Principal component analysis of event-by-event fluctuations

Rajeev S. Bhalerao,[1] Jean-Yves Ollitrault,[2] Subrata Pal,[3] and Derek Teaney[4]

[1] *Department of Theoretical Physics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*

[2] *CNRS, URA2306, IPhT, Institut de physique theorique de Saclay, F-91191 Gif-sur-Yvette, France*

[3] *Department of Nuclear and Atomic Physics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*

[4] *Department of Physics and Astronomy, State University of New York, Stony Brook, NY 11794, USA*

(Dated: October 29, 2014)

We apply principal component analysis to the study of event-by-event fluctuations in relativistic heavy-ion collisions. This method brings out all the information contained in two-particle correlations in a physically transparent way. We present a guide to the method, and apply it to multiplicity fluctuations and anisotropic flow, using ALICE data and simulated events. In particular, we study elliptic and triangular flow fluctuations as a function of transverse momentum and rapidity.

I. INTRODUCTION

Anisotropic flow, v_n, is one of the most striking observations in heavy-ion collisions at ultrarelativistic energies [1, 2]. It is an azimuthal asymmetry of particle production, which is interpreted as a sign of the system's hydrodynamic response to the initial geometry. Event-by-event fluctuations of the initial geometry have long been recognized to play a crucial role in the interpretation of elliptic flow v_2 [3] and they are solely responsible for triangular flow v_3 [4]. However, the methods used to analyze anisotropic flow, namely the event-plane method [5] and cumulants [6], were devised before the importance of fluctuations was recognized. There are several discussions in the literature of how fluctuations may bias measurements obtained with these methods [7–12]. Here we argue that flow fluctuations can be obtained directly from data by fully exploiting the information contained in two-particle correlations [13].

The flow picture is that particles are emitted independently with an underlying probability distribution which varies event to event [14]. In each event we write the single-particle distribution with $d\mathbf{p} = dp_t\, d\eta\, d\varphi$ as

$$\frac{dN}{d\mathbf{p}} = \sum_{n=-\infty}^{+\infty} V_n(p)e^{in\varphi}, \qquad (1)$$

where φ is the azimuthal angle of the outgoing particle momentum, $V_n(p)$ is a complex Fourier flow coefficient whose magnitude and phase fluctuate event to event, and p is a shorthand notation for the remaining momentum coordinates, p_t and η. $V_0(p)$ is real and corresponds to the momentum distribution, and $V_n^* = V_{-n}$. Note that the usual definition of anisotropic flow v_n is real and normalized: $v_n = |V_n|/V_0$.

The statistics of the flow harmonics, $V_n(p)$, are measured from the distribution of particle pairs. Specifically, in the flow picture the pair distribution is determined (predominantly) by the statistics of the event-by-event single particle distribution

$$\left\langle \frac{dN_{\text{pairs}}}{d\mathbf{p_1}d\mathbf{p_2}} \right\rangle = \left\langle \frac{dN}{d\mathbf{p_1}}\frac{dN}{d\mathbf{p_2}} \right\rangle + \mathcal{O}(N), \qquad (2)$$

where angular brackets denote an average over events, and the term $\mathcal{O}(N)$ corresponds to correlations not due to flow ("nonflow"), which are small for large systems.

If the pair distribution is also expanded in a Fourier series

$$\left\langle \frac{dN_{\text{pairs}}}{d\mathbf{p_1}d\mathbf{p_2}} \right\rangle = \sum_{n=-\infty}^{+\infty} V_{n\Delta}(p_1, p_2)e^{in(\varphi_1-\varphi_2)}, \qquad (3)$$

then the measured series coefficients $V_{n\Delta}$ determine the statistics of V_n:

$$V_{n\Delta}(p_1, p_2) = \langle V_n(p_1)V_n^*(p_2)\rangle, \qquad (4)$$

where we have neglected nonflow correlations. The right-hand side of Eq. (4) is a covariance matrix, hence it is positive semidefinite. Thus a nontrivial property of flow correlations is that the measured pair correlation matrix $V_{n\Delta}(p_1, p_2)$ has only non-negative eigenvalues.[1]

The current Letter uses the eigenmodes and eigenvalues of the two-particle correlation matrix, $V_{n\Delta}(p_1, p_2)$, to fully classify flow fluctuations in heavy-ion collisions. Specifically, we show how a Principal Component Analysis (PCA) [16] of $V_{n\Delta}(p_1, p_2)$ can be used to fully extract information on the pseudorapidity and transverse-momentum-dependence of flow fluctuations. We first test the applicability of the method with Monte-Carlo simulations using the transport model AMPT [17] in both rapidity and transverse-momentum. The correlation analysis reveals at least one important subleading mode in both rapidity and p_t for the momentum distribution and

---

[1] Back-to-back jets, on the other hand, typically result in large negative diagonal elements for odd n [15], hence negative eigenvalues.

its second and third harmonics. Then we analyze AL-ICE data [13] in transverse-momentum and determine the first subleading elliptic and triangular flow coefficients, $v_2^{(2)}(p_t)$ and $v_3^{(2)}(p_t)$.

