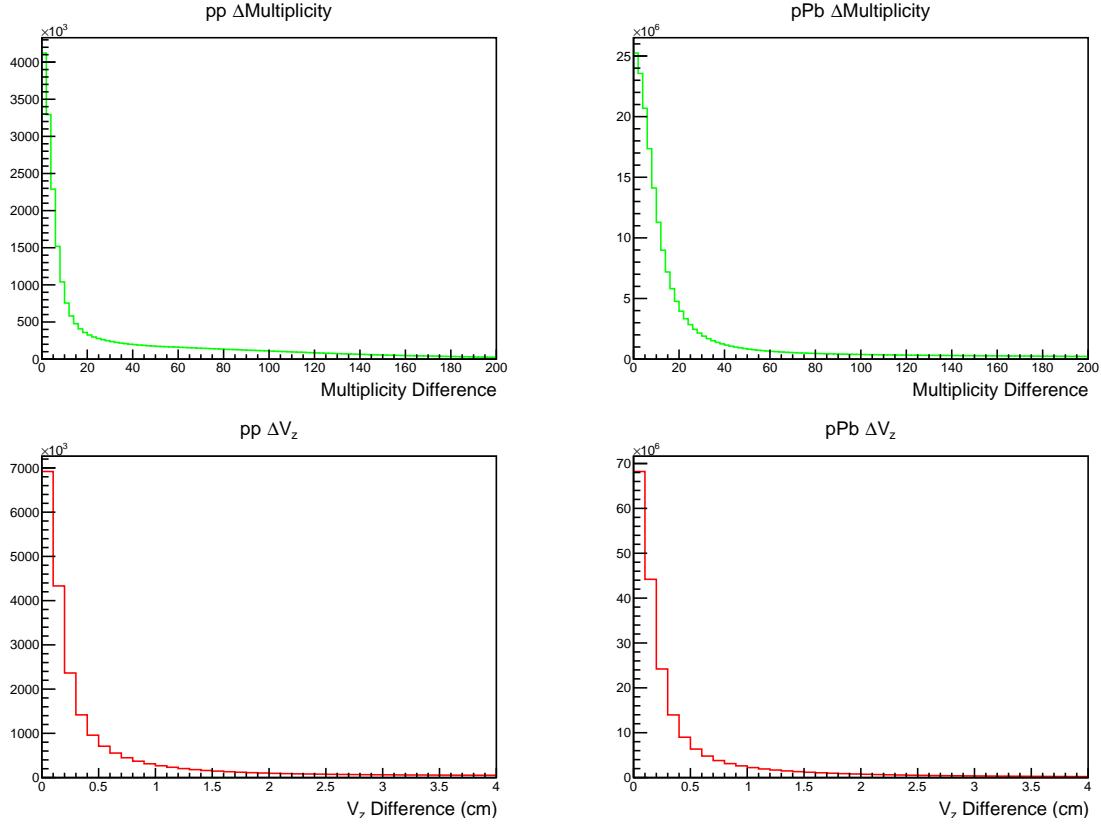


Figure 3.18: Difference in V0 multiplicity (upper row) and longitudinal vertex position (bottom row) between paired events in pp (left column) and p–Pb right. The pairing algorithm results in sharp peak near zero for these difference distributions, particularly in the longitudinal vertex difference. As described in the text, in the correlation analysis we apply a further selection to cut the large tails observed in these distributions.

from the same event. The mixed-event distribution,  $M(\Delta\varphi, \Delta\eta)$ , is given by

$$M(\Delta\varphi, \Delta\eta) = \alpha \frac{d^2 N_{\text{mixed}}}{d\Delta\varphi d\Delta\eta}. \quad (3.31)$$

Where  $\alpha$  is the normalization constant that sets the maximum value of the mixed event correlation to 1, and  $N_{\text{mixed}}$  is the number of mixed event cluster-track pairs. The term  $d^2 N_{\text{mixed}} / d\Delta\varphi d\Delta\eta$  is obtained by pairing trigger particles from  $\gamma$ -triggered events with tracks from minimum bias events matched in z-vertex and multiplicity.

Same event correlation functions are divided by the mixed event correlation function within the same  $z_T$  bins, shown for a single  $z_T$  bin in Figure ???. This procedure is carried out identically for clusters in the signal and background shower-shape regions.

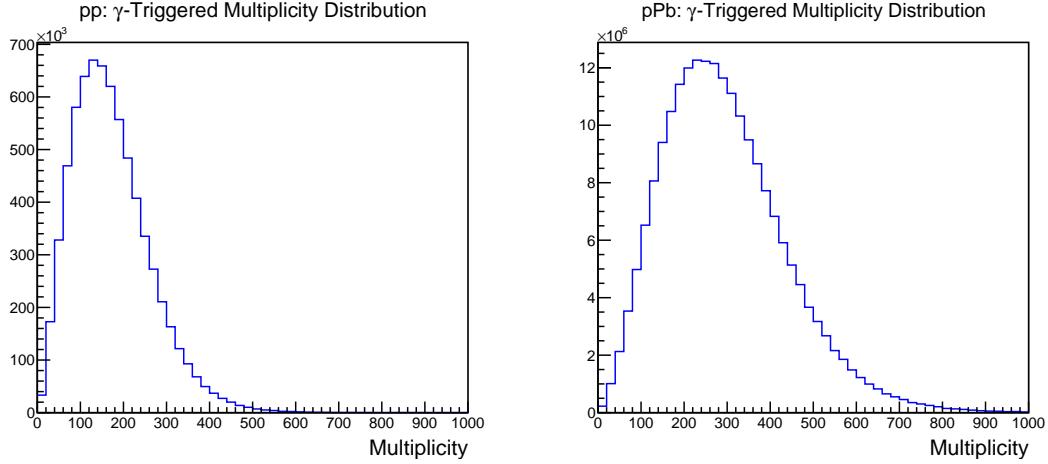


Figure 3.19: V0 multiplicity distribution, i.e. the sum of V0A and V0C amplitudes , in pp (left) and p–Pb (right) gamma-triggered data.

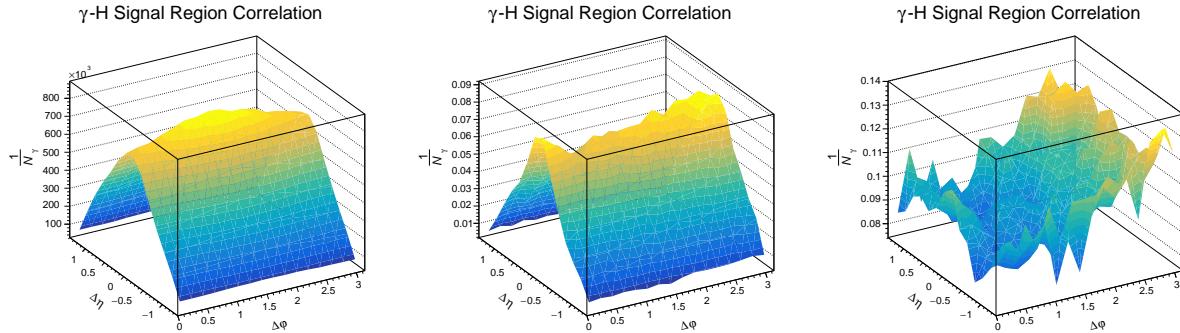


Figure 3.20: **Left** Mixed Event correlation for a single  $z_T$  bin for gamma-triggered, signal region clusters and hadrons from minimum bias events. **Middle** 2D Correlation for signal region clusters and hadrons from the same events. **Right** Signal region correlation function corrected for detector acceptance effects.

The triangular shape in  $\Delta\eta$  is due to the limited acceptance of the ITS and EMCal in pseudorapidity. A useful analogy is the integration of two intersecting square waves that result in a clear triangular signal. The round shape in  $\Delta\phi$  is more subtle, however. It is due to the inefficiencies and holes in the ITS, i.e. due to imperfections in the tracking system. This is further discussed in section ??.

### 3.9.6 Fully Corrected $C_{\text{SR}}$ and $C_{\text{BR}}$

The fully-corrected  $C_{\text{SR}}$  and  $C_{\text{BR}}$  correlations are shown in Figures ?? and ???. The shown  $\gamma^{\text{iso}}\text{--hadron}$  correlations are the difference between the scaled- $C_{\text{SR}}$  and the scaled- $C_{\text{BR}}$ , which are shown in blue and red respectively. While the statistical precision of both  $C_{\text{SR}}$  and  $C_{\text{BR}}$  is high in all  $z_T$  bins and datasets, this gets diluted in the subtraction. That is, the low-purity leads to the subtraction of two comparable numbers, which results in a large statistical uncertainty.

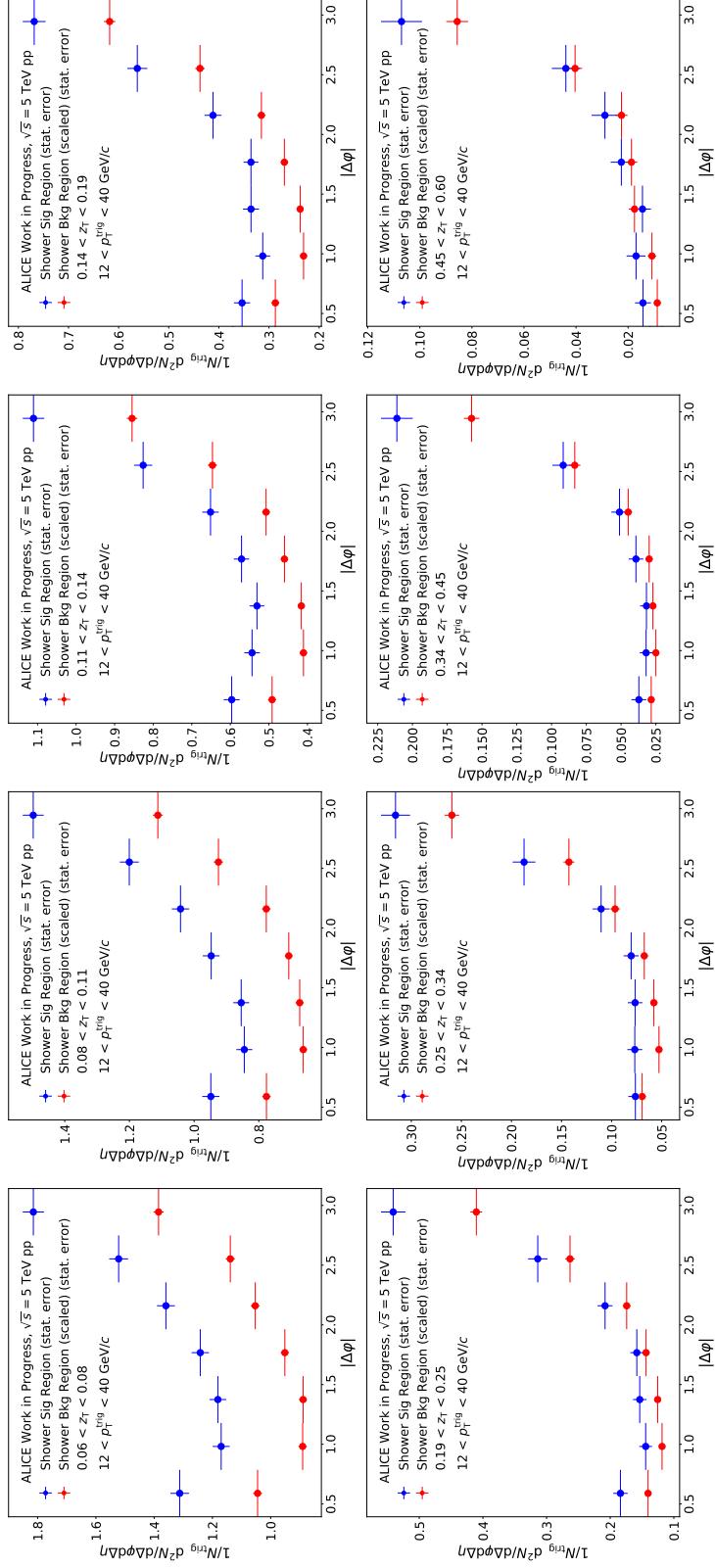


Figure 3.21: Signal-region correlation  $C_{\text{SR}}$  (blue) and background-region correlation  $C_{\text{BR}}$  (red) in pp collisions for various  $z_T$  intervals. The error bars represent statistical uncertainties only.

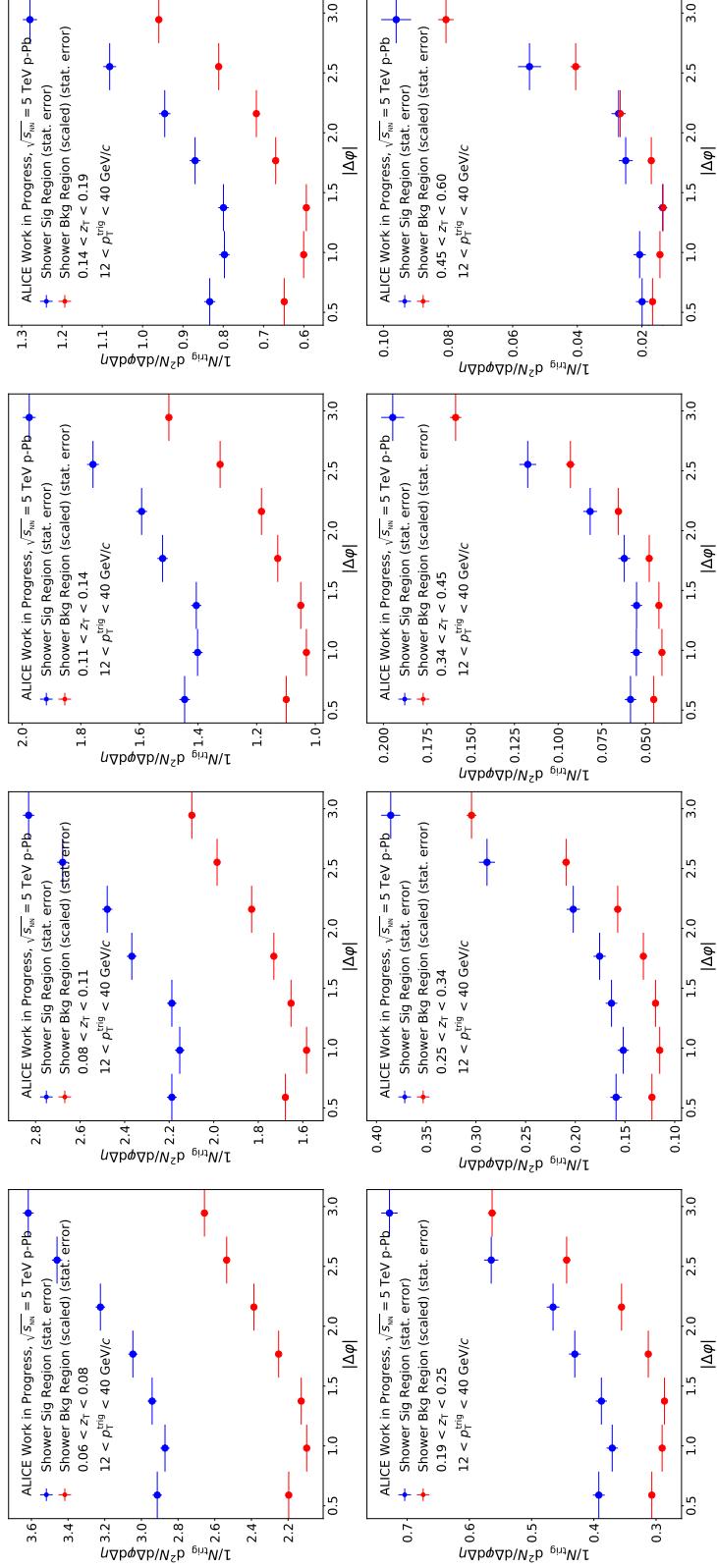


Figure 3.22: Signal-region correlation  $C_{SR}$  (blue) and background-region correlation  $C_{BR}$  (red) in p–Pb collisions for various  $z_T$  intervals. The error bars represent statistical uncertainties only.

### 3.9.7 Underlying Event Estimation

As mentioned in Section ??, the underlying event corresponds to all the activity in the event that does not directly relate to the hard scattering in the initial collision. In particular, low  $p_T$  hadrons that are not physically correlated with hadronization of the scattered parton make up a large portion of the underlying event. For this reason, this background is also called the *the uncorrelated background*. The first step in this measurement, however, is to take the correlation between the trigger photon and all charged hadrons in the event. There is therefore a large contribution to the measured correlation functions from hadrons in underlying event that must be subtracted.

To subtract this underlying event contribution, we use the zero-yield at minimum method. As the name suggests, this method assumes zero true signal at the minimum of the correlation function [include reference that the referee recently asked for]. The distribution of hadrons arising from the underlying event in pp and in p-Pb collisions is conveniently isotropic in  $\varphi$ . Therefore, the distribution of  $\Delta\phi$ , the angle between the trigger photon and the charged hadrons from the underlying event, will be flat in  $\Delta\phi$ . As result, the contribution from the underlying event can be estimated as a flat pedestal in  $\Delta\phi$ .

In dihadron and dijet measurements, the minimum occurs most often around  $\Delta\phi = \pi/2$ . This is because a struck parton is kinematically unlikely to scatter at roughly  $90^\circ$  from another parton in the initial collision. An illustration of this method for *dihadron* measurements is shown in Figure ??.

The isotropic nature of hadrons in the underlying event tells us the shape of the background is a pedestal, while the ZYAM assumption indicates the overall height of the pedestal. Once the shape and magnitude of the underlying event contribution is understood, this background can be subtracted from the correlation functions.

In this analysis, however, we use a modified version of the ZYAM method. One of the most prominent features of Figure ?? is a near side peak made up of the autocorrelation of charged hadrons within a jet. The triggers in this analysis, however, are isolated prompt photons that have little to no surrounding hadronic activity. Therefore, the near side jet peak shown in Figure ?? is completely absent in isolated prompt photon-hadron correlations. While some surrounding hadronic activity can still be present, either due to fragmentation photons from a jet that are surrounded by the jet constituents, or by decay photons within a jet, the latter is subtracted away by subtracting the decay-photon hadron correlation.

As a result, the minimum of the  $\gamma^{\text{iso}}$ -hadron correlation function spans a much larger region in  $\Delta\phi$  than in the dihadron case. We modify the standard ZYAM method by take the average value of the correlation function in the region of  $0.4 < \Delta\phi < \pi/2$  to estimate the underlying event pedestal, rather than a much narrower region centered around  $\Delta\phi \approx \pi/2$ . A minimum  $\Delta\phi$  of 0.4 is used in order to avoid the region of the isolation cone used in the photon isolation calculation – avoiding an artificially low pedestal estimate. The maximum of  $\pi/2$  is used to avoid the tail of the away side jet peak.

This larger region in  $\Delta\phi$  has the advantage of higher statistical precision in the underlying event estimate. The magnitude of the underlying event is estimated for both SR and

BR correlation functions, and subtracted as constant in  $\Delta\phi$ . In order to show the effect of pedestal subtraction on the correlation functions in pp and p–Pb data, the correlation functions in both systems are overlayed in Figure ?? . By construction, the points at small  $\Delta\varphi$  are consistent with zero as demonstrated by the dark grey bands. Additionally, the figures demonstrate the larger underlying event in p–Pb data, as well as the agreement in away side yields in the two systems after pedestal subtraction. This also shows visually the fraction of signal to background, particularly at low  $z_T$  in p–Pb collisions.

After the pedestal subtraction, one can begin to see the similarities between the correlation functions in pp and p–Pb.

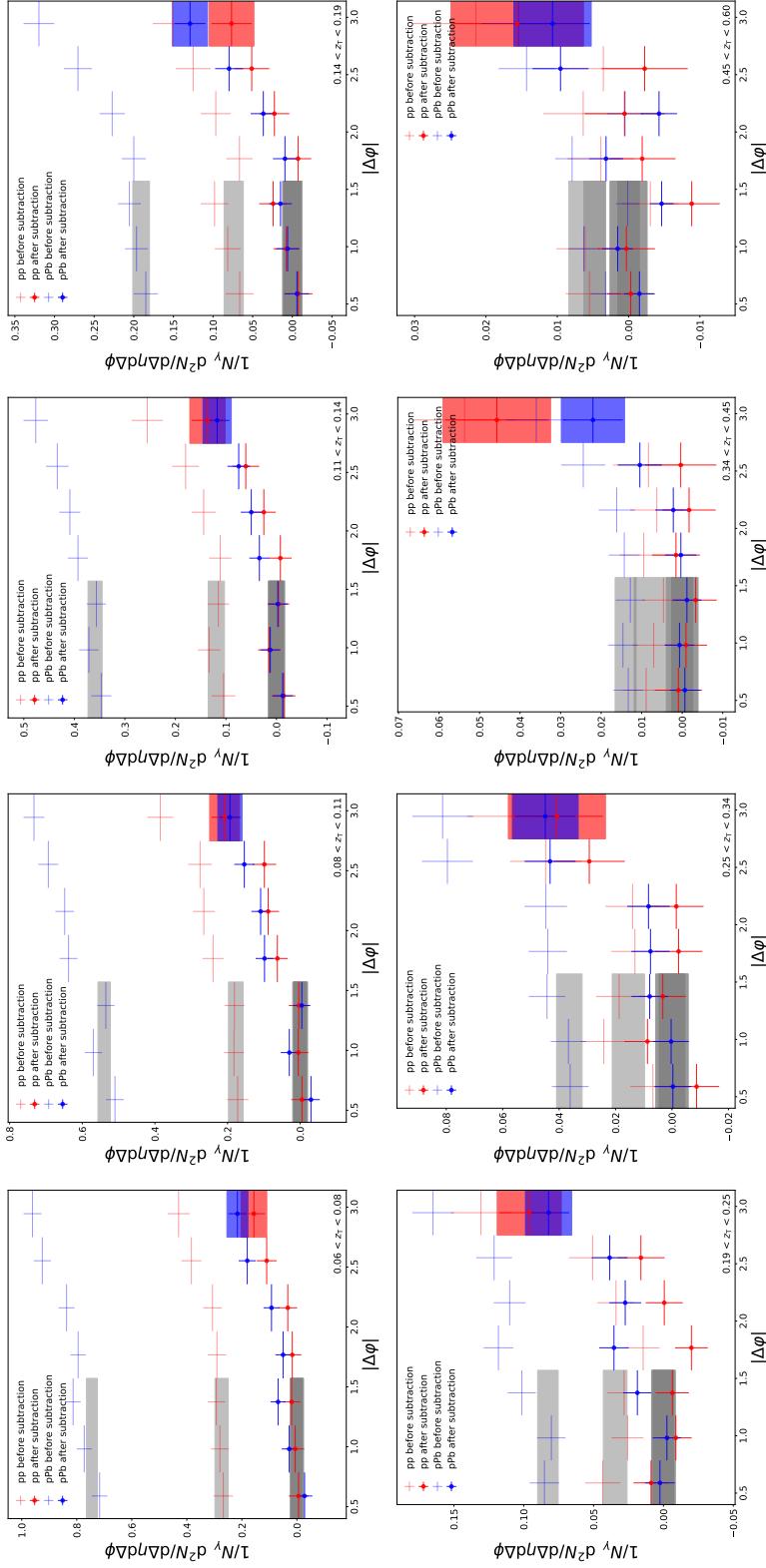


Figure 3.23: Correlation functions in pp (red) and p-Pb (blue) before and after pedestal subtraction. The light grey bands represent the ZYAM estimates, while the dark gray bands represent the near side average after subtraction. The colored bands represent the away side average after pedestal subtraction

# Chapter 4

## Checks and Systematics

### 4.1 Event Mixing

Event mixing is used in the analysis in order to correct for the pair-acceptance of photon-hadron pairs. In other words, the convolution of the EMCAL acceptance with the  $\eta - \varphi$  dependant efficiency of the standalone ITS tracking is non-trivial. It can be corrected for, however, using the data driven method outlined in Sec. ??.

Because this correction is data-driven, a systematic on due to any mis-modeling of the detector performance is not applied. In theory, there should be a statistical uncertainty on the event mixing that would propagate as an uncertainty on the final analysis. For this reason, a very large number of MB events are mixed with gamma triggered events such that the statistical uncertainty on the mixed event correlations is negligible, and thus is not propagated into the analysis. Furthermore, for the high  $z_T$ bins in the correlation and fragmentation function analysis, MB events where at least one track with a  $p_{\text{T}}$ of 14GeV/care chosen to be in an event pool (this  $p_{\text{T}}$ requirement roughly corresponds to the  $z_T$ bin for a photon at 40GeV/c). This provides very good statistics, even in the normally statistics limited bins at higher  $z_T$ .

Additionally, because the mixing is used as an acceptance correction, the normalization is set such that the highest value of the mixed-event correlation is set to 1.0, and no uncertainty is applied to this normalization.

Two cross checks are done on the mixing, however. The first is to validate the overall shape of the mixing, and to check that the origin of the non-uniformity in  $\Delta\varphi$ is in fact due to the  $\varphi$ -dependant efficiency of the ITS-only tracking. The second check applies the slower but more standard binned-mixing technique as a check on the novel use of the gale-shapley algorithm in pairing events.

#### 4.1.1 Toy Monte Carlo for Validating Event Mixing

While the ALICE EMCAL has a limited acceptance in  $\varphi$ , the ITS has full azimuthal coverage.

Given a perfect detector with limited acceptance, the mixed-event correlations functions is expected to be flat is  $\Delta\eta$ . This is because the minimum bias tracks in detected by an idealized version of the ITS will have a flat distribution in  $\varphi$ . While the EMCal has a limited acceptance in  $\varphi$ , the  $\Delta\phi$  distribution will be flat, as any trigger cluster will be correlated with tracks that are homogeneously spaced in  $\varphi$ ; minimum bias tracks are just as likely to be near a trigger photon as they are to be opposite the trigger photon. This is show in Figure ?? (Left), where trigger photons are constrained in  $\varphi$  and  $\eta$  according to the EMCal acceptance, while track are limited only in  $\eta$  to correspond to a perfect ITS with full azimuthal coverage.

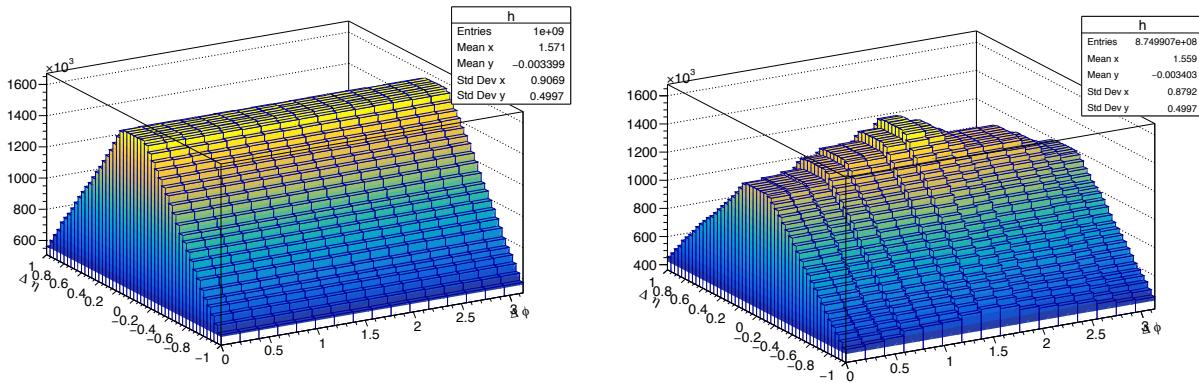


Figure 4.1: Toy Monte Carlo mixed-event correlation function. (Left) Photons are produced randomly within  $0^\circ \leq \varphi \leq 107^\circ$  and  $-0.67 \leq \eta \leq 0.67$  to roughly match the ALICE EMCal acceptance. Tracks are produced randomly for all values of  $\varphi$  (between 0 and  $360^\circ$ ) and  $-0.8 < \text{eta} < 0.8$  to match the acceptance of an ideal ITS. (Right) Photons are generated in the same range as the left panel, however tracks have the additional constraint of excluding the range  $25^\circ < \varphi < 45^\circ$  and  $245^\circ < \varphi < 270^\circ$  in order to roughly approximate holes in the charged particle tracking efficiency, relative to the EMCal acceptance.

However, this flat distribution cannot be assumed, as previously seen in Figure ?? and now in Figure ?? (right). Dead areas or spots with a lower overall tracking efficiency accumulate over years of use at the LHC, and if large enough they can effect the pair-efficiency and therefore the mixed event correlations. This is shown in Figure ?? . Both track and trigger photon  $\Delta\eta$  and  $\Delta\phi$  are constrained according to their respective detector acceptances, however additional holes in the  $\varphi$  distribution of tracks are placed according to 2D tracking efficiency plot, Fig. ??, Sec. ??.

#### 4.1.2 Binned Event Mixing

Typically, events are binned according the their event information, centrality, z-vertex, flow (in PbPb), and then added to larger pools of events according to these bins. While it depends

on the analysis, very common bin sizes are 5%, 2cm for centrality and z-vertex. The gale-shapley pairing algorithm is novel for this use case, and thus a check was carried out in which the mixed event correlation function was calculated using the standard binning method for event mixing.

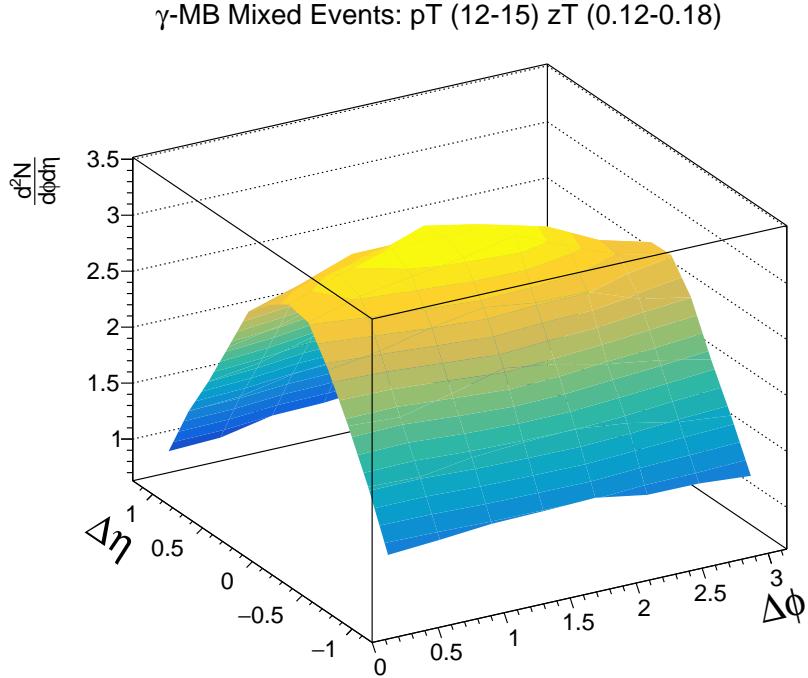


Figure 4.2: 10 Mixed-Event correlation using the traditional binning method. The major features presented in the Gale-Shapley-Paired Mixed event correlation are reproduced in this low-statistics binned mixing check; The trapezoidal shape in  $\Delta\eta$  and the non-uniformity in  $\Delta\phi$  from holes in the ITS are reproduced.

## 4.2 Underlying Event Estimation

The underlying event must be corrected for in two separate steps in this analysis. One is the pedestal in the correlation functions, attributed to correlating soft particles in the underlying event with the trigger isolated photon. The second is in the estimation of the underlying event density used in the isolation criteria. This section discusses checks on both corrections.

### 4.2.1 Large $\Delta\eta$ Check

The ZYAM assumption is incredibly useful, as it indicates the overall magnitude of the background arising from the underlying event. It is, however, an assumption that must be checked. To check that the pedestal estimate using the ZYAM assumption truly corresponds to the amount of uncorrelated background, it is cross checked with another reasonable background estimation method.

As a check on the ZYAM procedure, one can take advantage of the fact that the UE-estimation is independent of  $\Delta\eta$  and that genuine correlations due to hard-scatterings decrease as  $\Delta\eta$  increases. To this end, a region that is dominated by UE is selected, and then extrapolate back to the region that would normally contains both UE and hard-scattering contribution. The UE is estimated by projecting the large  $\Delta\eta$  region defined as  $0.8 < |\Delta\eta| < 1.4$  onto the  $\Delta\varphi$  axis. To minimize bias from the isolation cut as well as the away side jet peak, the uncorrelated background is estimated from the projection in the region  $0.4 < \Delta\varphi < 1.4$ . This  $\Delta\varphi\Delta\eta$  region is illustrated in Figure ??.

This region is chosen because particles from the same hard scattering are very unlikely to have a small  $\Delta\phi$  and large  $\Delta\eta$  in the absence of flow effects. This is because the nearside jet peak arises from the autocorrelation of the particles within a jet, i.e. it is sensitive only to the individual characteristics of a single parent parton and its fragmentation process. Therefore, the near side jet peak is observed as a sharp peak at small  $\Delta\phi$  and  $\Delta\eta$ . The away side jet ridge however, is the result of correlating particles between jets, and is therefore sensitive to the  $k_T$  asymmetry of the two colliding partons: The partons in the initial system do not *necessarily* have 0  $p_T$ . Both partons can have an initial transverse momentum,  $k_T$ , that makes up a component of their overall momentum fraction of the nucleon, Bjorken-x, and the resulting scattering becomes more spread out in  $\Delta\eta$ . For this reason, a region in small  $\Delta\phi$  and large  $\Delta\eta$  will be dominated by the underlying event, as it avoids the away-side ridge, and the sharp near side jet peak.

The statistical uncertainty in the UE estimate method is taken as a systematic uncertainty for  $\Delta\varphi$  correlations as it is completely correlated bin-to-bin in  $\Delta\varphi$ . Figure ?? shows the two UE estimates compared with the isolated photon-hadron  $\Delta\varphi$  correlations for only 2  $z_T$  bins in order to show detail. The full detail of the two UE estimates is shown in Tables ?? for pp and p-Pb data. The two estimates are consistent within uncertainties for almost all  $z_T$  bins in both pp and p-Pb data. For the only case where a significant disagreement is observed, which is for the lowest  $z_T$  bin in p-Pb data, the difference is summed in quadrature as an additional systematic uncertainty.

### 4.2.2 Checks on UE Estimate with Standard ALICE Tracking

As discussed in Section ??, the TPC had space-charge distortions during the 2013 p-Pb run that resulted in a drop in efficiency for tracking beyond 4 GeV, which limits the ability to use it for the correlation measurements. However, the TPC tracks can still be used for low  $p_T$  tracking, which is the relevant region for underlying-event and isolation measurements.

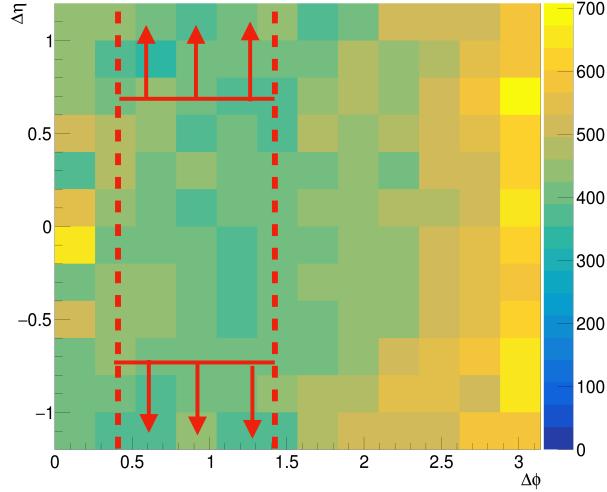


Figure 4.3: The 2D region used to calculate the uncorrelated background. The  $\Delta\varphi$  region is chosen to avoid the away side jet peak, as well as the isolation region of  $R=0.4$ . The  $\Delta\eta$  region is chosen assuming that genuine correlations from hard-scatterings decrease as  $\Delta\eta$  increases. The large  $\Delta\eta$  is projected onto the  $\Delta\varphi$  axis, and then averaged within region of  $0.4 < \Delta\varphi < 1.4$ . ZYAM is estimated in the region  $0.4 < \Delta\varphi < 1.4$ , but for the full  $\Delta\eta$  range ( $-1.2 < \Delta\eta < 1.2$ ).

Figure ?? shows the  $\rho$  and isolation distributions measured with ITS-only and TPC+ITS tracks in p–Pb data. The  $\rho$  distributions are very similar; the means of  $\rho$  to be 3.129 GeV and 3.202 GeV for ITS and TPC+ITS  $\rho$  values respectively. This is expected because the UE-estimation is dominated by tracks with low momentum, and while the ITS tracking resolution is poorer, the smearing effects are relative small at low momentum. There are some differences in the isolation distribution, however, that can be attributed to the worse momentum resolution for the ITS-only tracks as the isolation is sensitive to higher  $p_T$  tracks where the momentum resolution worsening is more significant.

For simplicity the same threshold of 1.5 GeV/ $c$  is used for the  $\gamma^{\text{iso}}$  candidates for both the ITS-only and ITS+TPC tracks. The ITS+TPC tracks leads to a better rejection of the background, which leads to an increased photon purity. This is shown in Figure ??.

The main results (correlation function) were checked in p–Pb data by performing the analysis with isolation variable, UE estimate, and corresponding purity values calculated separately for ITS-only tracks and for ITS+TPC tracks. As shown in Figure ??, the results are consistent. A slightly better statistical uncertainty is obtained when including TPC (a relative uncertainty of 22% to 41% depending on  $z_T$  vs 24% to 51% for the ITS case), which can be attributed to the corresponding higher purity. However, these slightly better statistical uncertainties do not change the main result of this study. For consistency with pp results (where one cannot use ITS+TPC tracks because the TPC was not read out), results for

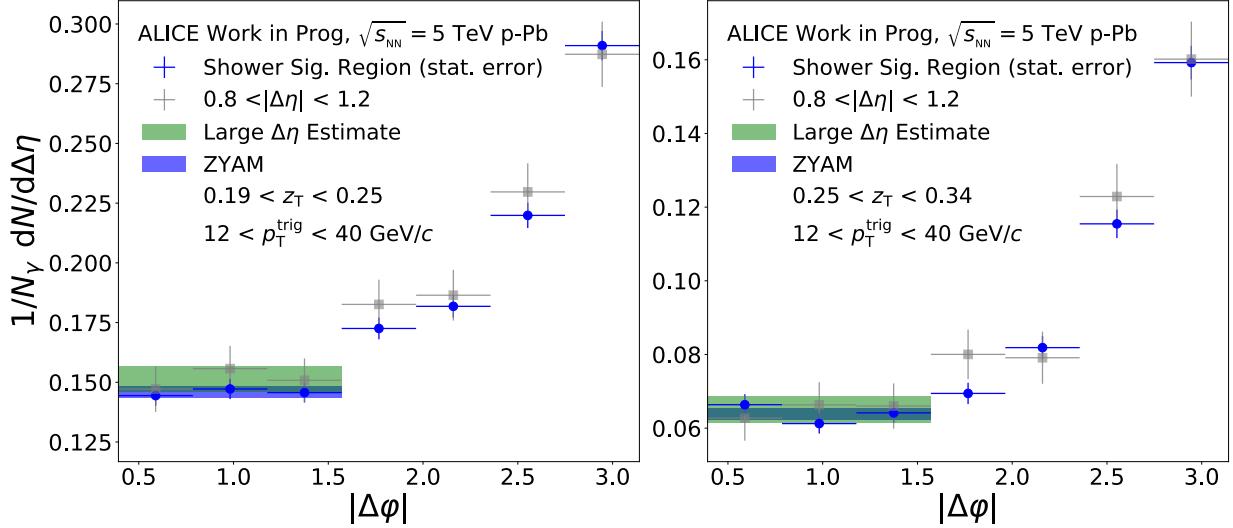


Figure 4.4: Projections of the  $\gamma^{\text{iso}}$ -hadron correlations in p-Pb collisions in 2  $z_T$  bins after correlated subtraction with UE estimates plotted. The grey points represent the large  $\Delta\eta$  region ( $0.8 < |\Delta\eta| < 1.2$ ) projected onto the  $\Delta\varphi$  axis. The blue band represents the region used to calculate ZYAM and the green band represents the region of large  $\Delta\eta$  points used to calculate the Large  $\Delta\eta$  estimate.

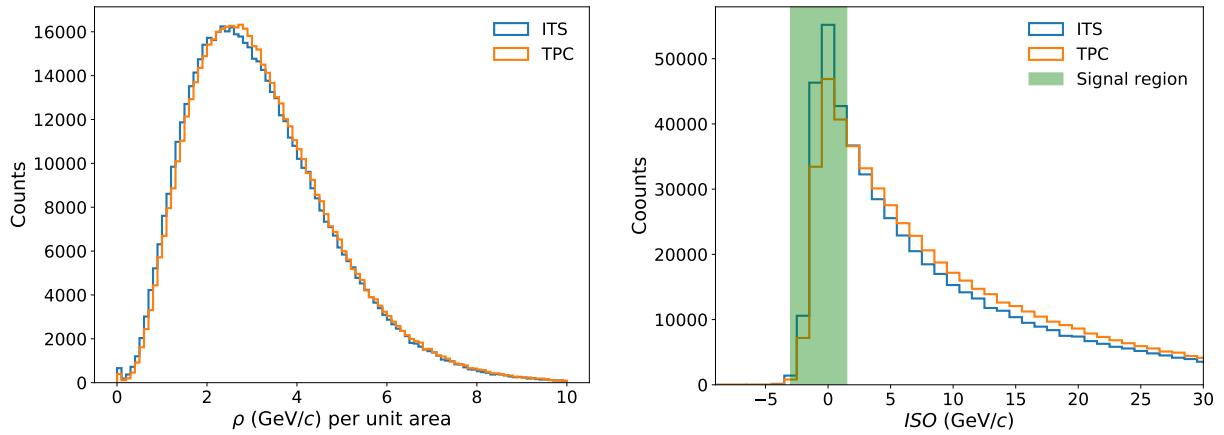


Figure 4.5: Transverse momentum density (left panel) and isolation distributions (right panel) determined with ITS tracks (in blue) and TPC+ITS tracks (in orange) in p-Pb data.

Table 4.1: Summary of UE-pedestal estimates with ZYAM and the  $\Delta\eta$  method for various  $z_T$  bins, as well as the difference between LE and ZYAM estimates. The background estimate is shown in units of pairs per trigger. The uncertainty quoted is statistical only.

$z_T$ interval	ZYAM	Large $\Delta\eta$	Difference
<b>pp</b>			
0.06 - 0.08	$0.480 \pm 0.007$	$0.472 \pm 0.015$	$0.008 \pm 0.016$
0.08 - 0.11	$0.347 \pm 0.006$	$0.346 \pm 0.012$	$0.000 \pm 0.014$
0.11 - 0.14	$0.219 \pm 0.005$	$0.209 \pm 0.010$	$0.009 \pm 0.011$
0.14 - 0.19	$0.131 \pm 0.004$	$0.129 \pm 0.008$	$0.002 \pm 0.008$
0.19 - 0.25	$0.063 \pm 0.002$	$0.058 \pm 0.005$	$0.005 \pm 0.006$
0.25 - 0.34	$0.030 \pm 0.002$	$0.025 \pm 0.003$	$0.005 \pm 0.004$
0.34 - 0.45	$0.013 \pm 0.001$	$0.015 \pm 0.003$	$0.002 \pm 0.003$
0.45 - 0.60	$0.006 \pm 0.001$	$0.005 \pm 0.002$	$0.001 \pm 0.002$
<b>p-Pb</b>			
0.06 - 0.08	$1.142 \pm 0.006$	$1.190 \pm 0.013$	$0.047 \pm 0.014$
0.08 - 0.11	$0.855 \pm 0.005$	$0.864 \pm 0.011$	$0.010 \pm 0.012$
0.11 - 0.14	$0.557 \pm 0.004$	$0.566 \pm 0.009$	$0.009 \pm 0.010$
0.14 - 0.19	$0.318 \pm 0.003$	$0.317 \pm 0.007$	$0.001 \pm 0.007$
0.19 - 0.25	$0.151 \pm 0.002$	$0.159 \pm 0.005$	$0.009 \pm 0.005$
0.25 - 0.34	$0.062 \pm 0.001$	$0.062 \pm 0.003$	$0.000 \pm 0.003$
0.34 - 0.45	$0.022 \pm 0.001$	$0.021 \pm 0.002$	$0.001 \pm 0.002$
0.45 - 0.60	$0.007 \pm 0.000$	$0.006 \pm 0.001$	$0.001 \pm 0.001$

ITS-only tracks was chosen for the final analysis.

This study comparing ITS-only tracking and ITS+TPC tracks serves as a check against possible biases due to worse momentum resolution or fake rate of the ITS-only tracking. Because consistent results are obtained with the standard ITS+TPC tracking, an additional systematic uncertainty is not assigned for the tracking performance on UE-estimate and isolation variables.

## 4.3 Purity

### 4.3.1 Systematic Uncertainties of the Purity Measurement

There are two assumptions underlying the template fit procedure. The first is that the signal template from simulations is correct. The second is that the shape of the background estimated from the anti-isolated sideband, with the correction estimated from simulations, reflects the shape of the background in the signal region. In other words, the assumption is that the correlation between the shower shape and isolation variables can be corrected for via an appropriate simulation. The dominant sources of systematic uncertainty on the

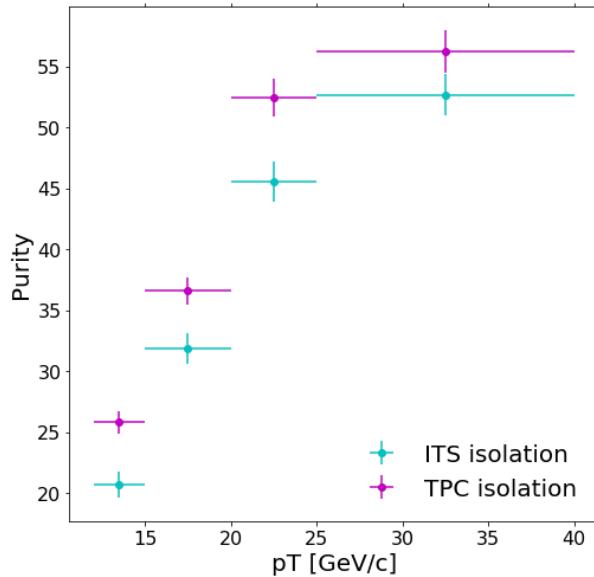


Figure 4.6: Comparison of purity obtained in p–Pb collisions with isolation variable obtained with ITS-only tracks (purple) and with ITS+TPC tracks (cyan). The error bars represent statistical uncertainty only.

purity calculation are described in this section and can be summarized as follows: the signal template, the sideband region selection, and the background template correction. The effect of varying the cluster selection requirements were investigated, but it was found that the variations on the purity measurements are much smaller than the other sources of systematic uncertainties investigated here. Therefore, they are neglected. This is shown in Sec. ??.

### 4.3.2 Signal Template

The systematic uncertainty on the purity calculation arising from imperfections in the signal template is estimated by using a data-driven template fit. To this end, the range of the  $\chi^2$  fit to the background-dominated region of the shower-shape distribution is restricted (0.4–1.5 for  $\sigma_{\text{long}}^2$ ) and the background template is used only to fit the isolated data with the normalization as the only free parameter. In order to factorize the effect of the MC-correction to the background template, that correction is not applied here for this study. Once the background normalization is fitted, the signal is considered to be the integral of the isolated data minus the integral of the background, both in the signal region of the shower-shape variable.

Figure ?? shows the results obtained with this method in pp and p–Pb data. There appears to be some systematic pattern in the residuals, which can be attributed to the decision of not applying the MC-correction to the background template for this study.