II. METHOD

Divide the detector acceptance into N_b bins in transverse momentum and/or pseudorapidity, $p = (p_t, \eta)$. The sample estimate for $V_n(p)$ in a given event (usually referred to as the flow vector [5]) is

$$Q_n(p) \equiv \sum_{j=1}^{M(p)} \exp(in\varphi_j), \qquad (5)$$

where $M(p)$ is the number of particles in the bin and φ_j is the azimuthal angle of a particle. The pair distribution is

$$V_{n\Delta}(p_1, p_2) \equiv \langle Q_n(p_1) Q_n^*(p_2) \rangle - \langle M(p_1) \rangle \, \delta_{p_1, p_2} \\ - \langle Q_n(p_1) \rangle \langle Q_n^*(p_2) \rangle, \qquad (6)$$

where the second term of the right-hand side subtracts self-correlations [18]. If self-correlations are not subtracted, $V_{n\Delta}(p_1, p_2)$ is positive semidefinite by construction. After subtraction, eigenvalues may have both signs. However, the eigenvalues will be positive if the correlations are due to collective flow.

The last term on the right-hand side of Eq. (6) subtracts the mean value in order to single out the fluctuations. For $n > 0$ and an azimuthally symmetric detector, this term vanishes by symmetry. For an asymmetric detector, subtracting the mean value corrects for azimuthal anisotropies in the acceptance [19]. Note that we define $V_{n\Delta}(p_1, p_2)$ as a *sum* over all pairs, as opposed to the usual normalization [4, 13] where one *averages* over pairs in each bin.[2]

The principal component analysis approximates the pair distribution as:

$$V_{n\Delta}(p_1, p_2) \approx \sum_{\alpha=1}^{k} V_n^{(\alpha)}(p_1) V_n^{(\alpha)*}(p_2), \qquad (7)$$

where each term in the sum corresponds to a different component (mode) of flow fluctuations, and $k \leq N_b$. If there are no flow fluctuations, the pair distribution $V_{n\Delta}(p_1, p_2)$ factorizes [13] and there is only one component, *i.e.* $k = 1$ in Eq. (7). Flow fluctuations break factorization [23]. Higher-order principal components then reveal information about the statistics and momentum dependence of flow fluctuations.

In practice, the principal components are obtained by diagonalizing $V_{n\Delta}(p_1, p_2)$ and ordering eigenvalues $\lambda^{(\alpha)}$ from largest to smallest, $\lambda^{(1)} > \lambda^{(2)} > \lambda^{(3)} \cdots$. If $\psi^{(\alpha)}(p)$ denotes the normalized eigenvector corresponding to the eigenvalue $\lambda^{(\alpha)}$, the corresponding principal component is given by

$$V_n^{(\alpha)}(p) \equiv \sqrt{\lambda^{(\alpha)}} \psi^{(\alpha)}(p). \qquad (8)$$

Note that the principal component analysis assumes $\lambda^{(\alpha)} > 0$. Up to statistical fluctuations, $V_n^{(\alpha)}(p)$ can be chosen to be real if parity is conserved.

The flow in a given event can be written as

$$V_n(p) = \sum_{\alpha=1}^{k} \xi^{(\alpha)} V_n^{(\alpha)}(p), \qquad (9)$$

where $\xi^{(\alpha)}$ are complex, uncorrelated random variables with zero mean and unit variance, that is, $\langle \xi^{(\alpha)} \rangle = 0$ and $\langle \xi^{(\alpha)} \xi^{(\beta)*} \rangle = \delta^{\alpha, \beta}$. The rms magnitude and momentum dependence of flow fluctuations are determined by the corresponding properties of the principal components. The orientation angle of a flow fluctuation in a specific event is determined by the phase of $\xi^{(\alpha)}$.