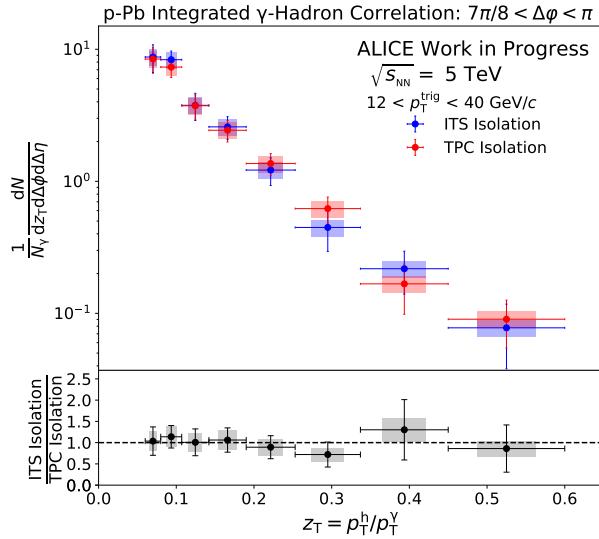
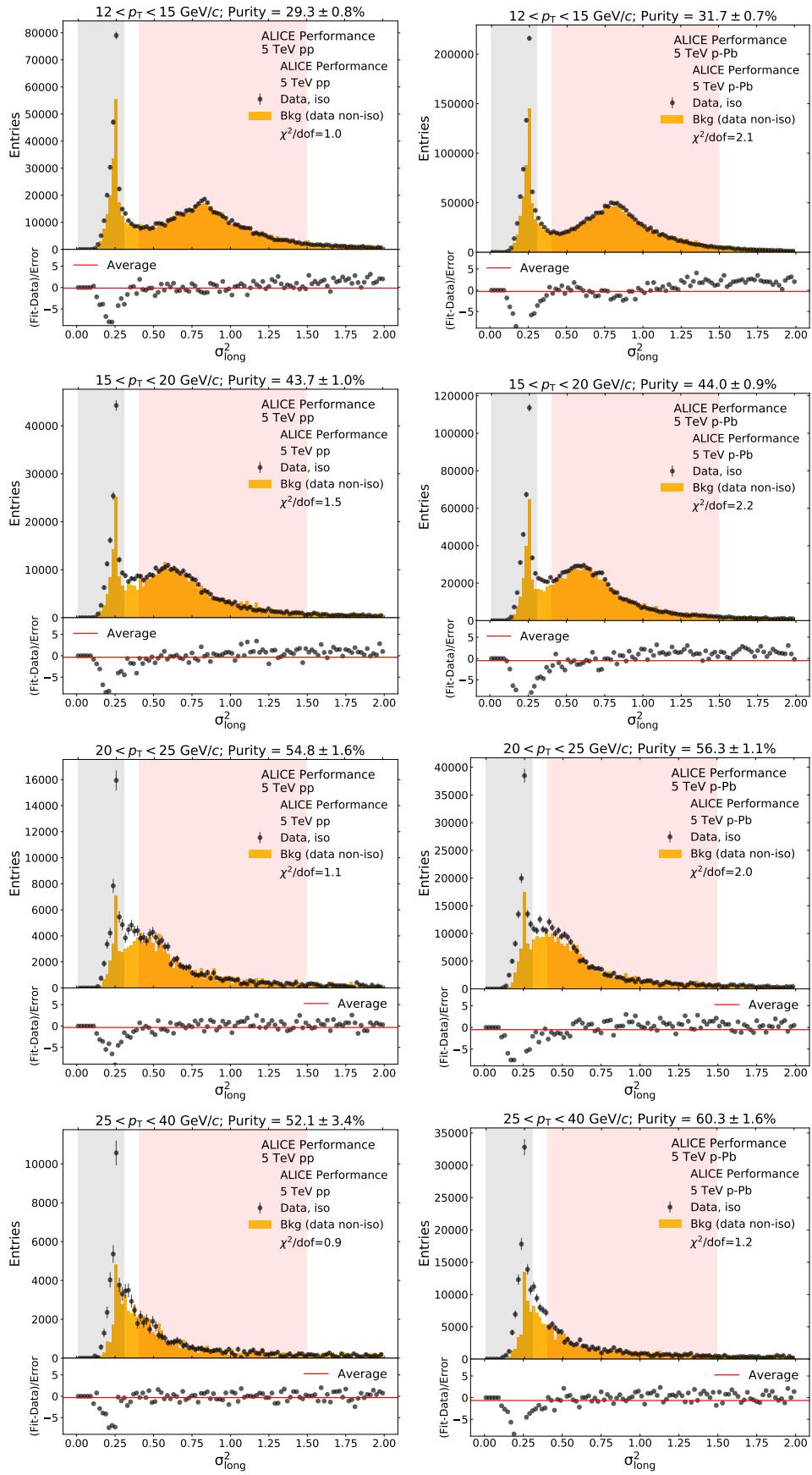


Figure 4.7: Comparison of fragmentation function measurement in p–Pb collisions with isolation variable obtained with ITS-only tracks (blue) and with ITS+TPC tracks (red).

This method makes the additional assumption that the fraction of signal misclassified as background is small. Strictly speaking, this method yields a lower limit on the extracted purity. The results agree with the nominal results within a few percent, indicating that the measurement is not particularly sensitive to the details of the modeling of the shower shape. As a conservative estimate, the full difference between the nominal results is taken as a systematic uncertainty in the signal template.

As an additional check, the signal template is smeared by multiplying the  $\sigma_{\text{long}}^2$  of each cluster by a random number selected from a Gaussian with a fixed width before then calculating the purity with this smeared distribution. This was done for a variety of widths up to 10%, which is much larger than the expected MC simulation mismodelling, and was found to yield a smaller uncertainty than the background-only fit (See Appendix ?? for more details). Thus, the uncertainty estimated by the background-only fits as described in the previous paragraph is taken as a final estimate of the systematic uncertainty on the purity arising from the signal template.



### 4.3.3 Sideband variation in the background template

To estimate the shower-shape distribution for the  $\gamma^{\text{decay}}$  background in the template fit, a sideband in the cluster isolation variable is used. Only the shape of this distribution is relevant, as the overall background normalization in the signal region (i.e. the purity) is measured with the template fit. As in any analysis using a sideband technique, nominally a sideband as close as possible to the signal region and as narrow as possible is used. Here, how the sideband region is chosen is discussed and the systematic uncertainty that arises from this arbitrary choice is addressed.

The cluster isolation distribution is divided into narrow ( $2 \text{ GeV}/c$ ) overlapping regions, each of which is used to estimate the background shower-shape distribution. A template fit is performed with each distribution and the  $\chi^2/\text{dof}$  and purity are calculated for each fit and plotted as a function of the anti-isolation region used to create the background template in Figure ???. Then, the  $\chi^2/\text{dof}$  distribution is examined to determine which regions of anti-isolation result in good fits in the template fit procedure:  $5\text{--}10 \text{ GeV}/c$  is chosen to be the sideband definition in the final purity calculation.

To calculate the systematic uncertainty on the purity due to this selection, the full range of purities reached by the narrow bands of anti-isolation that fall within  $5\text{--}10 \text{ GeV}/c$  is considered. Converting the full extent to a systematic uncertainty is a matter of dividing by  $\sqrt{12}$  (i.e, the  $1\sigma$  for a uniform distribution). This results in an absolute uncertainty on the purity of 0.7–5.8%, depending on the collision system and cluster  $p_T$  range.

### 4.3.4 Background template correction

Due to the correlation between the isolation and shower shape, the template extracted from the anti-isolated sideband does not exactly reflect the shape of the background in the signal region. Clusters in the isolation sideband have more associated activity than those in the true isolated background and thus emphasize the non-signal region of the shower-shape distribution. Consequently, using the isolation sideband instead of the true isolated background yields systematically higher purities. It should be noted that a similar observation was made for example by the CMS collaboration in their template-fit purity measurements (e.g. Ref. [Sirunyan:2017qhf])

A dijet MC simulation is used to correct this bias, as described in Equation ???. However, this correction is only valid to the extent that the dijet MC reproduces the data. To estimate the systematic uncertainty on this correction, a technique based on a method used in the ABCD calculation [Erwann] is implemented. In particular, a double ratio is used to check to which extend does the dijet MC describe the background-dominated region in data:

$$\text{Double ratio} = \frac{\text{Iso}_{\text{data}}/\text{Anti-iso}_{\text{data}}}{\text{Iso}_{\text{MC}}/\text{Anti-iso}_{\text{MC}}} \quad (4.1)$$

In the signal region of the shower shape distribution ( $0.0\text{--}0.3$  for  $\sigma_{\text{long}}^2$ ), this double ratio will be far from unity, as the data have prompt photons and the dijet MC do not.

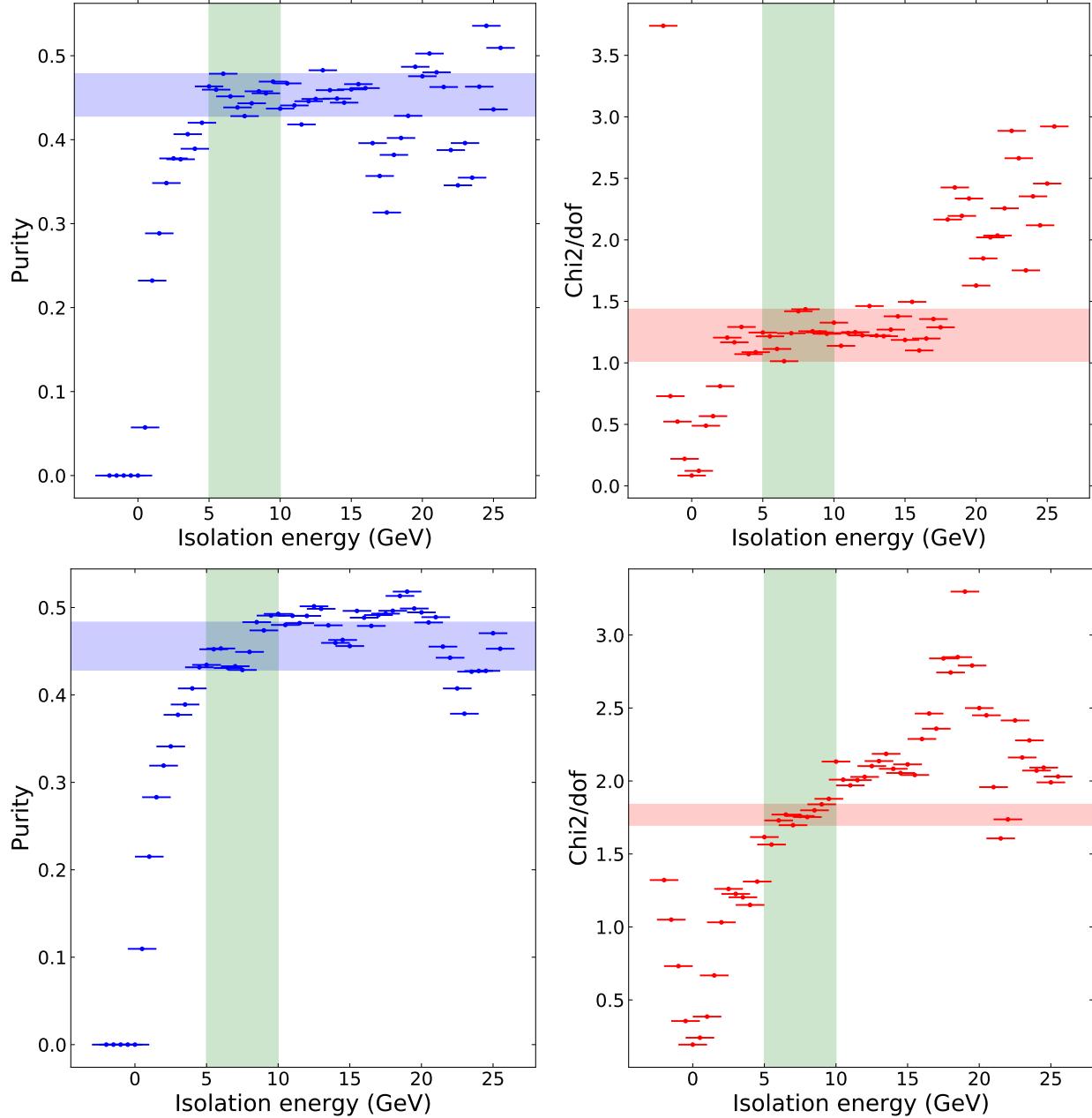


Figure 4.9: Template fit results (purity and  $\chi^2/\text{dof}$ ) as a function of anti-isolation region for clusters with  $15 < p_T < 20 \text{ GeV}/c$  in pp (top) and p-Pb (bottom). The green band shows the selected sideband region. The blue and red bands show the full extent of the purity within the selected sideband region.

However, away from that region, where background dominates, the double ratio should be flat (i.e. have no slope) if the dijet MC reproduces the background shower-shape of the data. It should be noted that for this analysis, only the shape is important and overall normalization is irrelevant. At a minimum, the variation in the double ratio is expected to be smooth. Thus the double ratio is fit to smooth functions (linear and exponential) in a shower shape range away from the signal region and extrapolate the fit back into the signal region. A similar procedure was used isolated-photon purity measurements with the ABCD method [**Acharya:2019jkx, Erwann**].

Fits to the double ratio are shown in Figures ?? and ??, for pp and p–Pb data respectively. It was found that the linear and exponential fits gave nearly identical results. In particular, the slope was sufficiently small that the higher-order terms in the exponential were negligible. Thus for the purposes of estimating the systematic uncertainty due to the background template correction, only linear fits to the double ratio were done. In order to remove covariance effects between the slope and intercept, the fits were forced to go through the weighted average of the double ratio value within the fit range at the center of the fit range, making it a single-parameter linear fit with only the slope as a free parameter. This allowed us to propagate the fit uncertainty on the slope to an uncertainty on the purity.

These linear fits to the double ratio were done in two fit ranges: 0.5–1.5 and 0.5–1.75 for  $\sigma_{\text{long}}^2$ . In all cases, it was found that the slopes were consistent with 0 within the fit uncertainties and thus concluded that the dijet MC was consistent with the data. Therefore, an additional double-ratio correction is not applied to the Weights function in Equation ???. It was also found that the double ratio fits with the different fit ranges gave purities consistent with each other. So in order to minimize the amount of extrapolation, the fit in the largest reasonable fit ranges for each of the variables was done (the larger of each of the ranges described at the beginning of this paragraph).

The uncertainty on that double ratio fit is then taken and propagated to a purity uncertainty. This purity uncertainty was then taken to be the systematic uncertainty on the background correction. It varies between 1.2–3.4% (absolute) depending on cluster  $p_{\text{T}}$  and collision system.

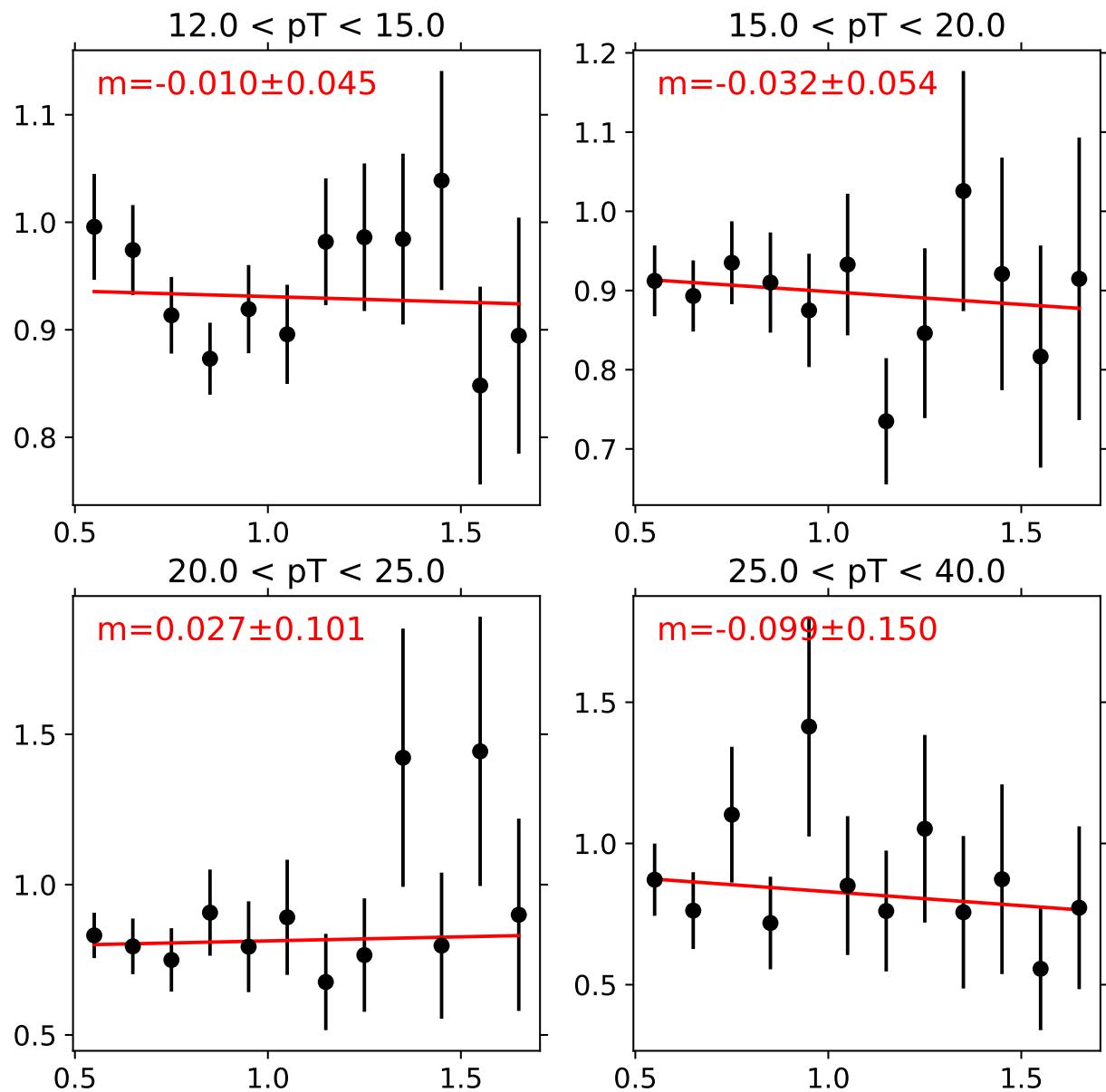


Figure 4.10: Linear fits for the double ratio (as described in Equation ??) for the  $\sigma_{\text{long}}^2$  variable in pp data. Included are the value and uncertainty of the fitted slope (in red).

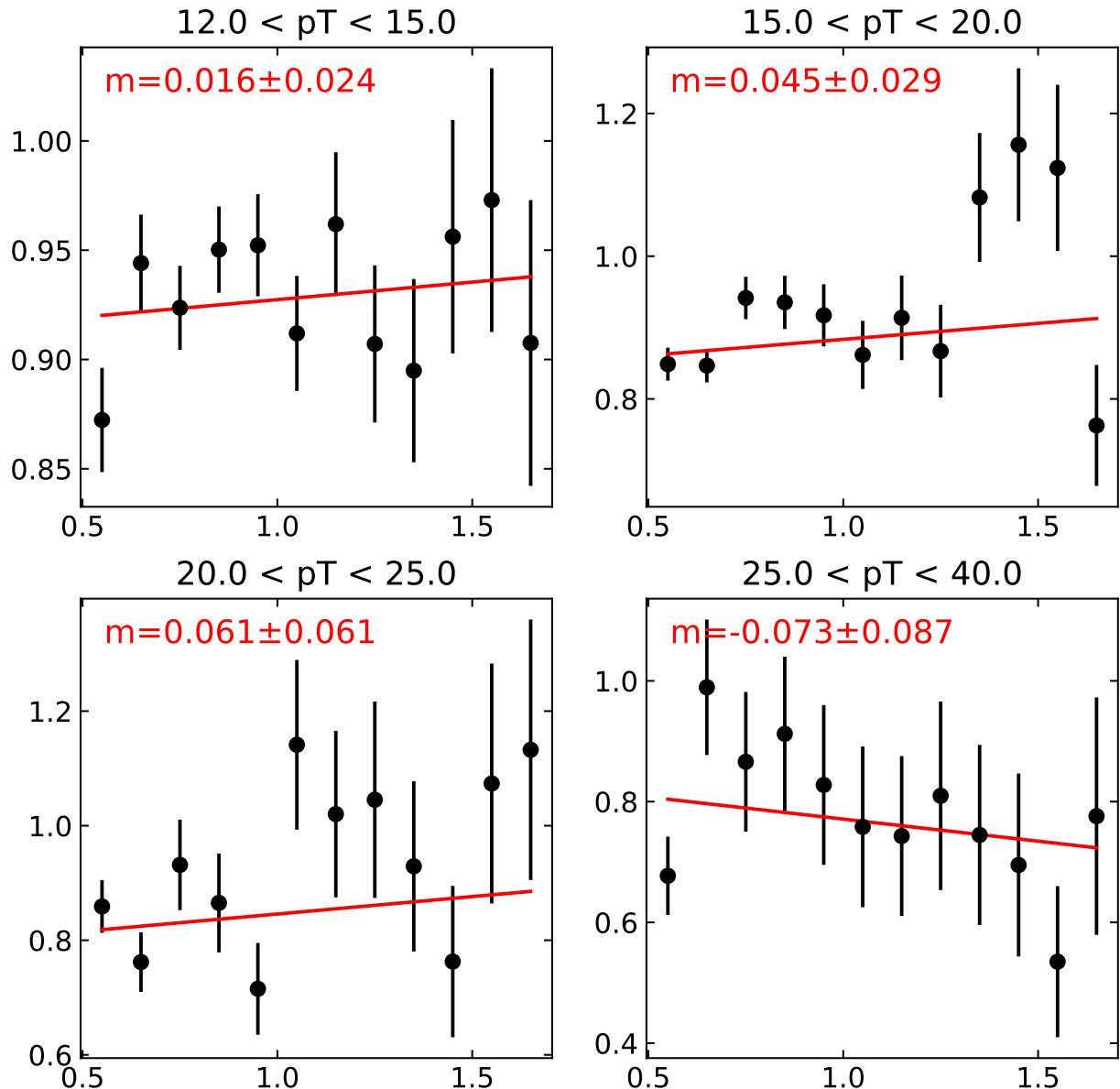


Figure 4.11: Linear fits for the double ratio (as described in Equation ??) for the  $\sigma_{\text{long}}^2$  variable in p–Pb data. Included are the value and uncertainty of the fitted slope (in red).

### 4.3.5 Summary of Systematic Uncertainties on Purity Measurement

Tables ?? and ?? give the full estimates of the systematic uncertainties in both collision systems. No single source of systematic uncertainty dominates across  $p_T$  ranges or collision systems.

Table 4.2: Summary of the systematic uncertainties on the purity as measured with  $\sigma_{\text{long}}^2$  in p–Pb collisions. All values are in absolute percentage. “Stat.” refers to the statistical uncertainty; “Signal” refers to the signal template uncertainty; “Anti-iso” refers to the uncertainty due to the sideband selection; “Bkg” refers to the uncertainty due to the background template correction; “Total” is the sum of the previous three columns in quadrature.

$p_T(\text{GeV}/c)$	Purity	Stat.	Signal	Anti-iso	Bkg	Total syst
12.0-15.0	20.7	1.1	1.1	0.8	1.5	2.0
15.0-20.0	34.2	1.2	2.0	1.6	1.2	2.8
20.0-25.0	47.6	1.7	1.9	1.1	1.7	2.7
25.0-40.0	54.6	1.8	2.3	2.4	2.1	3.9

Table 4.3: Summary of the systematic uncertainties on the purity as measured with  $\sigma_{\text{long}}^2$  in pp collisions. All values are in absolute percentage. “Stat.” refers to the statistical uncertainty; “Signal” refers to the signal template uncertainty; “Anti-iso” refers to the uncertainty due to the sideband selection; “Bkg” refers to the uncertainty due to the background template correction; “Total” is the sum of the previous three columns in quadrature.

$p_T(\text{GeV}/c)$	Purity	Stat.	Signal	Anti-iso	Bkg	Total syst
12.0-15.0	20.1	1.7	2.0	1.2	2.9	3.7
15.0-20.0	31.7	2.0	2.5	1.5	2.4	3.8
20.0-25.0	47.3	2.9	0.8	3.0	2.8	4.2
25.0-40.0	48.5	3.5	5.9	4.0	3.4	7.9

## 4.4 Tracking

### 4.4.1 Comparison to Published Data

In this section results of the closure test, which is described in Section ??, are shown with the goal to validate the MC simulation description of track efficiency, fake rate and momentum smearing corrections using the published data.

Figure ?? shows the correction procedure on published data at various stages: the pink line is the published charged-particle spectrum in p–Pb collisions at  $\sqrt{s_{\text{NN}}} = 5$  TeV from Ref. [Acharya:2018qsh]; the red line is the published data spectra after multiplying by the

efficiency; the orange line is the red spectrum multiplied by the response matrix according to Equation ??:

$$R_j = \sum M_{ij} \cdot P_i \quad (4.2)$$

where  $M$  is the response matrix,  $P$  is the published data times the efficiency,  $R$  is smeared  $p_T$  spectrum. The smeared  $p_T$  spectrum should be consistent with the measured spectrum after removal of the fake-track contribution.

A "folding" of the published spectra is shown instead of an "unfolding" of the measured spectra because the former is a unique transformation and avoids the need for systematic studies on the stability of the unfolding procedure. This allows us to focus on unambiguously testing the response matrix.

The published data has a total uncertainty (quadrature sum of statistical and systematic uncertainties) that ranges from 1.8% at 1  $\text{GeV}/c$ , reaches 4.8% by 10  $\text{GeV}/c$  and grows quickly to about 20% at 15  $\text{GeV}/c$ , where it is dominated by the statistical uncertainty.

Figure ?? shows the ratio of the measured fake-subtracted spectrum and the smeared published spectra. The ratio of the published data to the smeared published data is shown to illustrate the impact of the momentum smearing, which is less than a 2% effect for TPC+ITS tracks but it reaches up to a factor of two in the ITS-only case. The closure-test ratios from TPC+ITS tracking are consistent with unity within uncertainties, which is expected. The more interesting result is that the ratios due to ITS-only tracking are within  $\pm 5\%$  of unity in the range between 0.85–10  $\text{GeV}/c$  and within  $\pm 8\%$  of unity in the range between 0.5–0.85  $\text{GeV}/c$ , shown by the dashed lines in the bottom plot of Figure ??, which is the range used for the  $\gamma^{\text{iso}}$ -hadron analysis. This difference from unity is used as a systematic uncertainty on the tracking.

The statistically-significant deviation from the published data with ITS-only tracking at high  $p_T$  (blue curve in lower panel of Figure ??), could be due to several reasons including improper modeling of a rapid deterioration of the momentum resolution and underestimation of fake rate. Further work would be needed to understand and correct these and other effects at high  $p_T$ , but that lies beyond the scope of this work.

It should be noted that these spectra are normalized per-minimum bias event and not by integral, so the fact that the ratios are close to unity reflect the fact that the efficiency calculations shown in Figure ?? are an accurate description of the detector response. Based on the results shown in this section, the  $p_T$  range is restricted to no more than 10  $\text{GeV}/c$ , which is beyond the limit of the statistical power of the  $\gamma^{\text{iso}}$ -hadron analysis.

#### 4.4.2 Validation of $\varphi$ -dependence of efficiency

In order to validate the description of the ITS-only tracking holes, a test with minimum-bias data is done. Apart from the effect of  $\varphi$ -dependent holes, the measured azimuthal angle distribution of tracks is expected to be uniform in minimum-bias data. Thus the track  $\varphi$  spectrum is measured and then corrected for the  $\varphi$ -dependent efficiency and checked whether

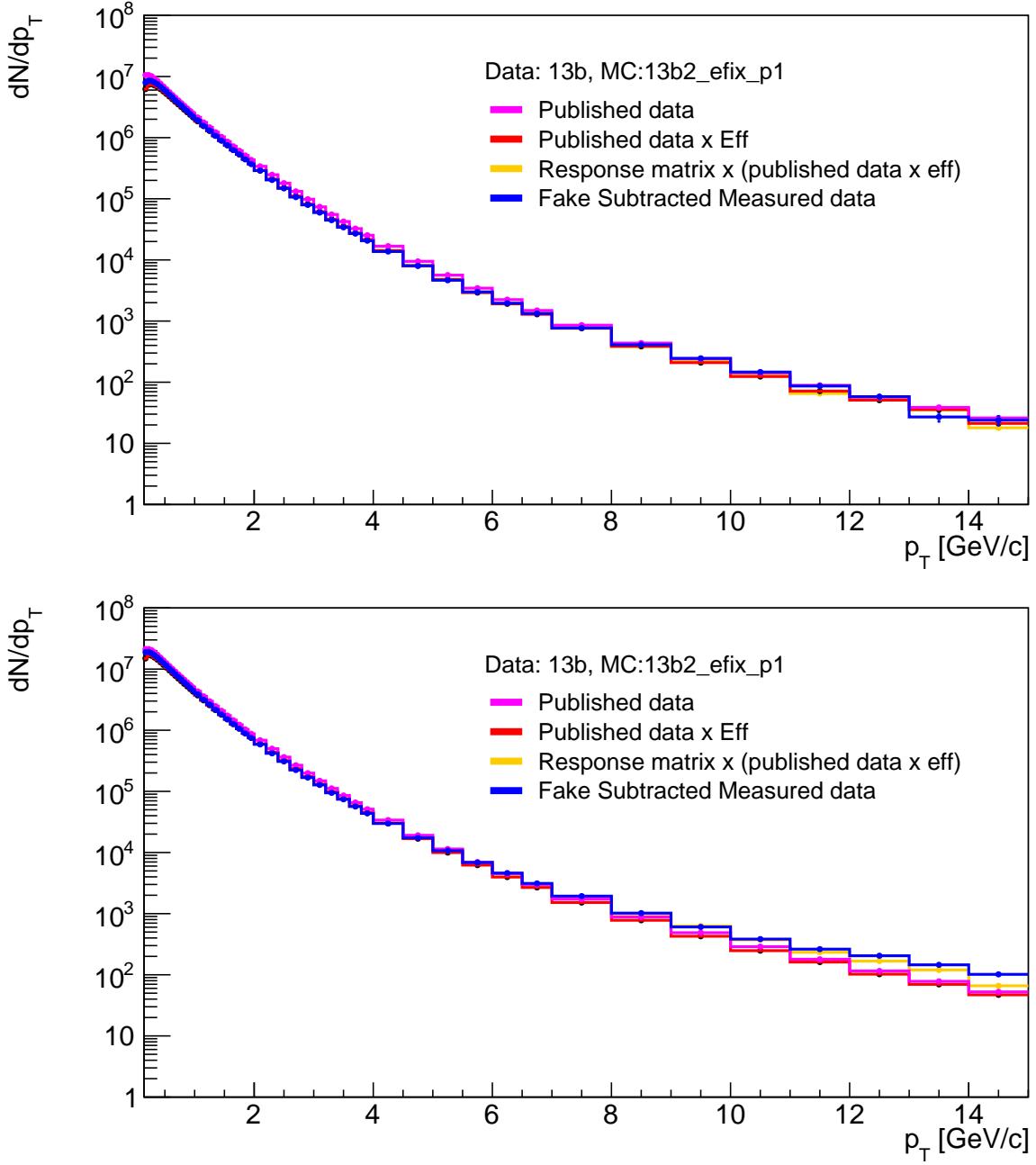


Figure 4.12: Smearing of published data at various stages, using TPC+ITS (top) and ITS only (bottom) response matrices. The pink is the published data  $p_T$  spectrum. The red is after the pink has been multiplied by the efficiency. The orange is the spectra after applying the response matrix in order to induce the smearing on the red. The blue is fake-rate-subtracted measured data.

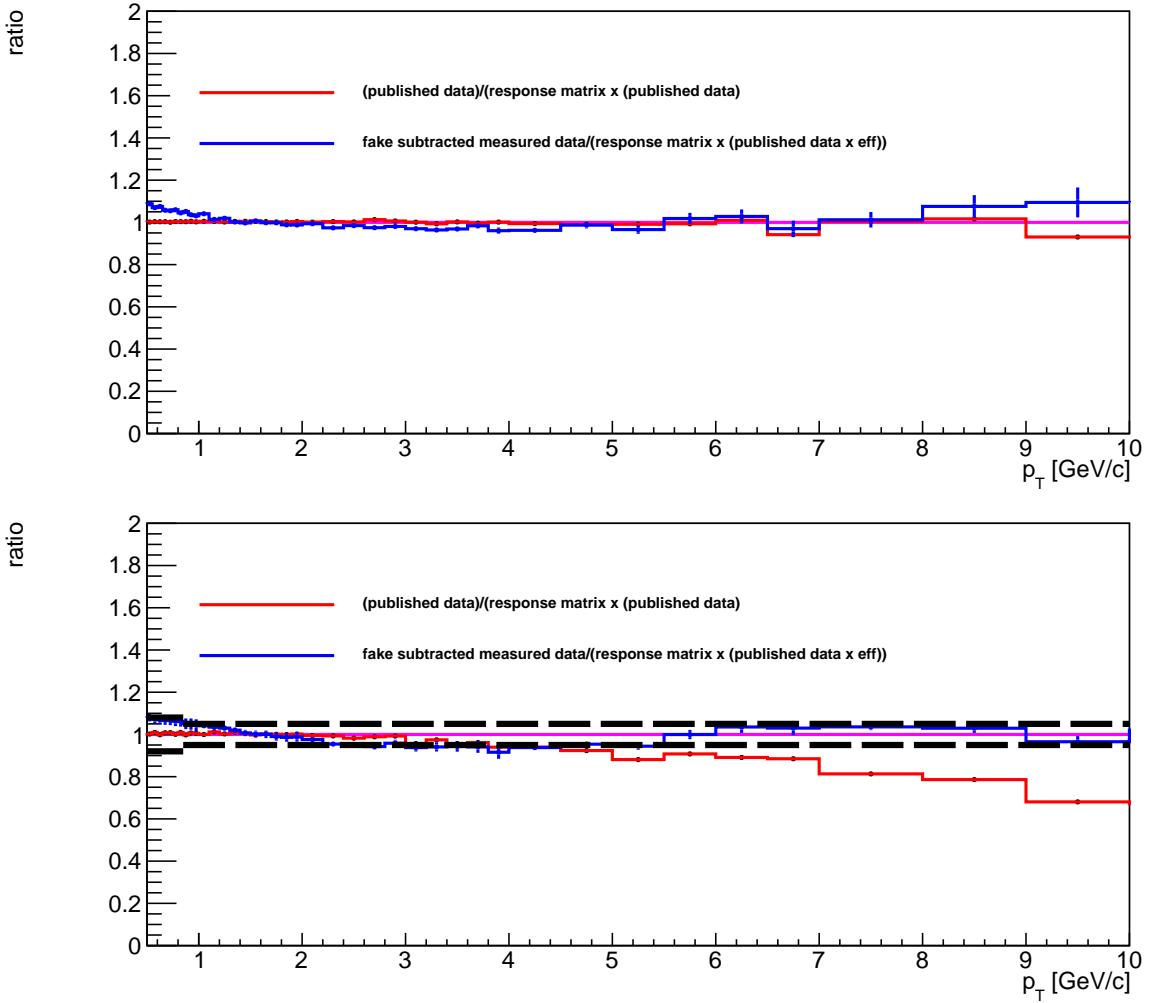


Figure 4.13: Result of closure test comparing measured data and published data, for TPC+ITS tracking (top) and ITS-only tracking (bottom). The red curves show the ratio of the reference spectra to the smeared reference spectra. The blue curves show the ratio of the fake-subtracted measured data and the smeared reference spectra. Ideally the blue curve would be flat at unity. The error bar represents statistical uncertainty only for the blue curve, and the quadrature sum of statistical and systematic uncertainties for the red curve. Additionally, the dashed lines from 0.5 to 0.85 GeV/c represent an 8% band around 1, while the dashed lines from 0.85 to 10.0 GeV/c represent a 5% band around 1.

the distribution is flat. The level of flatness gives us a sense of the systematic uncertainties associated with mis-modeling of the  $\varphi$ -dependent efficiency.

Figure ?? shows the  $\varphi$ -dependence of the tracking efficiency for TPC+ITS and ITS-only tracks. The efficiency is calculated using Equation ??, but as a function of  $\varphi^{\text{true}}$  instead of  $p_T^{\text{true}}$ . The TPC+ITS tracking efficiency is flat in  $\varphi$  as expected, but there are dips in the efficiency for the ITS-only tracking due to dead staves in the ITS. These have little impact on the TPC+ITS tracks because the selection does not have the strict requirement on the number of ITS hits,  $N_{\text{ITS}} \geq 4$ , that is applied for ITS-only tracks.

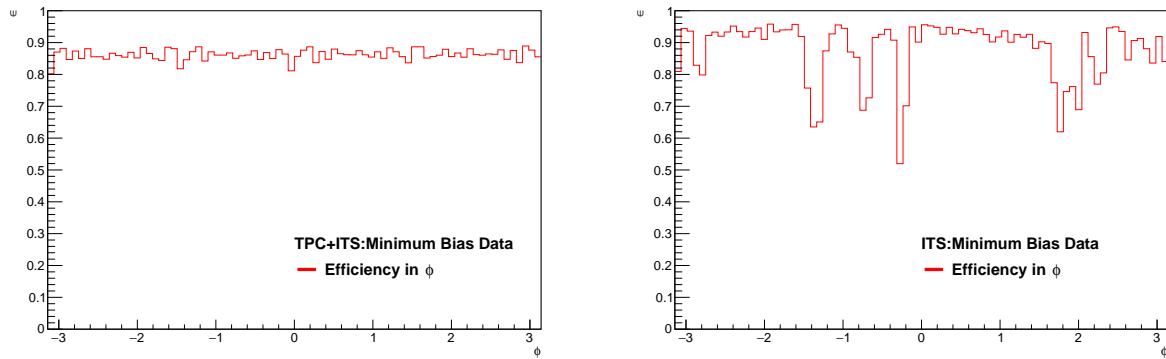


Figure 4.14: Tracking efficiency as a function of  $\phi^{\text{true}}$  for TPC+ITS tracks (left) and ITS-only tracks (right). In both cases, the efficiency is calculated for tracks with  $p_T^{\text{true}} > 1 \text{ GeV}/c$  using the LHC13b2\_efix\_p1 Monte Carlo simulation.

Figure ?? shows the  $\varphi$  distribution of TPC+ITS and ITS-only tracks in minimum-bias p-Pb data and the effect of applying the efficiency correction to the  $\varphi$  distribution. Before applying the efficiency correction, there are visible holes at  $\varphi = -1.04, -0.8, -0.2$  rad in ITS-only tracks which are not present in the TPC+ITS tracking. After applying the  $\varphi$  efficiency, the holes are corrected, and a distribution which is flat within  $\pm 2.5\%$  is obtained. The TPC+ITS remains flat after the efficiency correction, as expected. This shows that the description of dead channels in the ITS is well-described in the simulation.

#### 4.4.3 Hybrid tracking on isolation check

#### 4.4.4 Summary of the ITS-only Tracking Performance Studies

This section summarizes the findings of the studies on ITS-only tracking performance:

1. Tracking Efficiency:

The ITS-only tracking efficiency is 75% at 150 MeV/ $c$  and grows to 85% at 1 GeV/ $c$  and above.

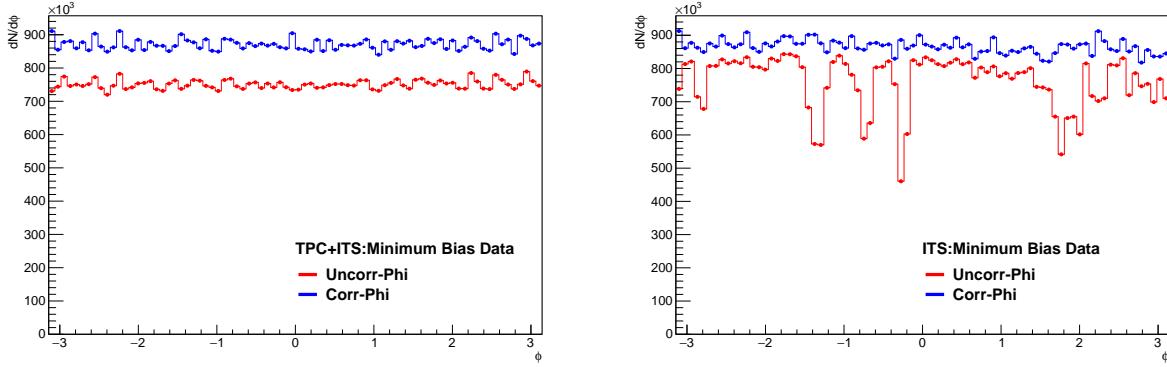


Figure 4.15: Left panel: track  $\varphi$  distribution measured in data for tracks with  $|\eta| < 0.8$  before (red) and after (blue) applying the efficiency correction for TPC+ITS tracks. Right panel: track  $\varphi$  distribution measured in data for tracks with  $|\eta| < 0.8$  before (red) and after (blue) applying the efficiency correction for ITS-only tracks.

## 2. Fake rate:

The fake rate of ITS-only tracking is about 10 times worse than for TPC+ITS tracks, but still less than 20% below 10  $\text{GeV}/c$ , which is the relevant range for the analyses presented in this note.

## 3. Momentum resolution:

The momentum smearing effects are significant for ITS-only tracking. The bin-to-bin correction factor due to smearing effects for ITS-only tracking is 0.7 at 10  $\text{GeV}/c$  and 0.5 at 15  $\text{GeV}/c$ . The smearing effects for ITS+TPC tracking are negligible.

## 4. Description of $\varphi$ holes

The efficiency as a function of  $\varphi$  shows inhomogeneity not present in the TPC+ITS tracking that are attributed to dead staves in the ITS. These are concentrated in specific  $\eta$  and  $\varphi$  regions. These are well described in the simulation.

These studies validate the MC corrections for tracking efficiency, fake rate and momentum smearing by comparing with published data. From that study, the combined systematic uncertainty on tracking performance is estimated to be a relative  $\pm 5\%$  for ITS-only tracks with  $0.5 < p_T < 10 \text{ GeV}/c$ .

## 4.5 Neutral Energy in Isolation Variable

In this analysis, the isolation variable was constructed using only charged-particles. In principle, neutral particles could have been added as well in the isolation definition. However,

that would have limited the acceptance the measurement. For example, the recent ALICE isolated photon paper [Acharya:2019jkx] restricted the pseudorapidity of the  $\gamma^{\text{iso}}$  to  $|\eta| < 0.27$  to ensure a good containment of the isolation cone that has a radius of  $R = 0.4$  (the EMCAL acceptance is  $|\eta| < 0.67$ ). An acceptance limitation would have a large impact of this analysis in terms of statistical precision, so a "charged-only" isolation was chosen. This is not different than several ALICE jet analyses that report "charged-only" jets and not "full-jets".

PYTHIA8 events are used to estimate the impact on the isolation variable of including neutral-particles. Figure ?? shows the comparison of the prompt-photon hadron correlations according to PYTHIA8 when using no isolation requirement; with an isolation variable based on charged particles (used in this analysis); and with an isolation variable based on both charged-particles and neutral particles. In all cases, the charged-particles and neutral particles are final-state particles and have  $p_T > 150 \text{ MeV}/c$  and  $|\eta| < 0.8$ . No significant difference between the selection of  $1.5 \text{ GeV}/c$  based on charged particles and the selection based on  $2.0 \text{ GeV}/c$  based on charged and neutral particles was observed. Therefore, the  $\text{ISO} < 1.5 \text{ GeV}/c$  selection is enough to suppress the near-side peak in the correlation functions coming from fragmentation photons, and that using neutral-particles in the isolation variable would not yield any significant improvement.

## 4.6 Impact of Acceptance Difference Between pp and p–Pb due to Boost

In this section, the impact of the acceptance difference between pp and p–Pb data that arises due to the boost in p–Pb data is estimated. The boost in p–Pb data arises due to the energy difference between the proton and lead beam, and it amounts to a rapidity difference of  $\Delta y = 0.47$  in the proton-going direction. That means that in p–Pb collisions, the lab acceptance for photons that is  $-0.67 < \eta < 0.67$  corresponds to  $-0.2 < \eta < 1.14$  in the center-of-mass frame, whereas the charged-particle acceptance of  $-0.8 < \eta < 0.8$  corresponds to  $-0.33 < \eta < 1.27$  in the center-of-mass frame.

PYTHIA8 events are used to estimate what is the difference between  $\gamma^{\text{iso}}$ –hadron correlations with the acceptance of  $\gamma^{\text{iso}}$  and charged particles  $-0.20 < \eta < 1.14$  and  $-0.33 < \eta < 1.27$  instead of the nominal ranges of  $-0.67 < \eta < 0.67$  and  $-0.8 < \eta < 0.8$ . This is shown in Figure ???. The boosted acceptance results in about a 5% lower  $\gamma^{\text{iso}}$ –hadron correlation compared to the nominal acceptance, irrespective of  $z_T$  range. For illustration purposes, the impact of a boost of  $\Delta y = 1.0$  is plotted, which shows a decrease of about 15% with respect to the nominal acceptance, irrespective of  $z_T$  range.

The chosen integration window (dashed lines) in the figure above makes it clear what effect this may have on the final away side yields. Thus the effect of acceptance the mismatch in this analysis is limited due to the relatively small boost of  $\Delta y = 0.47$  and the limited acceptance of EMCAL and ITS, which even with boost is within mid-rapidity region where

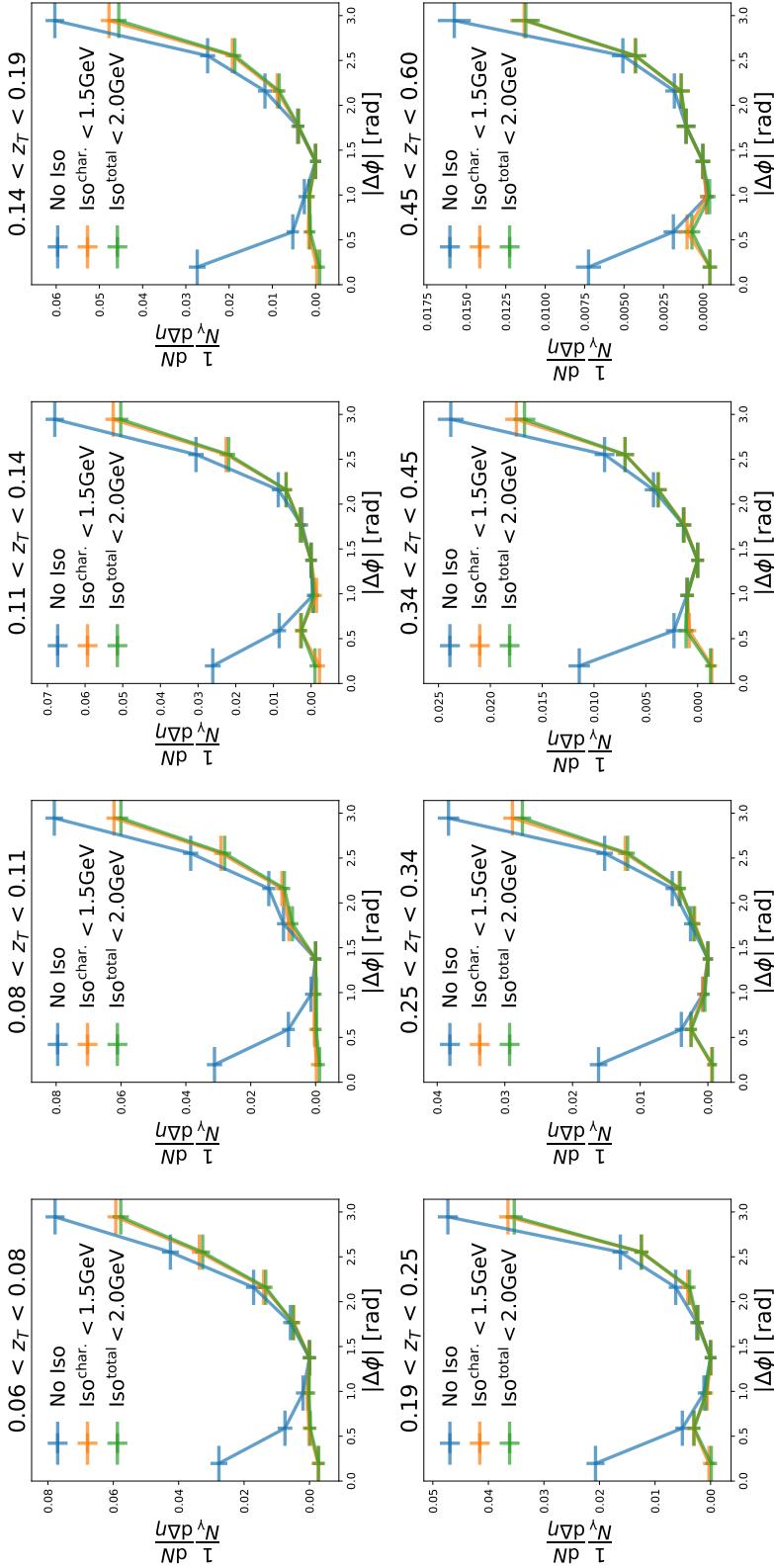


Figure 4.16: Correlation function between prompt photons and hadrons from PYTHIA8 for various  $z_T$  bins. Three selections on the prompt photons based on isolation are presented: no isolation (blue); ISO  $< 1.5\text{GeV}/c$  that considers only charged-particles (orange); and ISO  $< 2.0\text{ GeV}/c$  that considers both charged and neutral particles (green). In all cases, the uncorrelated background has been subtracted using the ZYAM method. The error bar represent statistical uncertainty only.

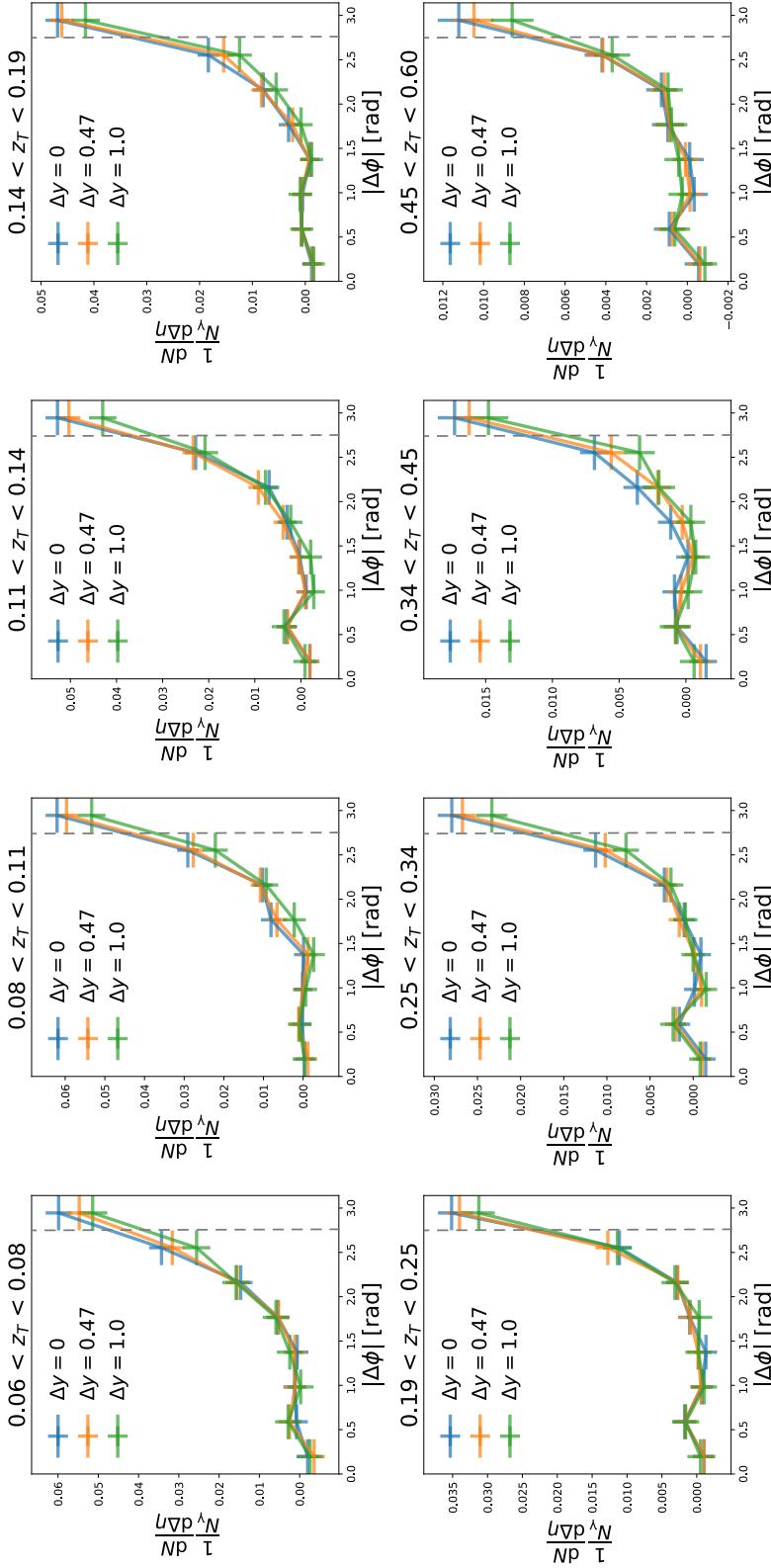


Figure 4.17: Correlation function between isolated prompt photons and hadrons from PYTHIA8 for various  $z_T$  bins. The nominal result ( $\Delta y = 0$ , in blue) is compared with results obtained with a kinematic selection that mimics the boost of p–Pb data (orange). For illustration purposes, the impact of a boost that is larger than the one of p–Pb data ( $\Delta y = 1.0$  in green) is shown. The dashed lines at  $\varphi = \frac{7\pi}{8}$  indicates the integration window used to obtain the away side yields.

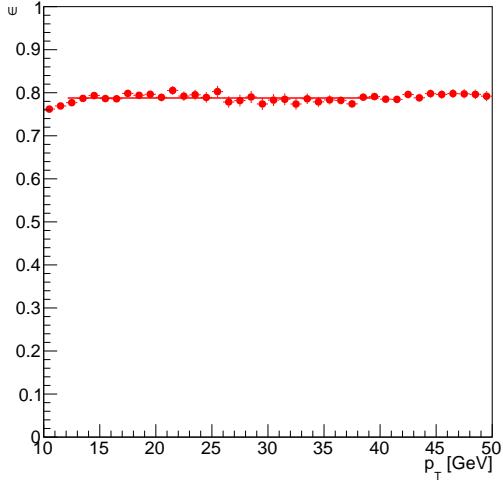


Figure 4.18: Isolated-photon efficiency obtained with p–Pb simulation.

the cross-sections do not change drastically.

## 4.7 Isolated-photon Efficiency

Because the correlation functions are normalized per photon trigger, the photon efficiency cancels. In principle, a bias may be introduced if the photon efficiency varied rapidly within the photon  $p_T$  range being used (12–40 GeV/ $c$ ) but this section will show that this is negligible.

The efficiency the isolated-photon selection is shown in Figure ???. The efficiency is rather independent of  $p_T$  in the range relevant for this analysis. Less than 1% variation is observed between the high and low ranges of the energy distribution of the photon triggers (77.7% at 12 GeV/ $c$  and 78.5% at 40 GeV/ $c$ ). This level of variation has a negligible impact in the correlation analysis.

## 4.8 Comparison to ABCD method results

In this section, the results obtained with the template fit method are compared to the ABCD method results obtained by Erwann Masson for his isolated-photon analysis in p–Pb data [[Erwann](#)]. That study is performed with the same p–Pb data using the same event and cluster selections and a similar isolation criterion that uses only charged-particles (shown in the Appendix of Ref. [[Erwann](#)]).

Figure ?? shows the results compared to the results of the ABCD method. Both the results using ITS-only tracks and ITS-TPC tracks (which is more directly comparable to the ABCD result) are shown. The results are systematically below the ABCD method, but they are consistent within  $1\sigma$  systematic uncertainty for most of the  $p_T$  range.

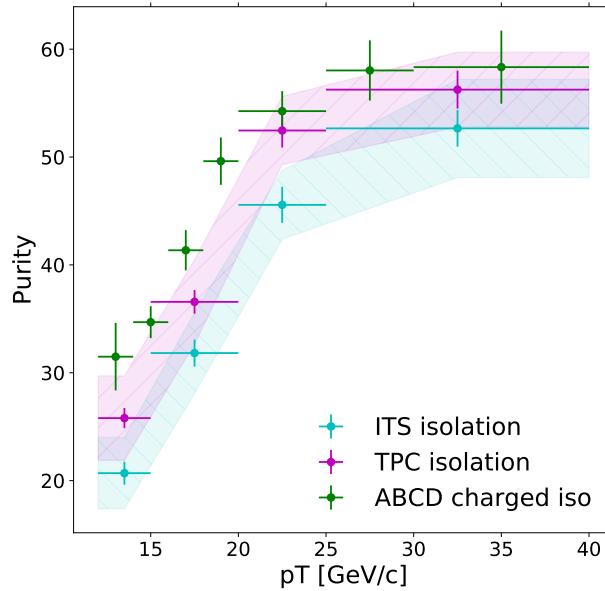


Figure 4.19: Comparison between purity measured with the template fit and ABCD method, from Ref. [Erwann], in p–Pb data. The error bars represent statistical uncertainty only and the bands represent the systematic uncertainty only.

## 4.9 Check by splitting clusters in $|\eta| < 0.4$ and $0.4 < |\eta| < 0.67$

In this analysis, only clusters with  $|\eta| < 0.67$  (Section ??) are used, with an isolation variable constructed with tracks with  $|\eta| < 0.8$  (Section ??) and a cone size of  $R = 0.4$  (Section ??). The isolation cone is thus fully contained in the tracking acceptance only for clusters with  $|\eta| < 0.4$ ; the isolation cone for clusters with  $0.4 < |\eta| < 0.67$  is only partially covered in pseudorapidity angle. Note that the tracking acceptance covers the full azimuthal angle so this is not an issue in azimuth.

To check for possible biases that this partial containment of isolation cone in pseudorapidity all measured clusters were split into two categories:  $|\eta| < 0.4$  and  $0.4 < |\eta| < 0.67$ . In principle, if the bias introduced by the lack of total coverage of the isolation cone would lead to higher background in the  $0.4 < |\eta| < 0.67$  region (background might appear less-isolated

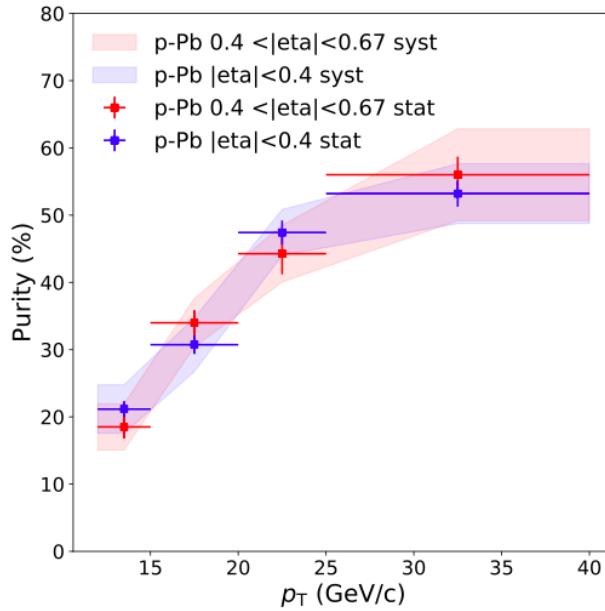


Figure 4.20: Purity measurement in p–Pb collisions for clusters with  $|\eta| < 0.4$  and  $0.4 < |\eta| < 0.67$ .

than in reality, and might pass the selection), it should thus lead to a lower purity. This study is performed in p–Pb data, which has better statistical precision than the pp data and would allow us to better constrain small biases. The comparison of purity measurements is shown in Figure ???. Both sets of measurements are compatible within statistical uncertainties. Thus, the potential bias due to incomplete coverage of isolation cone is negligible and any systematic uncertainty to this source is not assigned to the measurement. This observation may be explained by noting that the azimuthal angle is fully covered, and that most of the energy in the isolation cone is within small angles of the neutral-mesons that dominate the background (the background is primarily high-z neutral mesons in jets).

## 4.10 Cluster Selection Variations

This section studies the impact of variations in the cluster selection (Section ??) on the purity measurement. It should be noted that the purity by itself is not a physical quantity. It is entirely cut-dependant, where a higher purity often results in a lower efficiency. As a result, the variations in this analysis are not used to estimate an uncertainty on the purity.

First, the impact of the number-of-local maximum criteria for clusters is studied. The NLM < 3 selection was used in previous isolated photon analyzes (e.g. Ref [Acharya:2019jkx, Erwann]), where it was found to help to improve the simulation description of the back-

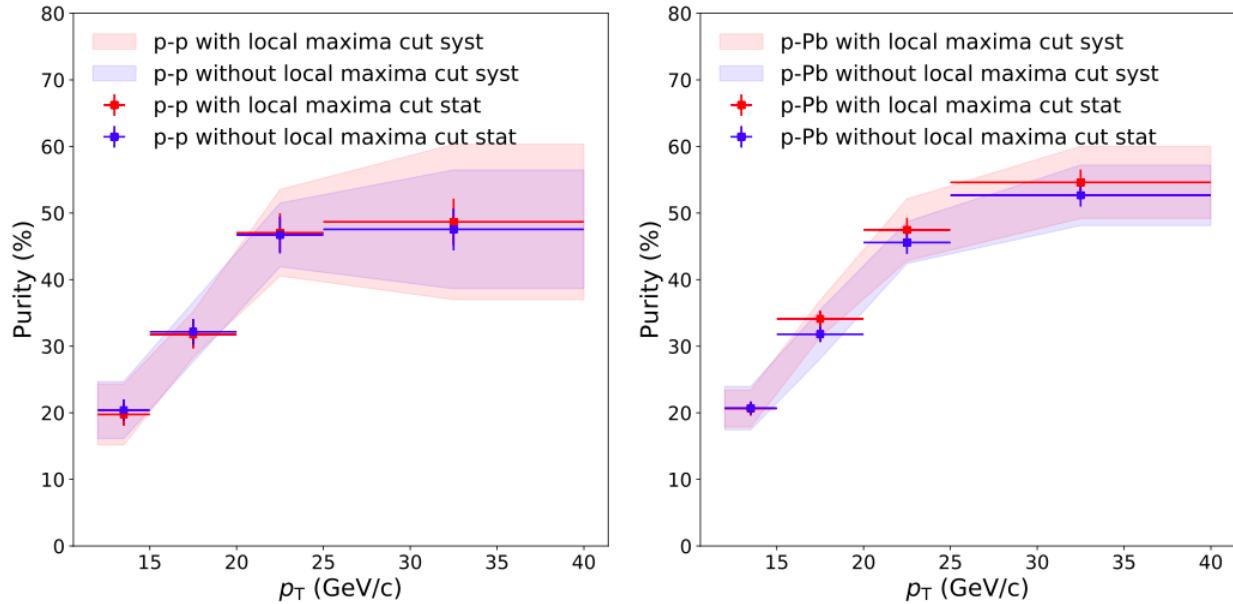


Figure 4.21: Purity measurement with and without cluster selection based on number-of-local maxima in pp (left) and p–Pb (right) collisions.

ground shower-shape. Figure ?? shows the purity measurement with and without this selection in pp and p–Pb data. No significant difference is observed.

Figure ?? shows the purity measurement by varying the distance-to-bad channel cut from the nominal  $\geq 1$  to  $\geq 2$ . This change would remove about 20% of the  $\gamma^{\text{iso}}$  candidates. Once again, no significant difference is observed.

Figure ?? shows the purity measurement with and without cluster time selection. No significant difference is observed.

Figure ?? shows the purity measurement with an exotoxicity cut with a threshold of 5% (nominal) and a variation of 3%, with no significant difference shown.

Figure ?? shows the purity measurement with different isolation thresholds. The nominal threshold of 1.5 GeV/c is varied by  $\pm 0.2$  GeV/c. As expected the higher (lower) isolation threshold results in a lower (higher) purity. However, the difference with respect to the nominal result is small with respect to the statistical and systematic uncertainties. This indicates that the purity measurement has only a weak dependence on the choice of isolation threshold.

Given that the impact of these cluster variations on the purity measurement is small compared to the estimated systematic uncertainties (described in Section ??), an additional systematic uncertainty due to cluster selection is not assigned.

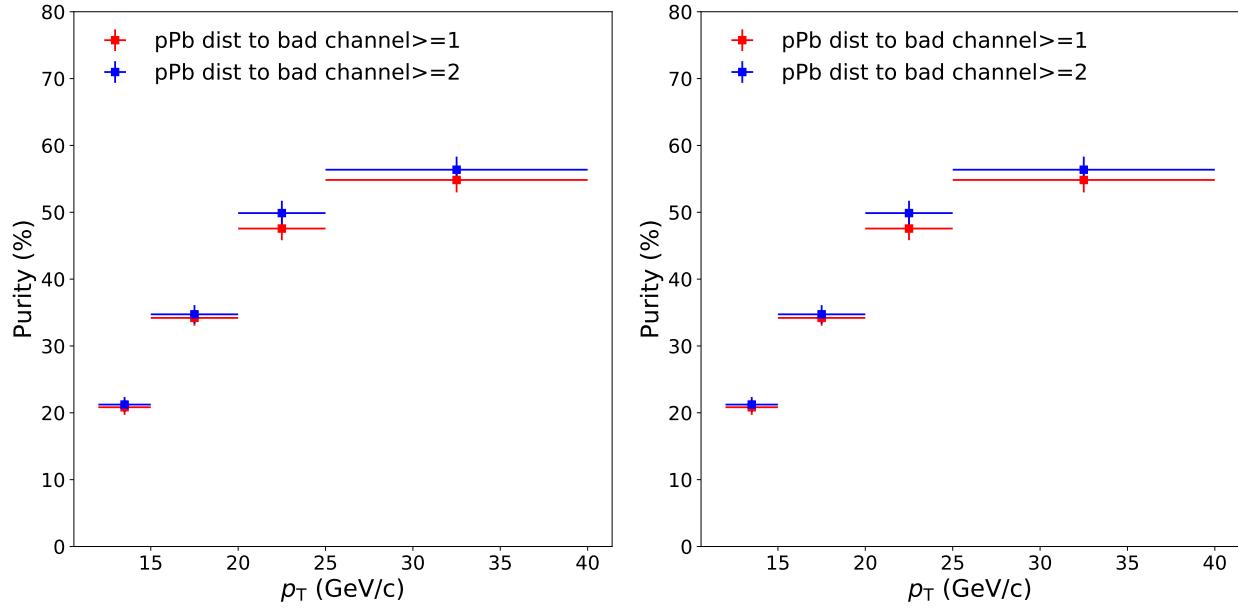


Figure 4.22: Purity measurement with distance-to-bad channel  $\geq 1$  (nominal) and  $\geq 2$  in pp (left) and p-Pb (right) data.

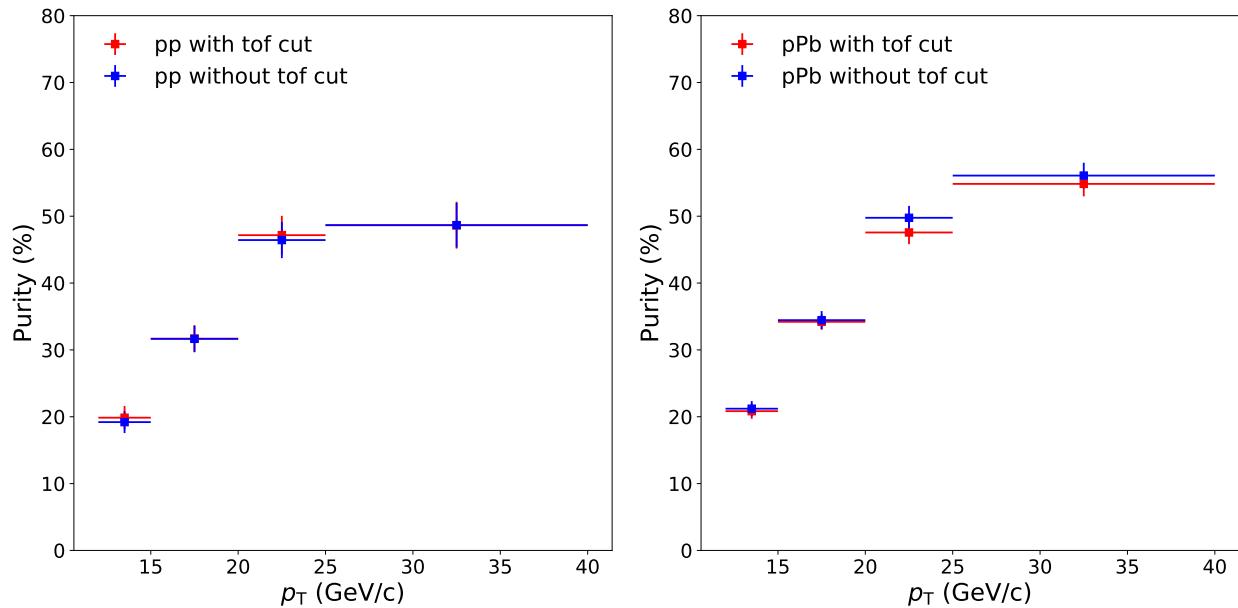


Figure 4.23: Purity measurement with and without time cut ( $\Delta t < 20$  ns) in pp (left) and p-Pb (right) data.

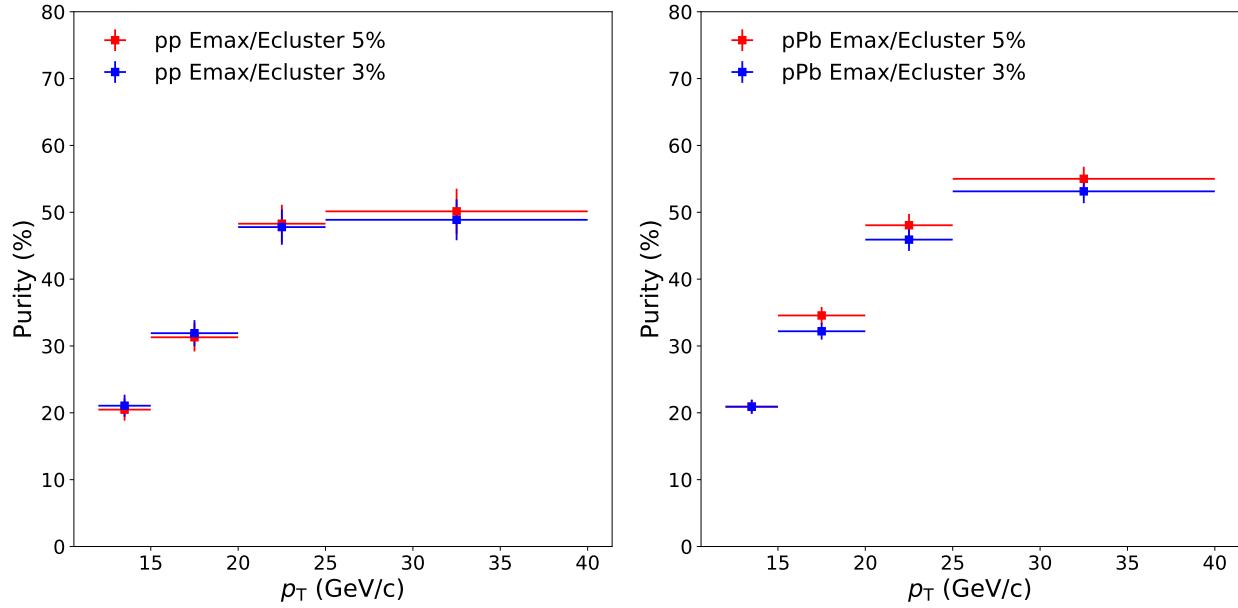


Figure 4.24: Purity measurement with an exotoxicity cut of 5% (nominal) and 3% in pp (left) and p–Pb (right) data.

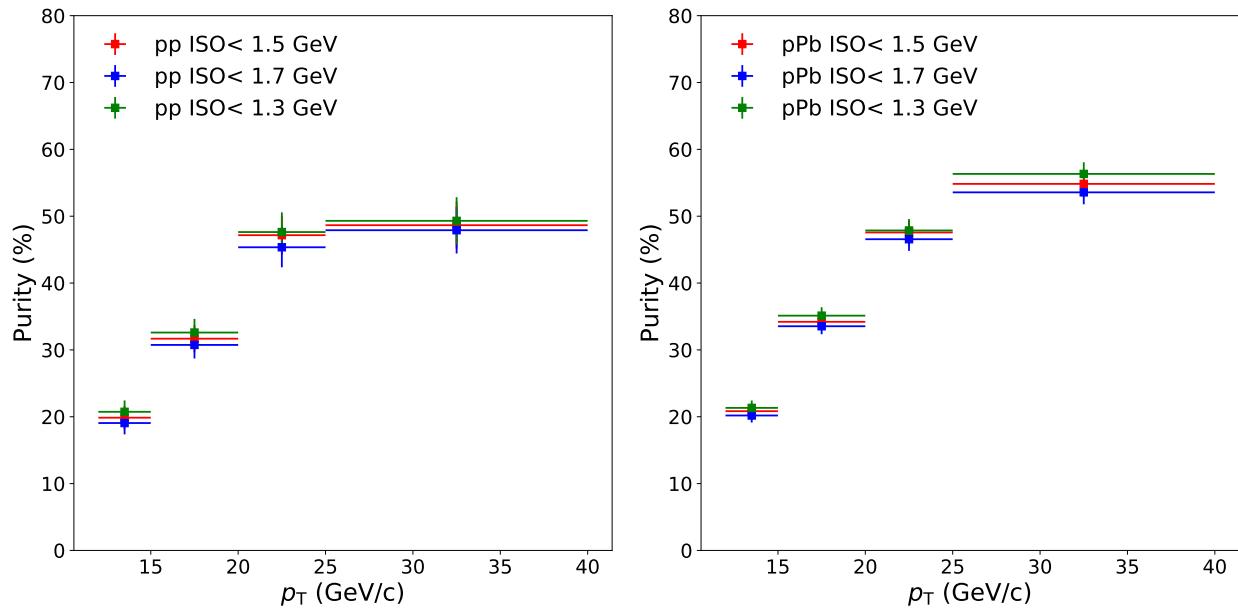


Figure 4.25: Purity measurement with different isolation threshold requirements in pp (left) and p–Pb (right) data.

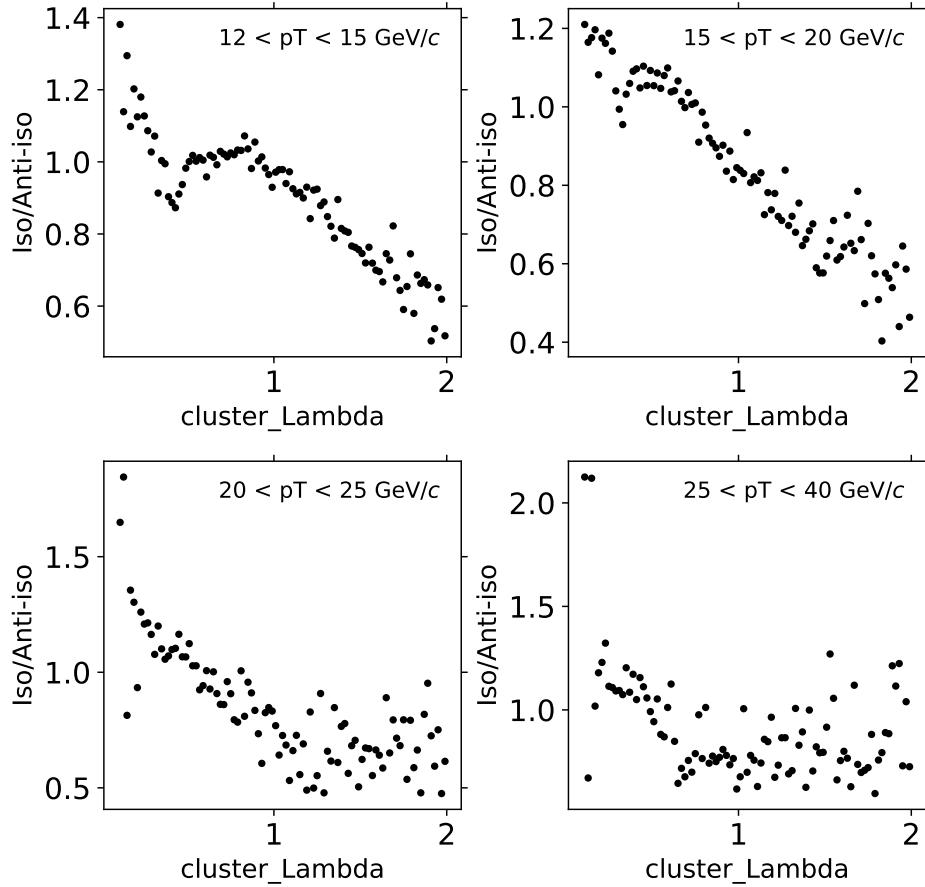


Figure 4.26: MC-based correction applied to the background shower-shape template for  $p-$   
 $Pb$  collisions in various  $p_T$  ranges.

## 4.11 MC-based correction for background template

Figure ?? shows the weight correction (see Equation ??) applied to the anti-isolation shower-shape distribution, which is obtained from dijet MC simulation. As described in more detail in Section ??, only the shape of this correction is relevant as the normalization is actually fixed in the template fit. The systematic uncertainty associated with this correction is obtained with a double-ratio using data in the background-dominated region ( $\sigma_{\text{long}}^2 > 0.4$ ), as described in more detail in Section ??.

## 4.12 Checks on Error Function Fits to Purity

The purity is fit to a 3-parameter error function, shown in Section ??, in order to capture the quickly rising behavior at low photon- $p_T$  and to avoid bin-edge effects. The use of an error function was chosen at it best represents the data as shown in Figure ??.

It is important to note that the uncertainty on the fit is not propagated as an uncertainty on correlation analysis. The uncertainty arising from the purity is discussed in detail in Section ?? . The fit is used for convenience and enables a cluster-by-cluster weighting when constructing the decay- $\gamma^{\text{iso}}$  hadron correlation, followed by the correlated subtraction. The error function agrees well within the purity uncertainties, seen in Section ?? as well as in the last panel in Figure ??.

Additionally, a check was performed to study the effect that shifting the bin-centers horizontally would have on the fit. A comparison was made between using the linear center of the bins as the bin-center and using the mean  $p_T$  for the photon-bin (this essentially shifts the bin centers towards lower values of  $p_T$  due to the quickly falling photon  $p_T$  spectrum). This comparison is shown for pp and p-Pb in Figure ??.

## 4.13 Smearing signal template

As part of the investigation into evaluating the template fit systematic uncertainty, studies were performed on "smearing" of the signal-shape template. That is, for each cluster the measured  $\sigma_{\text{long}}^2$  is smeared by multiplying it by some random number drawn from a Gaussian centered at unity with a given standard deviation "smearing width".

Figure ?? shows the purity obtained for  $15 < p_T < 20 \text{ GeV}/c$  as a function of smearing width used and the corresponding  $\chi^2/\text{DOF}$ . A rather significant deterioration of the  $\chi^2$  with increasing smearing width is observed, which indicates that smearing beyond  $\approx 3\%$  is strongly disfavoured by the data. To account for the deterioration of the  $\chi^2$  that is observed with the smearing, the statistical error of the purity is multiplied by a scaling factor of  $\sqrt{\chi^2/\text{DOF}}$ . A trend of decreasing purity with increasing smearing width is found, but this trend can be accounted for by the artificial worsening of the goodness-of-fit.

A constant is fit to the purity values vs smearing widths (shown as band in Figure ??). The error on that constant fit and its deviation of its central value from the nominal result (smearing width equal to zero) is below and absolute 1% difference, which is much smaller than the signal-template uncertainty based on the background-only template fit described in Section ?? . The same conclusion holds for all  $p_T$  ranges in both the pp and p-Pb data.

## 4.14 Checking sensitivity to $\Delta\varphi$ binning

Here, the impact of binning on the final correlation functions is quantified. The binning in principle could affect the ZYAM estimate and the shape of the correlation peak. Results are shown with double the number of  $\Delta\varphi$  bins. This is done for simplicity in order to be able

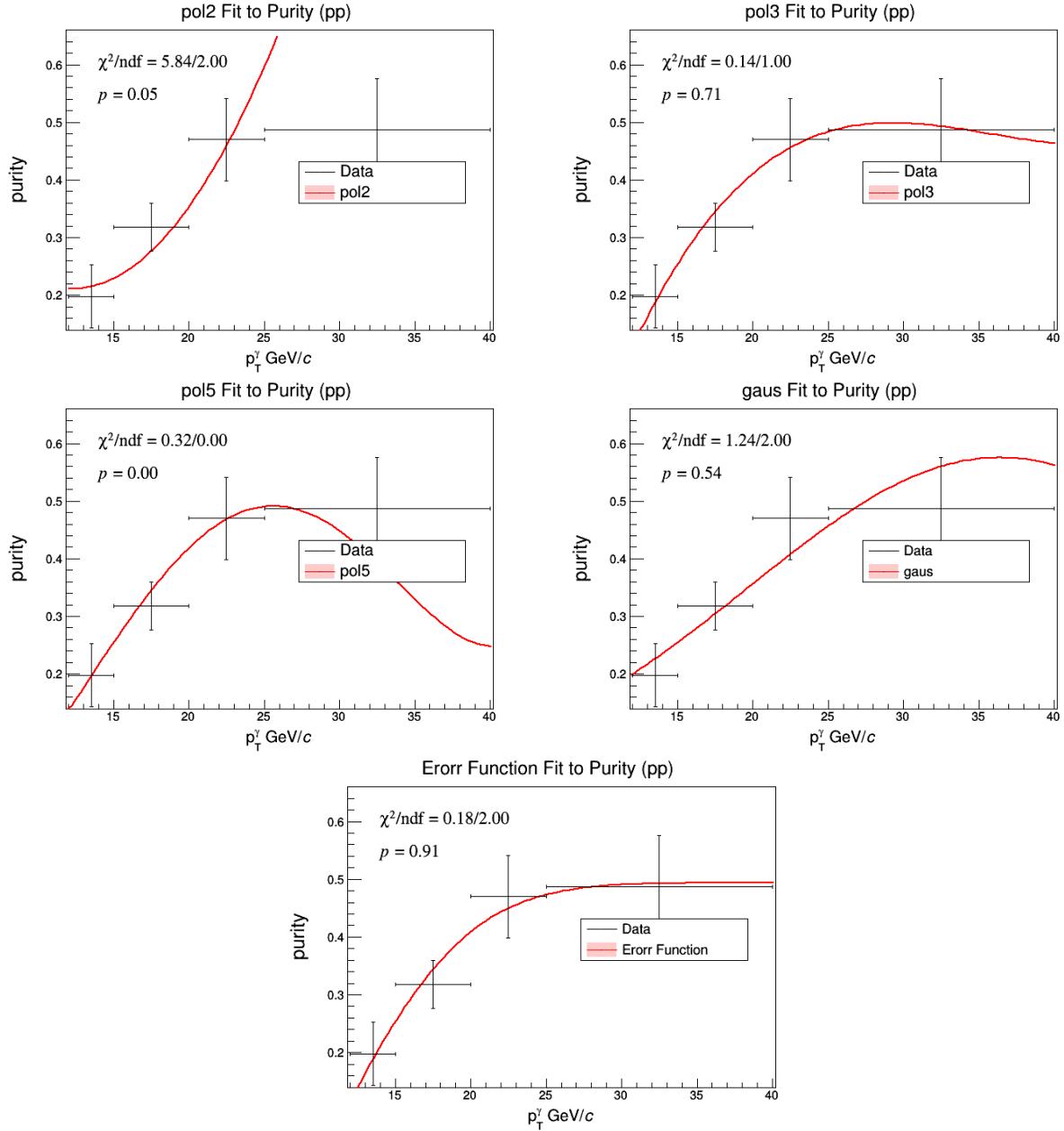


Figure 4.27: Various functions fit the purity in pp. While there is no simple physics-motivated reason to use the error-function, it was the only function to simultaneously capture the quick rise at low  $p_T$  as well as the plateau at high  $p_T$ .

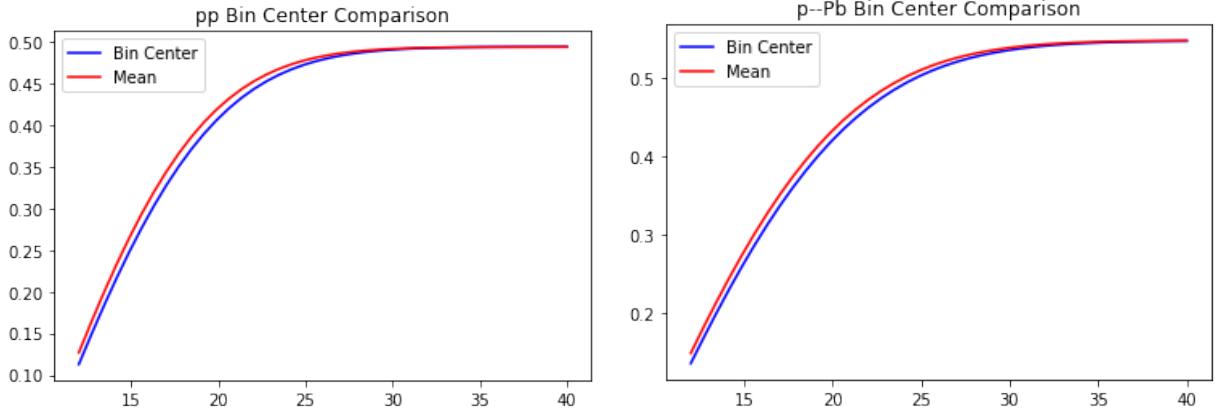


Figure 4.28: Comparison of two different bin-centers for used in the purity-error function fit. Blue indicates the fit using the linear center of the bin  $p_{\text{T}}\text{bin-width}$ . The red represents using the mean  $p_{\text{T}}$  for that  $\gamma^{\text{iso}}$   $p_{\text{T}}$ -bin. The left panel shows the comparison for pp, while the right panel shows the comparison for p-Pb. The x-axis is the  $\gamma^{\text{iso}} p_{\text{T}}$  in units of  $\text{GeV}/c$ . While a small difference is observed, it is well within the uncertainties on the purity, as well as in the fits' confidence intervals.

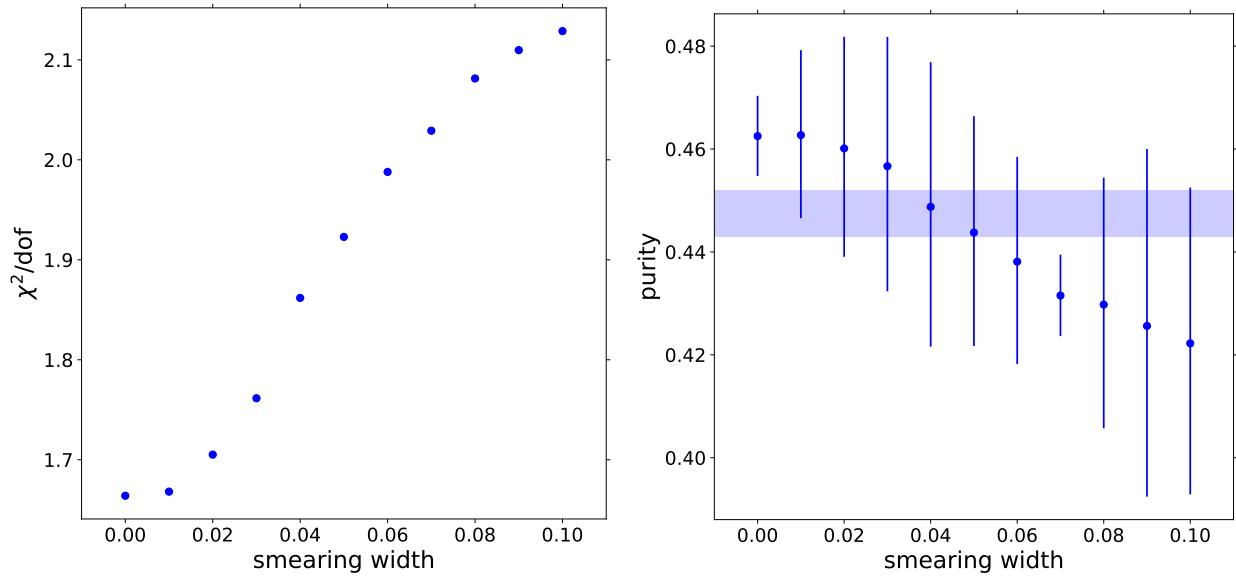


Figure 4.29: Left: Signal template distribution for various smearing widths. Right: purity vs smearing width. The error bar represents the statistical uncertainty, which has been scaled by  $\sqrt{\chi^2/\text{DOF}}$ .

to integrate the correlation in exactly the same window as reported in the nominal results ( $\Delta\varphi > 7\pi/8$ ). The integration window is varied separately from this study. Figure ?? shows the correlation function obtained with double the number of bins as the nominal result. The main conclusion of the study does not change: in every  $z_T$  bin, the pp and p–Pb data are compatible within uncertainties (as quantified in the figure by the  $\chi^2$  and corresponding p-value), and the PYTHIA8 describes the data within uncertainties.

## 4.15 Checking Sensitivity to Beam Flip

Figure ?? shows results for 13de and 13f periods. They differ in the beam configuration and have roughly similar integrated luminosity. No significant difference is observed.

## 4.16 Correlations including $\Delta\varphi = 0$

Figure ?? shows the correlation results down to 0 radians. The first bin is contained within the isolation requirement and therefore is biased, which is why is not reported in the main results.

## 4.17 Comparing ITS only tracking to ITS+TPC Tracking

This section compares results obtained using hybrid (ITS + TPC) tracks in the triggered 13def data (p–Pb). Other than the change in the appropriate tracking corrections, the analysis chain is identical to the one that uses the ITS-only tracking. Figure ?? shows the correlation functions for p–Pb that uses hybrid tracks and pp which only uses ITS tracks (the TPC was not read out during the 17q data taking).

Figure ?? Shows the fragmentation studies in p–Pb data with ITS only tracks and Hybrid tracks. It can be seen in the ratio of ITS Only/Hybrid that the hybrid tracks have a significantly lower yield at high  $z_T$ , which roughly corresponds to tracks with high  $p_T$ . This is unsurprising as there are issues with TPC charge distortions that have been documented at length [<https://alice.its.cern.ch/jira/browse/ATO-351>].

## 4.18 Checking sensitivity to $z_T$ binning

This section looks into varying the number of  $z_T$ bins. Figure ?? shows the resulting fragmentation function in 6  $z_T$  bins.

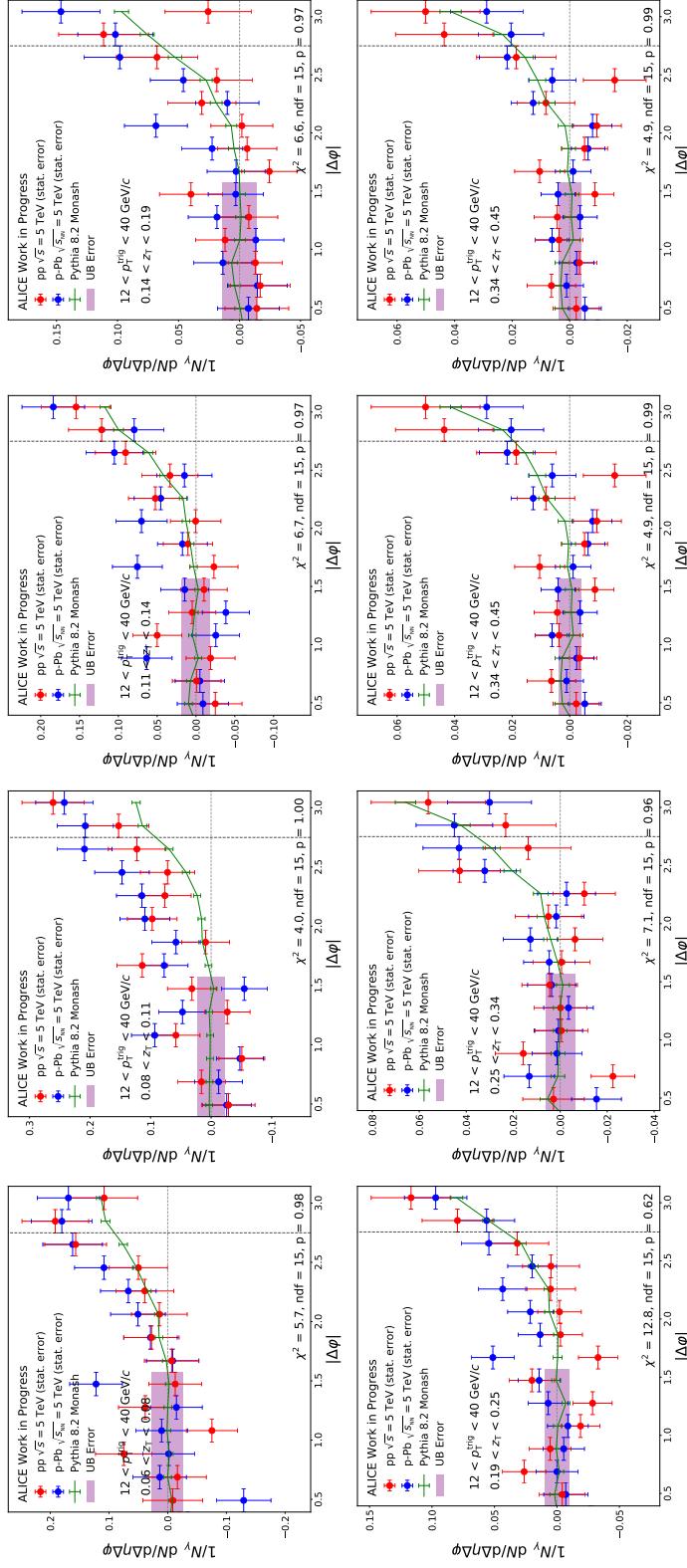


Figure 4.30: Fully-corrected  $\gamma^{\text{iso}} - \text{hadron}$  correlation function pp (red) and p-Pb (blue) data. The purple band represents the uncertainty from the underlying event estimate in pp and p-Pb. The error bars represent statistical uncertainty only. The green line is the  $\gamma^{\text{iso}} - \text{hadron}$  correlation function obtained with PYTHIA 8.2. The key difference with the nominal results shown in Section ?? the number of  $\Delta\varphi$  bins is doubled.

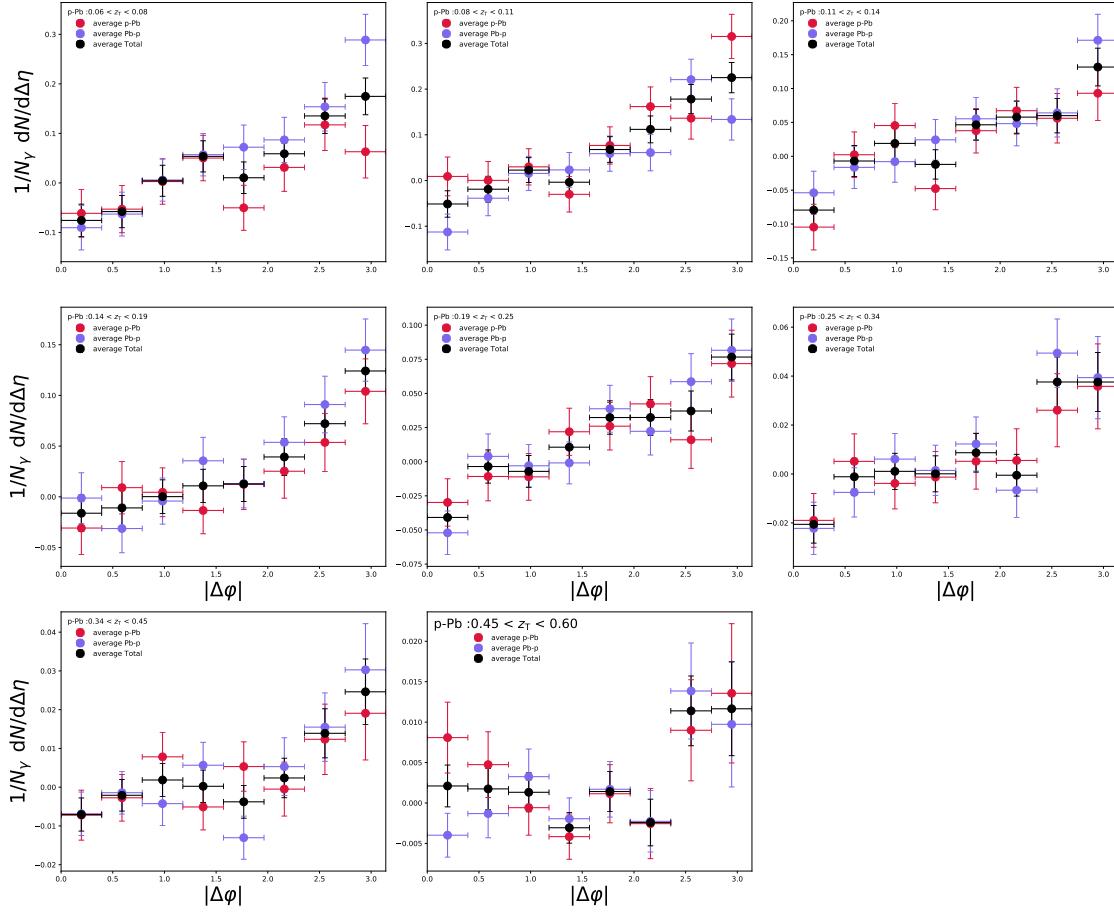


Figure 4.31: Comparison of final correlations functions for p–Pb (red) and Pb–p (blue) collisions.

## 4.19 Checking sensitivity to pile up cut in pp

A stringent pile up cut is applied to the pp dataset. Figure ?? shows the correlation functions in 6  $z_T$  bins to compensate for the loss of statistics in pp after applying the cut.

Figure ?? shows the resulting fragmentation function in 6  $zt$  bins with the pile up cut applied to the pp data. The figure shows no significant change as a result of this cut.

## 4.20 Checking Maximum Track $p_T$ Cut

This section explores the effect of varying the maximum  $p_T$  cut on the final away-side yields in pp and p–Pb and the effect of this variation on the final ratio is shown.

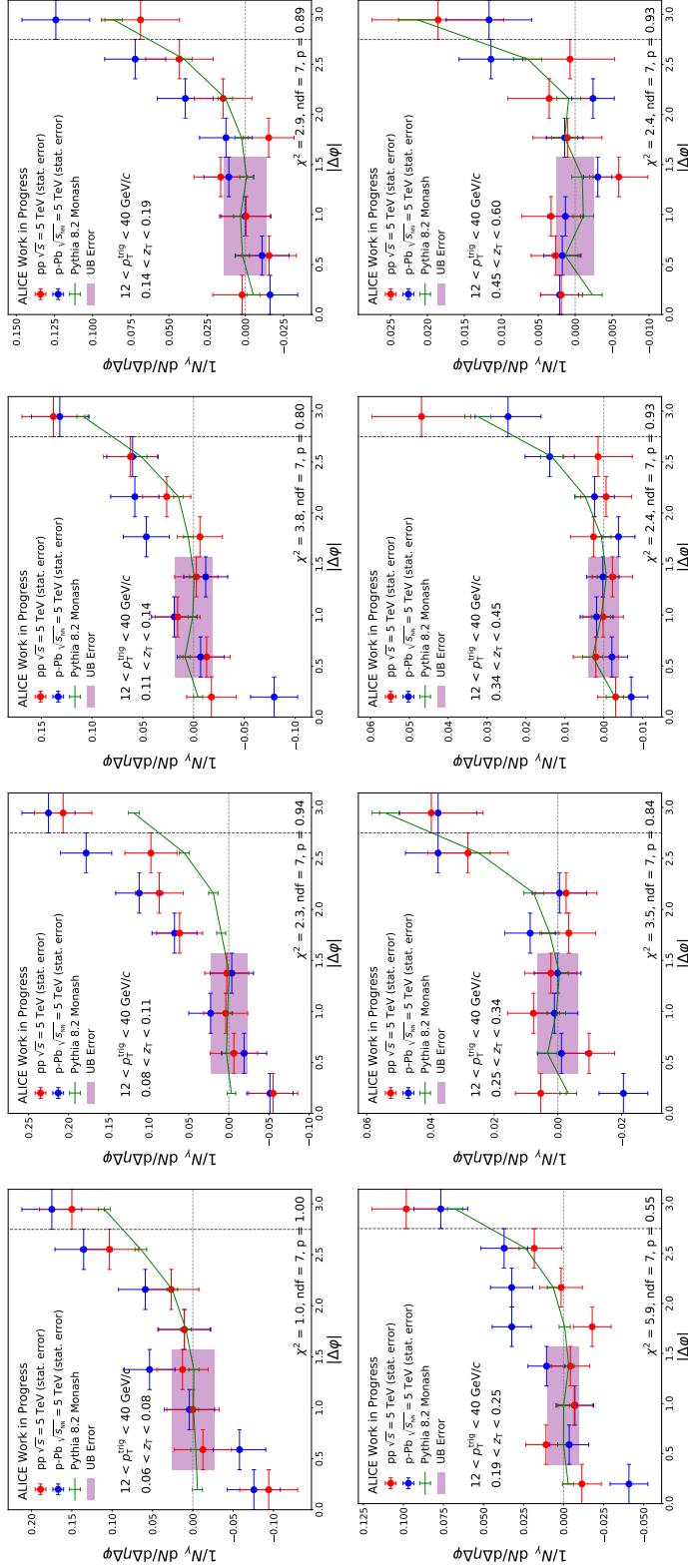


Figure 4.32: Fully-corrected  $\gamma^{\text{iso}}\text{-hadron}$  correlation function  $\text{pp}$  (red) and  $\text{p-Pb}$  (blue) data. Identical to Figure ?? with the exception that the correlation functions down to  $\Delta\varphi=0$  are shown.

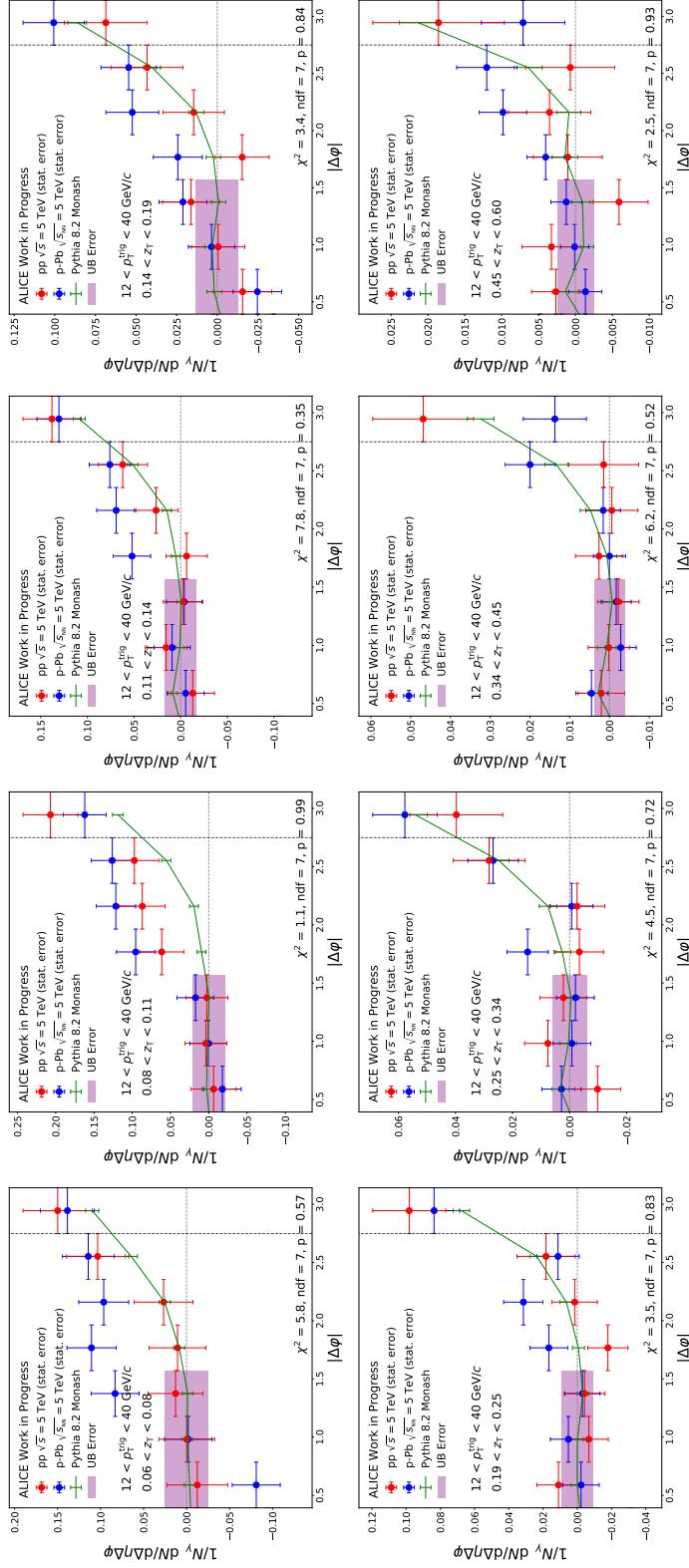


Figure 4.33: Fully-corrected  $\gamma^{\text{iso}}$ -hadron correlation function in pp using ITS only tracks (red) and p-Pb using hybrid tracks (blue).

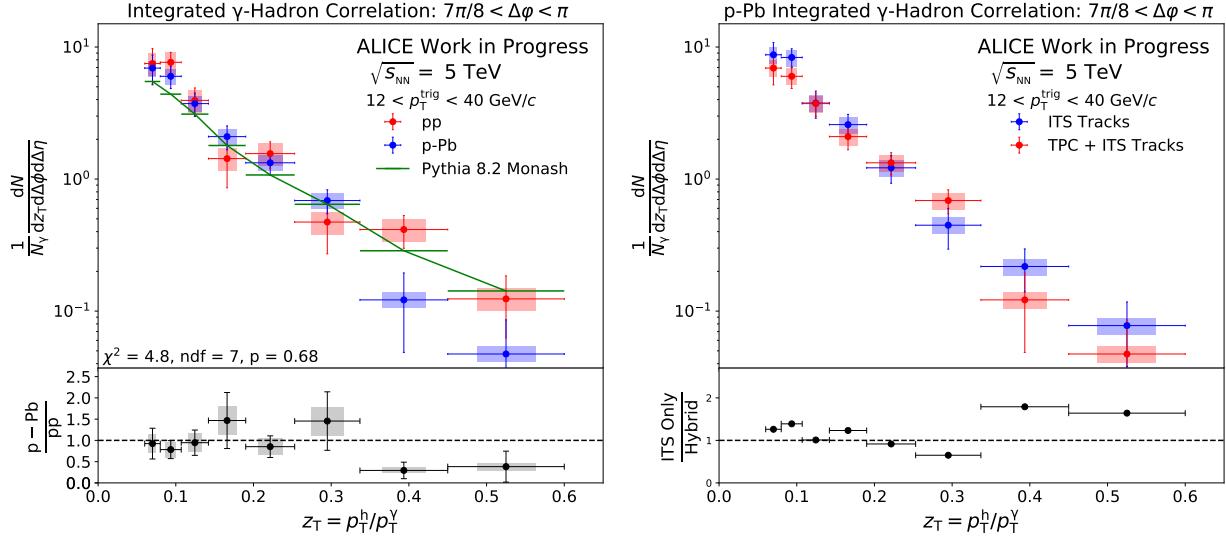


Figure 4.34: Left: Comparison of final correlations functions for p–Pb Hybrid tracks (blue) and pp ITS only tracks (red). Right: Comparison of final correlations functions for p–Pb Hybrid tracks (red) and p–Pb ITS only tracks (blue).

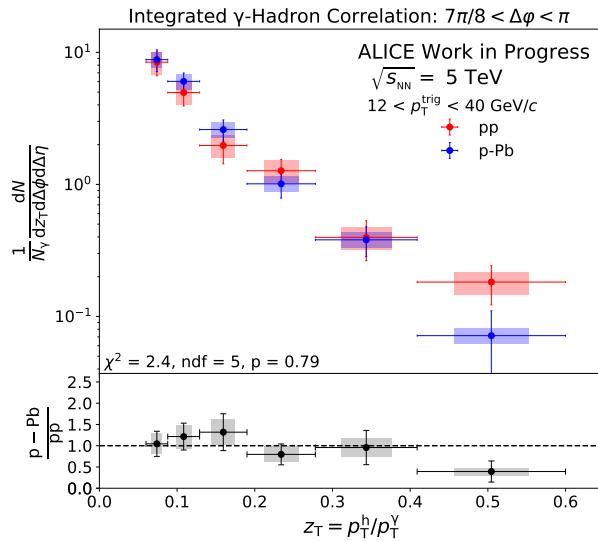


Figure 4.35: Results with variation of numbers of  $z_T$  bins.

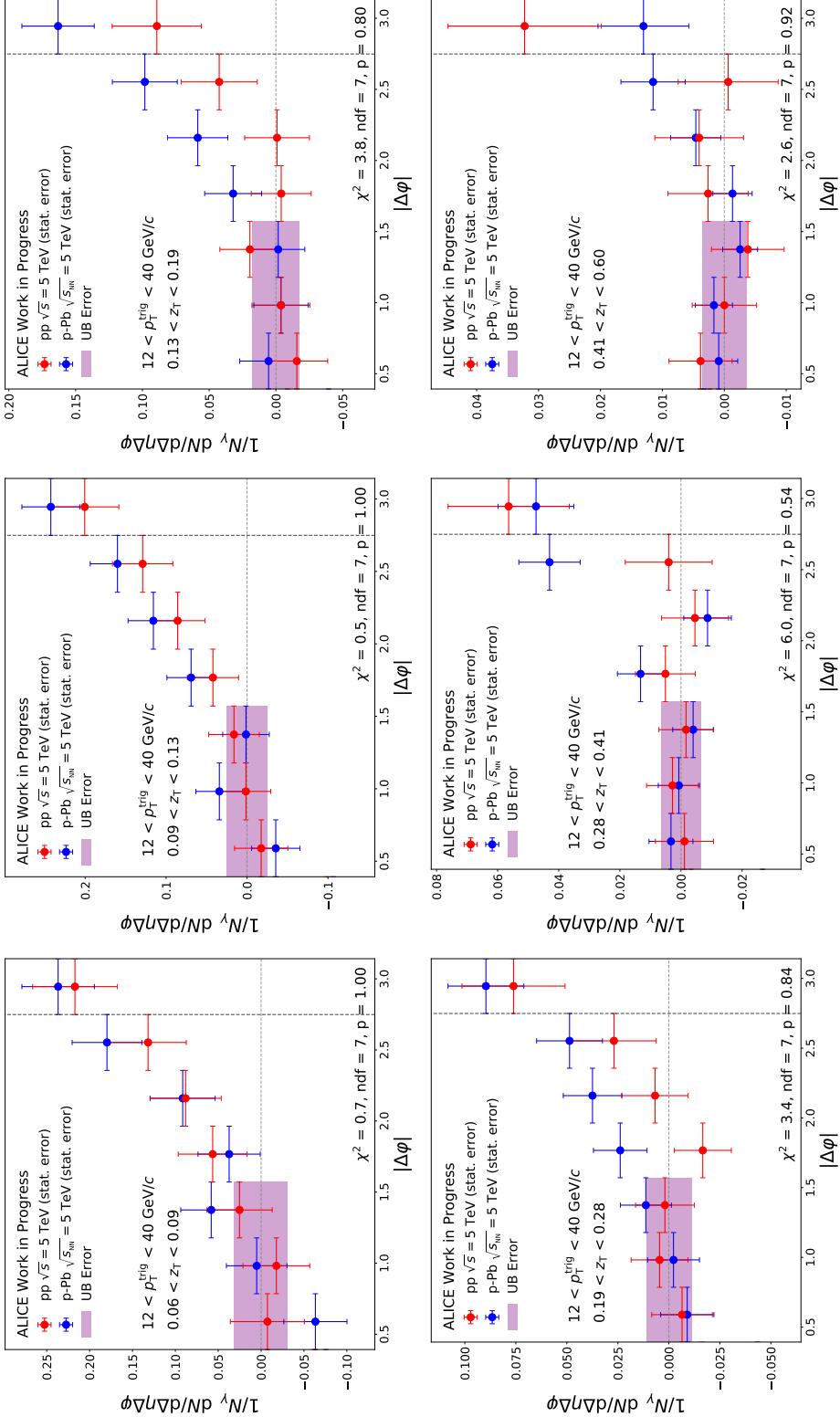


Figure 4.36: Fully-corrected  $\gamma^{\text{iso}}$ -hadron correlation function in pp (red) and p-Pb (blue) data with the pile up cut in pp only. The correlations have been rebinned to 6 zT bins to compensate for the loss in statistics after applying the cut.

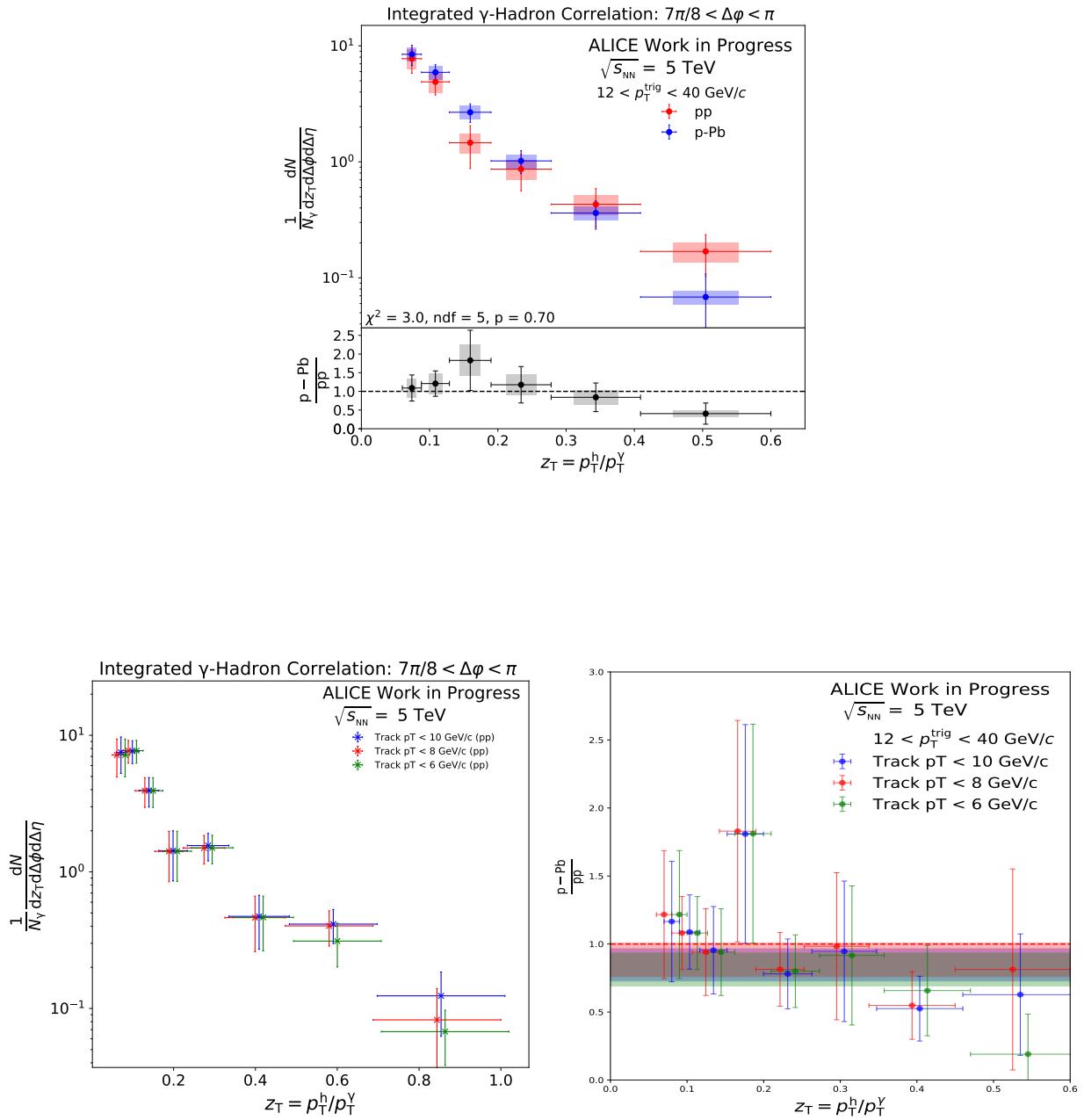


Figure 4.37: Left: The away side yields in pp with a maximum  $p_T^{\text{Track}}$  cut of 10, 8, and 6  $\text{GeV}/c$ . Right: The Ratio of p-Pb to pp with a maximum  $p_T^{\text{Track}}$  cut of 10, 8, and 6  $\text{GeV}/c$

## 4.21 Sensitivity to $\chi^2/\text{ITS}_{\text{clus}}$ in p-Pb

The effect of the  $\chi^2/\text{ITS}_{\text{clus}}$  cut on the tracking efficiency, fake rate, and bin migration is studied to select the most effective  $\chi^2/\text{ITS}_{\text{clus}}$  maximum limit. Four values have been tested:  $\chi^2/\text{ITS}_{\text{clus}} = 1, 2, 3$ , and  $36$ . The following figures ??, ??, ??, and ?? show the effect of  $\chi^2/\text{ITS}_{\text{clus}}$  cuts on the efficiency, fake rate, bin migration, and track correction weight. The weight is calculated from the efficiency, fake rate, and bin migration using equation ???. The efficiency seems to be similar for all cut variation except for  $\chi^2/\text{ITS}_{\text{clus}} \geq 1$ , where it is lower by 5%. However, the fake rate and bin migration effects are improved as the cut is tightened. It is to note that regardless of the threshold for the  $\chi^2/\text{ITS}_{\text{clus}}$  cut, as long as identical cuts are applied to both the Monte Carlo and the data sets used in the section ??, the fine agreement with published data is always present.

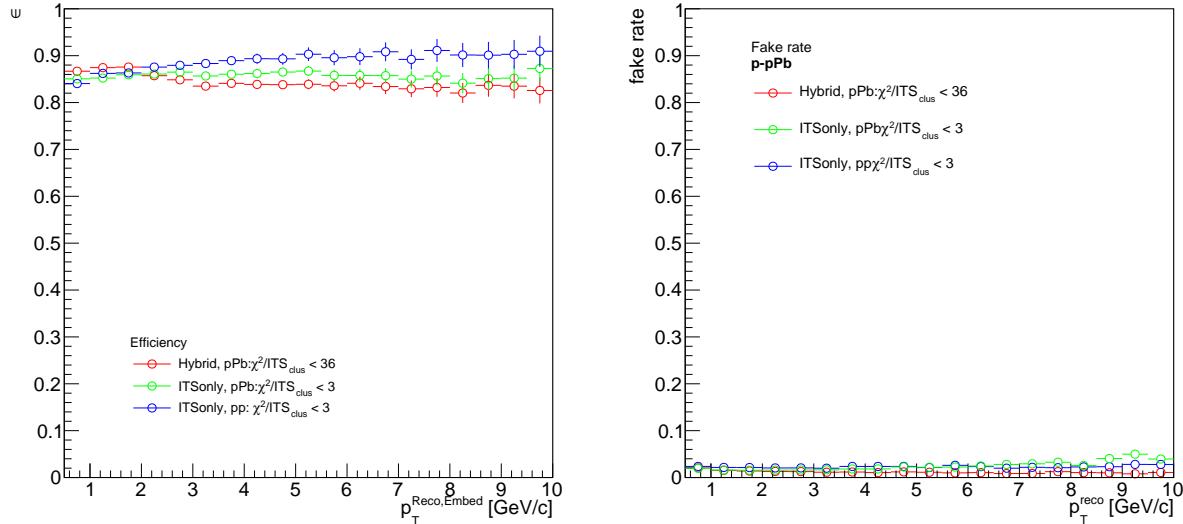


Figure 4.38: Left: The efficiency comparison for the various  $\chi^2/\text{ITS}_{\text{clus}}$  cuts. Right: The fake rate comparison for the various  $\chi^2/\text{ITS}_{\text{clus}}$  cuts

Additionally, figure ?? shows the effect of changing the  $\chi^2/\text{ITS}_{\text{clus}}$  from  $\chi^2/\text{ITS}_{\text{clus}} < 36$  to  $\chi^2/\text{ITS}_{\text{clus}} < 3$  on the track  $p_T$ spectra as the ratio of  $\chi^2/\text{ITS}_{\text{clus}} < 3$  and  $\chi^2/\text{ITS}_{\text{clus}} < 36$ . From the top two plots, one can see that the change in the cut does not affect the  $p_T$ spectra as the ratio is consistent with unity within 5% up to 8 GeV/ as along as the appropriate Monte Carlo anchored to the data is used. However, the bottom plots, uses LHC13def, but corrects using the 13b2 Monte Carlo, and divergence from unity shows the instability of the  $\chi^2/\text{ITS}_{\text{clus}}$  cut. Thus, the 13b2 Monte Carlo is not used in place of using the 17g6a1 simulation to correct the 13def data set.

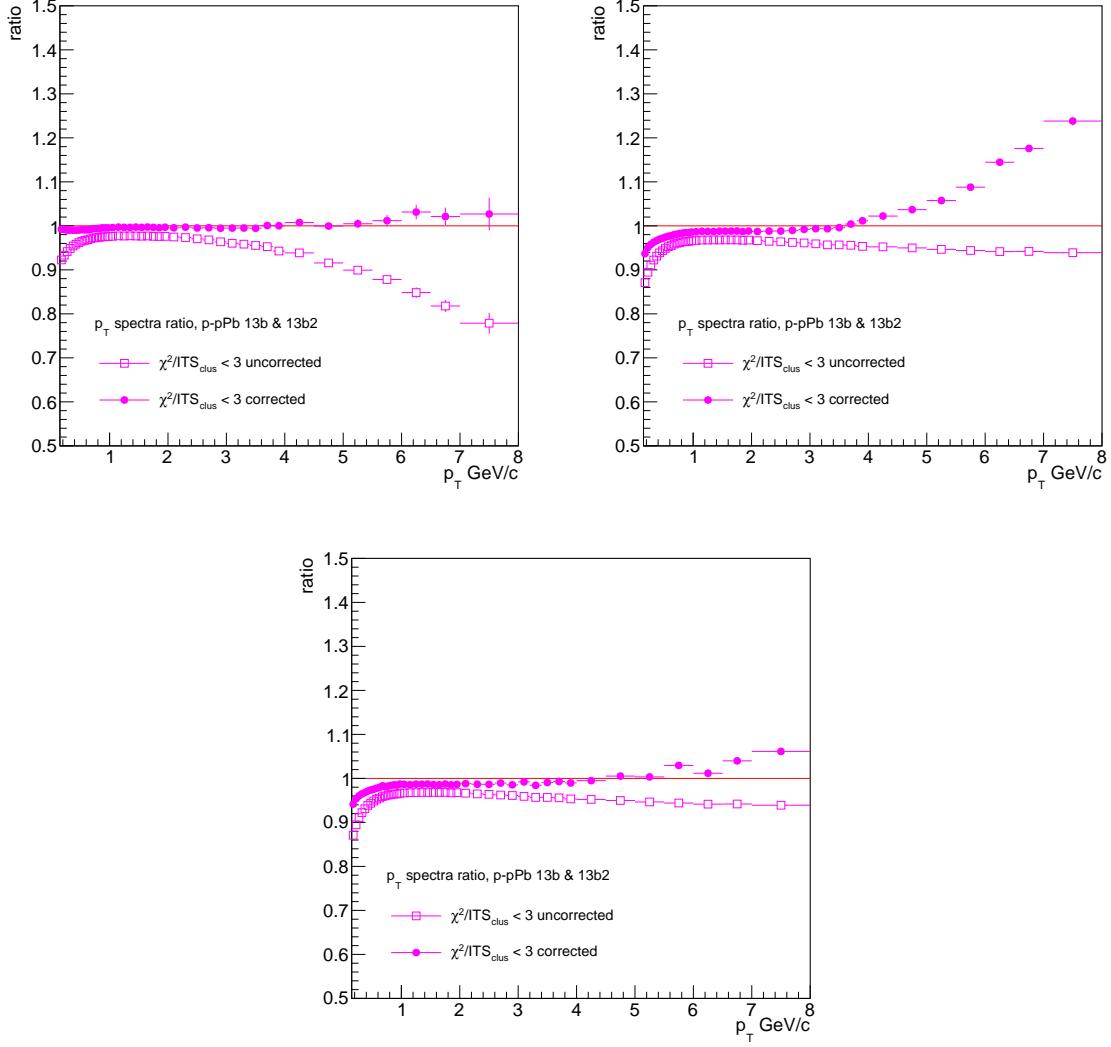


Figure 4.39: The changes in the tracking spectra with respect to the  $\chi^2/ITS_{clus}$  cut. The top left is the 13b data corrected using 13b2 Monte Carlo. The top right is 13def data corrected using 17g6a1 Monte Carlo. And the bottom plot is 13def data corrected using 13b2 Monte Carlo. [10.1007/978-3-642-18211-2\_166]

Finally, the efficiency, fake rate, bin migration and the weights are compared between ITS-only tracking in p-Pb and pp, and TPC+ITS tracking in p-Pb using 17g6a1 MC dataset. All the TPC+ITS and ITS-only tracking are comparable within uncertainties, while the ITS-only pp and p-Pb follow similar shapes. This result agrees with acceptations and trend demonstrated in section ???. The key difference is that in section ?? all the figures are

produced using 13b2, the minimum bias Monte Carlo, while in the case for figure ??, 17g6a1, the gamma-jet Monte Carlo is used.

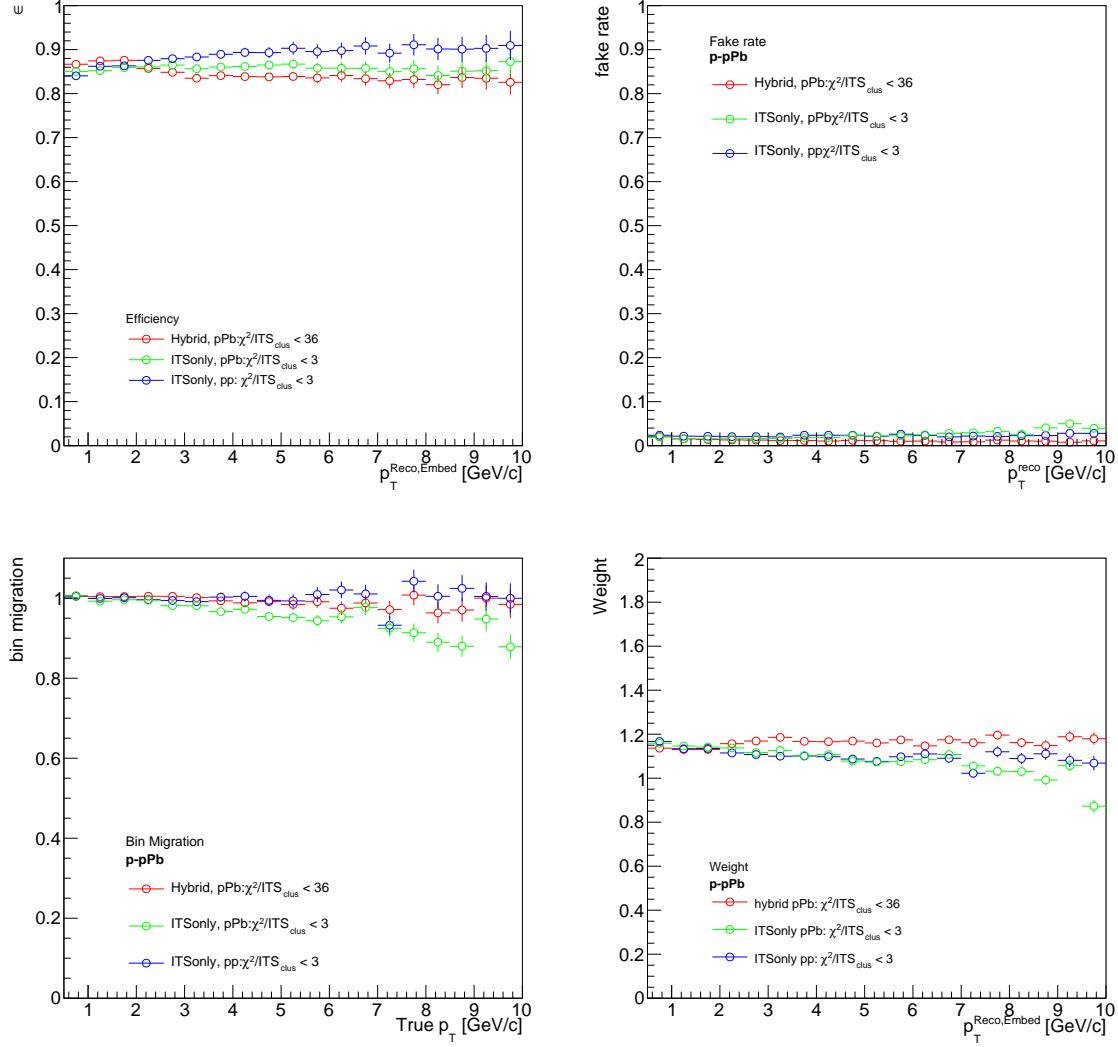


Figure 4.40: The comparison of efficiency (top left), fake rate (top right), bin migration (bottom left), and weights (bottom right) between ITS-only tracking for p-Pb and pp, as well as TPC+ITS tracking for p-Pb using the 17g6a1 gamma-jet Monte Carlo.[[PhysRevC.101.065204](#)]

## 4.22 Summary of Systematic Uncertainties

The source of systematic uncertainties of the  $\gamma^{\text{iso}}$ -hadron measurement are the following:

- Purity

The uncertainty of the purity measurement, is propagated to the correlation function measurement following Equation ???. The resulting uncertainty on the correlation function is a relative  $\pm 18\%$  for pp data and  $\pm 12\%$  for p–Pb data . As described in Section ??, a large fraction of the purity total uncertainty is either statistical uncertainty or systematic uncertainties that arise due to limited data sample. Therefore, the purity uncertainty in pp and p–Pb data are largely uncorrelated. As a conservative approach, they are taken to be totally uncorrelated.

- Underlying Event:

As described in Section ??, the uncertainty of the UE subtraction originates from statistical fluctuations in the ZYAM estimate. It propagates directly to the per-trigger yields. It ranges from 7% to 15% depending on the  $z_T$  bin and data. This uncertainty is fully correlated in  $\varphi$  for a given  $z_T$  bin, but totally uncorrelated among  $z_T$  bins, and totally uncorrelated between pp and p–Pb datasets.

- Tracking performance :

To estimate the systematic uncertainty of the charged-particle  $p_T$  measurement with ITS-only track reconstruction, MC simulation studies were performed compared with the published  $p_T$  spectra that used the ALICE standard tracking (i.e. including TPC) in pp and p–Pb collisions at 5 TeV [Acharya:2018qsh]. As described in Section ??, the combined uncertainty due to track efficiency, fake rate, and bin-to-bin migration corrections amounts to  $\pm 5\%$  added in quadrature with the total systematic uncertainty of the reference  $p_T$  spectra, which ranges from a relative 1.6 (2.1%) to 1.9% (2.5%) in the range  $0.5 < p_T < 10 \text{ GeV}/c$  for pp (p–Pb) collisions [Acharya:2018qsh].

Systematic uncertainties due to secondary-particle contamination and from modelling of the particle-type composition in MC simulations are small ( $< 2\%$ ) for the range  $0.5 < p_T < 10 \text{ GeV}/c$ . These were already estimated in Ref.[Acharya:2018qsh] for pp and p–Pb data sets and already included in the systematic uncertainty estimate described above. The tracking performance between pp and p–Pb datasets is very similar, but as a conservative approach the systematic uncertainties are taken to be completely uncorrelated.

- Acceptance mismatch due to boost:

As described in Section ??, our PYTHIA8 study of  $\gamma^{\text{iso}}$ -hadron correlations show that the impact of an acceptance mismatch between pp and p–Pb data that arises from the boost of  $\Delta y = 0.47$  amounts to about 5% effect irrespective of  $z_T$ . This estimate is subject to PDF uncertainties, which are the ones that dictate the shape of the differential cross-section of photons and associated hadrons in pseudorapidity. We chose

to not apply any correction for this effect, and assign a  $\pm 5\%$  systematic uncertainty on the per-trigger hadron yields. This systematic uncertainty is taken to be completely correlated with  $z_T$ . We assign this systematic uncertainty to our p–Pb measurements only.

- Luminosity, trigger, photon, and vertex reconstruction:

The observable is normalized per measured photon. Therefore the uncertainties related to overall normalization of the  $\gamma^{\text{iso}}$   $p_T$  spectra (such as luminosity scale, vertexing efficiency, trigger efficiency and photon reconstruction efficiency) cancel completely. Consequently, A systematic uncertainty associated with these sources is not applied to the measurement.

- Photon energy scale, resolution and material budget:

While the measurement is by construction totally insensitive to overall normalization, it is principle sensitive to bin-migration or scale uncertainties that affect the shape of the photon  $p_T$  spectra. This potential systematic uncertainty is reduced because of the large integration window over wide photon  $p_T$  range (12–40 GeV/ $c$ ). Moreover, the EMCal performance is such that these effects are small; for a 12 GeV cluster the resolution  $\sigma_E/E = 4.8\%/E \otimes 11.3\%/\sqrt{E} \otimes 1.7\%$  yields  $\sigma_E/E = 3.6\%$ , and at 40 GeV this yields  $\sigma_E/E = 2.4\%$ . The EMCal energy scale has been studied with beam-test data [[Allen:2009aa](#)] and comparison of  $\pi^0 \rightarrow \gamma\gamma$  events in data and simulation [[Adam:2016khe](#)], and has an associated uncertainty of 0.8%.

The uncertainties due to photon energy scale, resolution, and material budget have been estimated for the isolated photon cross-section measurement with 7 TeV pp and 5 TeV p–Pb data and are less than 3% in the  $p_T$  range covered in this analysis [[Erwann, Acharya:2019jkx](#)]. The effects on the per-trigger correlation functions would be even smaller. Given that this level of uncertainty is much smaller than other sources of systematic uncertainties for our measurement, they are neglected.

Table ?? presents as summary of all uncertainty estimates for our  $\gamma^{\text{iso}}$ –hadron correlation measurement. Table ?? shows the uncertainty in the p–Pb to pp ratio.

Table 4.4: Summary of uncertainties in  $\gamma^{\text{iso}}$ -hadron correlations, which are reported as per-trigger yields of correlated hadrons. The uncertainties quoted are relative. The ranges shown encompass the relative uncertainties for hadron  $z_T$  in 2 ranges: Low (0.06–0.18) and High (0.18–0.6)  $z_T$ .

	Low $z_T$ pp data	High $z_T$ pp data	Low $z_T$ p–Pb data	High $z_T$ p–Pb
Statistical Uncertainty	19%–40%	30%–49%	16%–23%	29%–39%
Photon Purity	18%	18%	12%	12%
Underlying Event	8%–15%	7%–12%	7%–10%	9%–9%
Tracking performance	5.6%	5.6%	5.6%	5.6%
Acceptance mismatch	–	–	5%	5%
Photon Energy Scale	<1%	<1%	<1%	<1%
Photon Energy Resolution	<1%	<1%	<1%	<1%
Material budget	<1%	<1%	<1%	<1%
Luminosity scale	–	–	–	–
Vertex efficiency	–	–	–	–
Trigger corrections	–	–	–	–
Photon reconstruction efficiency	–	–	–	–
Total Systematic Uncertainty	21%–24%	20%–22%	15%–16%	16%–16%
Total Uncertainty	28%–47%	34%–54%	22%–28%	31%–47%

# Chapter 5

## First Photon-Tagged Fragmentation Results in p–Pb

### 5.0.1 Fully Subtracted Correlation Functions

The final  $\gamma^{\text{iso}}$ -hadron correlations are reported in  $z_T$  bins for each trigger-photon  $p_T$  bin, where  $z_T$  is the ratio of the associated hadron,  $p_T^h$ , to isolated photon transverse momentum,  $z_T = p_T^h/p_T^{\gamma^{\text{iso}}}$ . The fully subtracted azimuthal correlations as a function of  $\Delta\varphi$ , the azimuthal angle between the photon and the hadron, are shown in Fig. ?? for pp and p–Pb data. With the measured  $\gamma^{\text{iso}}$  constraining the parton kinematics, the distribution of away-side associated hadrons with momentum fraction  $z_T$  represents the fragmentation function of the parton.

The darker colored bands at zero represents the uncertainty from the uncorrelated background estimate. The vertical bars indicate the statistical uncertainty only. The final correlation functions in each collision system demonstrate similar behavior: both show a signal consistent with zero at small  $\Delta\varphi$ , and a rising away-side peak at large  $\Delta\varphi$  arising predominantly from the hard-scattered parton opposite to the trigger photon.

Agreement within uncertainties between pp, p–Pb, and the PYTHIA 8.2 Monash Tune is observed. By measuring associated hadrons, correlations can be observed at much larger angles than would otherwise be possible for hadrons within a reconstructed jet. A  $\chi^2$  test between pp and p–Pb data and a p-value is calculated in each  $z_T$  bin for the null hypothesis that pp and p–Pb data follow the same true correlation function. In each bin, the null hypothesis cannot be rejected, indicating that there is no significant difference between the correlation functions in the two collision systems.

### 5.1 Parton Fragmentation Function

Similar to the methods described in Sec. 1.8.1 and Sec. 1.8.2, the correlation functions from Fig. ?? are then integrated in the region  $|\Delta\varphi| > \frac{7\pi}{8}$  for each  $z_T$  bin in order to obtain the

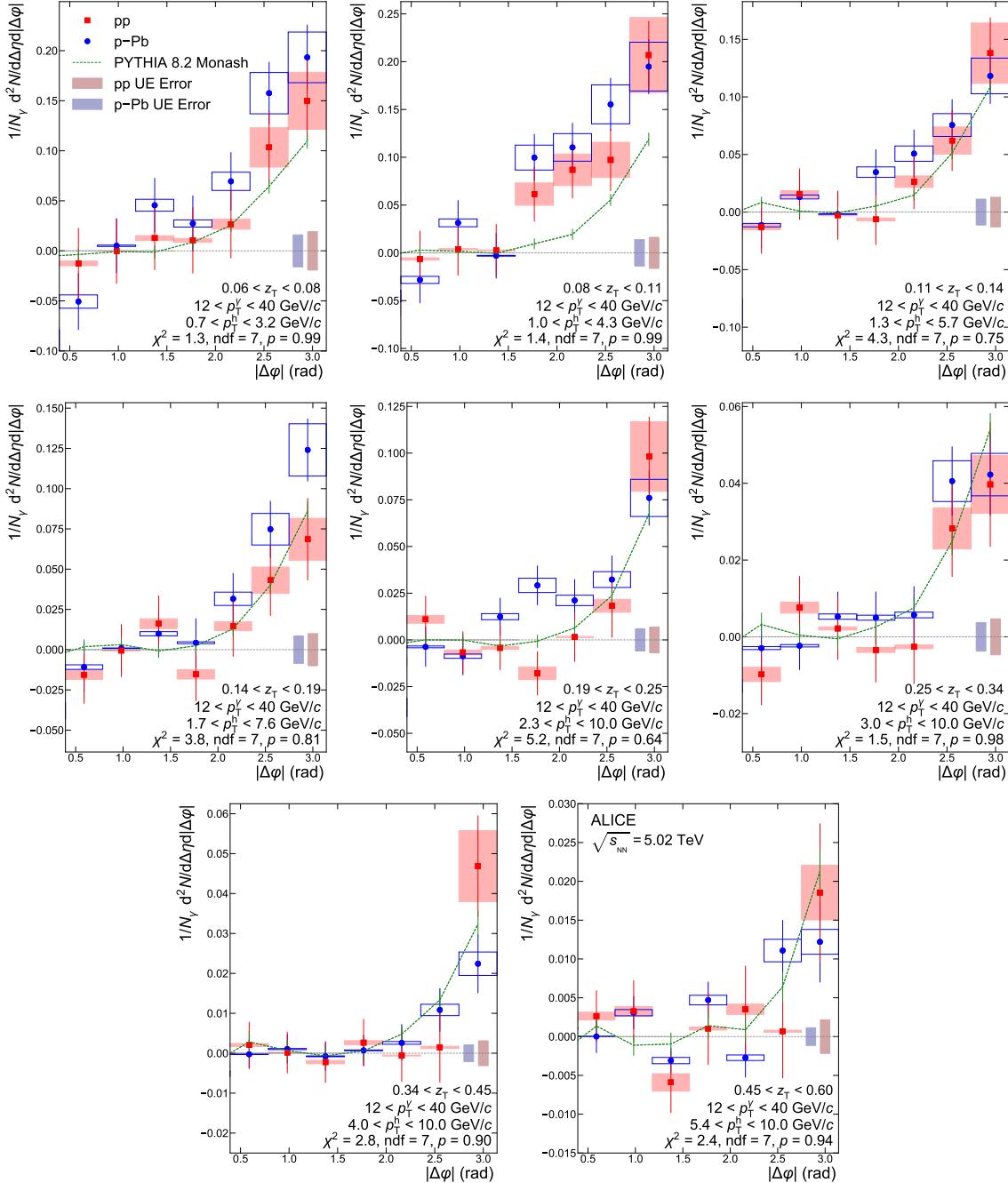


Figure 5.1:  $\gamma^{\text{iso}}\text{-hadron}$  correlation functions for pp (red) and p–Pb (blue) data at  $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$  as measured by the ALICE detector. The different panels represent three different  $z_T$  bins. The correlation functions are projected over the range  $|\Delta\eta| < 1.2$ . The darker bands at zero represents the uncertainty from the underlying event estimation in pp and p–Pb. The underlying event was estimated over the range  $0.4 < |\Delta\varphi| < 1.6$ . The vertical bars represent statistical uncertainties only. The boxes indicate the systematic uncertainties. The dashed green line represents the  $\gamma^{\text{iso}}\text{-hadron}$  correlation function obtained with PYTHIA 8.2 Monash Tune. “ $p$ ” is the p-value for the hypothesis that the pp and p–Pb data follow the same true correlation function.

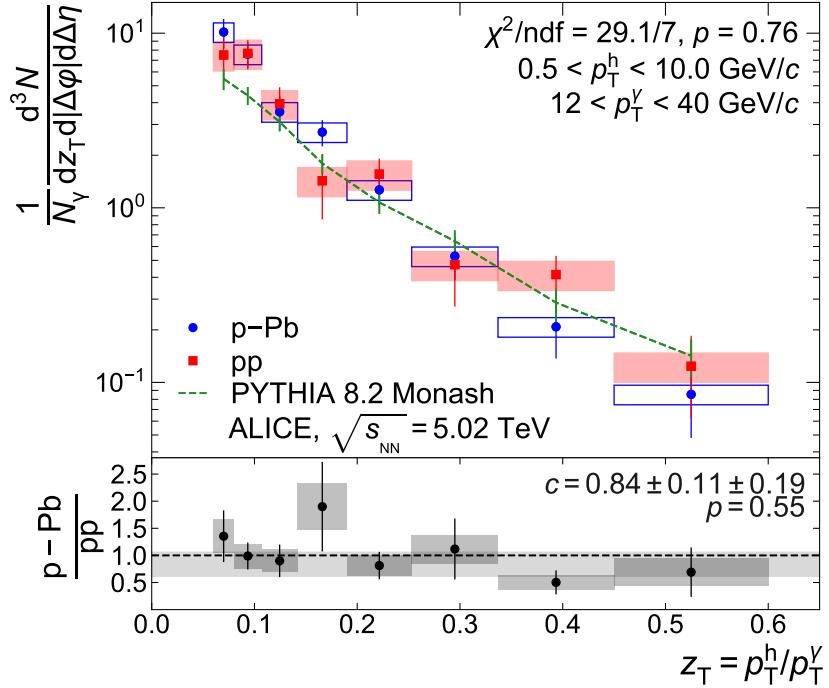


Figure 5.2:  $\gamma^{\text{iso}}$ -tagged fragmentation function for pp (red) and p–Pb data (blue) at  $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$  as measured by the ALICE detector. The boxes represent the systematic uncertainties while the vertical bars indicate the statistical uncertainties. The dashed green line corresponds to PYTHIA 8.2. The  $\chi^2$  test for the comparison of pp and p–Pb data incorporates correlations among different  $z_T$  intervals. A constant that was fit to the ratio including statistical and systematic uncertainties is shown as grey band, with the width indicating the uncertainty on the fit.

$\gamma^{\text{iso}}$ -tagged fragmentation function shown in Fig. ???. This range roughly corresponds to the azimuthal angle consistent with the commonly used radius of  $R = 0.4$  for jet measurements.

The statistical uncertainty on the away-side yields in each  $z_T$  bin is calculated from the statistical uncertainty in the fully subtracted correlation functions, along with the statistical uncertainty arising from the uncorrelated background subtraction. A maximum charged hadron  $p_T$  of 10 GeV/c and a photon trigger  $p_T$  up to 40 GeV/c could result in a potential bias of the associated  $z_T$  spectrum. However, by repeating the analysis in different photon trigger  $p_T$  bins, it was found that any such effects were negligible compared to other uncertainties. The two largest sources of systematic uncertainty are from the purity and the single track correction factors. For the chosen  $p_T^{\text{track}}$  interval, there is no strong  $p_T$  dependence for the uncertainty of the charged tracking efficiency.

The ratio of the fragmentation functions in p–Pb and pp collisions is shown in the lower panel of Fig. ???. The fit yields a constant factor of  $0.84 \pm 0.11(\text{stat}) \pm 0.19(\text{sys})$ . Thus,

within total uncertainties, the p–Pb to pp ratio is consistent with unity.

# Chapter 6

## Discussion

### 6.0.1 Integrated Statistical Uncertainty on Fragmentation Function Ratio

For the purpose of giving a single number to quantify how similar pp and p–Pb fragmentation functions are, an integrated statistical uncertainty on the ratio of the two was calculated (note: the p-value calculated from the two distributions only indicates that the null hypothesis, pp and p–Pb are the same, cannot be rejected). First, the fragmentation function in pp was integrated, and the statistical errors were added in quadrature. The summed statistical uncertainty was then divided by this integral to obtain the relative uncertainty. The same was done for the p–Pb fragmentation function. Then, the two relative uncertainties were added in quadrature and the ratio of the integrals was taken. This is shown in equation ?? below,

$$\begin{aligned} I &= \sum_i y_i \cdot z_i \\ \delta_{\text{abs.}} &= \delta_0 \oplus \delta_1 \oplus \dots \delta_n \\ \delta_{\text{rel.}} &= \frac{\delta_{\text{abs.}}}{I}, \end{aligned} \tag{6.1}$$

where  $I$  is the integral of the fragmentation function,  $y_i$  is the conditional yield of associated hadrons in  $z_T$  bin  $i$ , and  $z_i$  is the width of  $z_T$  bin  $i$ . Additionally,  $n$  is the number of  $z_T$  bins, and  $\delta_i$  is the statistical uncertainty of the  $i$ th  $z_T$  bin.  $\delta_{\text{rel.}}$  is the relative statistical error on the fragmentation function. Taking the ratio of the integrals and summing the uncertainties from pp and p–Pb in quadrature:

$$\delta_{\text{ratio}} = \frac{I_{\text{p–Pb}}}{I_{\text{pp}}} \cdot (\delta_{\text{rel,pp}} \oplus \delta_{\text{rel,p–Pb}}), \tag{6.2}$$

yields a total integrated statistical uncertainty on the ratio,  $\delta_{\text{ratio}}$  of 13%. Thus, modifications to the fragmentation function in p–Pb are constrained to be less than 13%.

## 6.1 Kinematic Range Probed

In this work, azimuthal correlations of charged hadrons with isolated photons,  $\gamma^{\text{iso}}$ , are presented in p–Pb and pp collisions with a center-of-mass energy of  $\sqrt{s_{\text{NN}}} = 5.02$  TeV. Isolated photons are measured at midrapidity,  $|\eta| < 0.67$ , and with transverse momenta in the range  $12 < p_{\text{T}} < 40$  GeV/ $c$ , which yields the scaling variable  $x_{\text{T}} = 2p_{\text{T}}/\sqrt{s_{\text{NN}}} = 0.005$ – $0.016$ . This  $x_{\text{T}}$  range is similar to RHIC measurements at forward rapidity [Adare:2011sc].

This study marks the first study of photon-tagged fragmentation in p–Pb collisions at the LHC. A precise measurement of  $Q^2$  in heavy ion collisions is not possible without reconstructing all particles in the event. Nonetheless, among the LHC experiments, ALICE is uniquely configured to measure low- $p_{\text{T}}$  charged particles. In the context of jet constituents and total jet  $p_{\text{T}}$ , ALICE is capable of measuring hard scatterings with a lower  $Q^2$  than other LHC experiments. This is of particular interest for studying cold nuclear matter effects, as they are expected to be largest at lower  $Q^2$ .

The agreement between pp and p–Pb in this kinematic range constrains modifications to the parton fragmentation function to be less than the integrated uncertainty of 13% on the p–Pb/pp ratio for  $0.005x_{\text{T}} < 0.016$ , and partons with  $p_{\text{T}}$  that roughly corresponds to the photon  $p_{\text{T}}$  measured in this analysis.

### 6.1.1 Insensitivity to Parton Distribution Function

?? detailed the factorization of the hadronic cross section into the product of the parton distribution function (PDF) and the fragmentation function (FF). A very important question to answer for the study modifications to the fragmentation function in pPb collisions compared to pp collisions is: How do we know if the modifications to the observed  $\gamma$ -tagged associated yields are due to the PDF, and not in fact modifications to the fragmentation function?

To answer this question, the prompt photon and hadronic cross sections need to be understood, and the importance of measuring *per-trigger* yields. Because the main observable is the conditional yield of hadrons per photon, modifications to the cross production cross sections cancel. Coupled with the fact the photons should not otherwise be effected by nuclear effects, this measurement is most relevant towards constraining the impact of cold nuclear modification on the parton fragmentation function. Eqn. ?? shows the factorized cross sections of photons and hadrons in nuclear collisions:

[prompt photon cross section]

[hadronic cross section in PbPb] (reference the FF section in the introduction 1.4).

With XXX being the nPDF. Because this term is shared between the two cross sections, it's contribution is cancelled when this ratio is taken. As the conditional yield observable is directly proportional to the ratio of these two cross sections, the nPDF are not expected to contribute to the final measurement.

Thus, the answer to the question is: The observable per-trigger hadrons yields is by construction insensitive to the differences between the PDF and nPDF.

## 6.2 Comparison to Theory

## 6.3 Conclusion

## **Appendix A**

# **High Performance Computing for Event Mixing**

**A.1 Block Implementation and Parallelization**

**A.2 Hiearchical Data Format 5**