For sake of compatibility with the usual definition of $v_n(p)$ which is the anisotropy per particle, we define

$$v_n^{(\alpha)}(p) \equiv \frac{V_n^{(\alpha)}(p)}{\langle M(p) \rangle}. \qquad (10)$$

Thus, $v_0^{(\alpha)}(p)$ describe relative multiplicity fluctuations, while $v_n^{(\alpha)}(p)$ describe fluctuations of anisotropic flow.

III. RESULTS

In order to illustrate the method, we analyze 10^4 Pb-Pb collisions at $\sqrt{s} = 2.76$ TeV in the 0-10% centrality range,[3] generated using the AMPT model [17]. Initial conditions are generated via the HIJING 2.0 model [24] which contains event-by-event fluctuations. Collective flow is effectively generated in AMPT by elastic scatterings. We have checked that the present implementation reproduces LHC data for anisotropic flow (v_2 to v_6) at all centralities [25].

We first construct the pair distribution, Eq. (6), for all particles in the $-3 < \eta < 3$ pseudorapidity window, in η bins of 0.5. We then diagonalize the 12×12 matrix corresponding to these pseudorapidity bins. The eigenvalues are in general strongly ordered from largest to smallest.

---

[2] The present normalization is required by the principal component analysis. Note that a similar normalization is now used in cumulant analyses of anisotropic flow, where it is better to give all pairs (or generally multiplets) the same weight [20–22].

---

[3] We only show one centrality bin for sake of illustration, but we have checked that results are similar for other centralities.
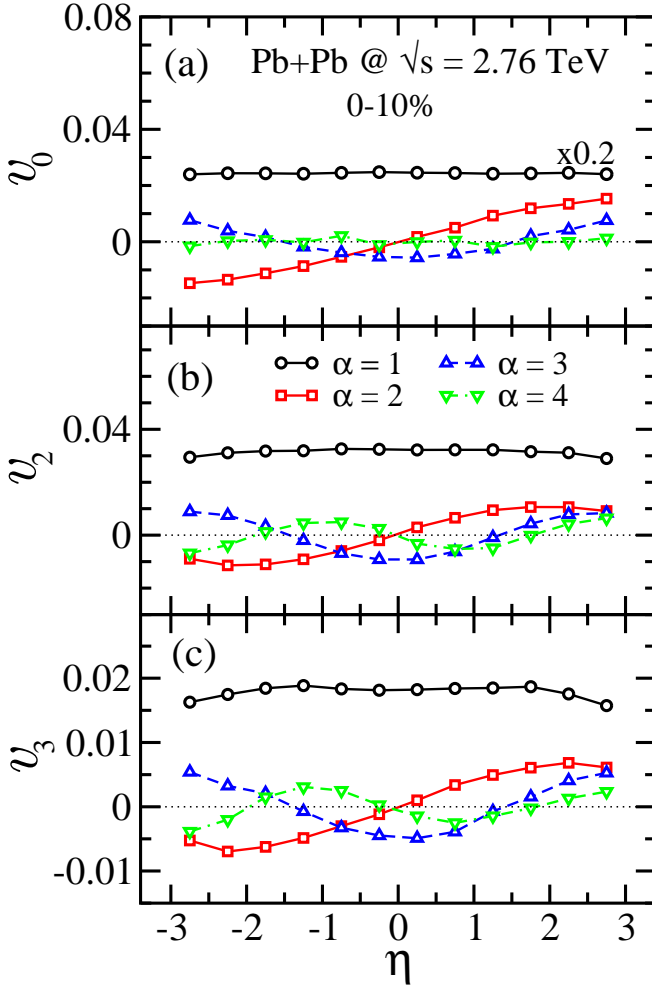
FIG. 1. Principal component analysis as a function of pseudorapidity for Pb+Pb collisions at $\sqrt{s} = 2.76$ TeV in the 0-10% centrality window generated with AMPT. (a) Multiplicity fluctuations; (b) Elliptic flow fluctuations; (c) Triangular flow fluctuations.

There are a few negative eigenvalues which can be attributed to statistical fluctuations.[4]

The leading principal components for $n = 0$, $n = 2$, and $n = 3$ are shown in Fig. 1. Fig. 1 (a) displays the principal modes of multiplicity fluctuations ($n = 0$) as a function of pseudorapidity. The leading mode $v_0^{(1)}(\eta)$ is a global 12% relative fluctuation, independent of η, corresponding to the fluctuation of the total multiplicity within the event sample. The next-to-leading mode $v_0^{(2)}(\eta)$ is odd and of much smaller amplitude, as shown by the eigenvalues, $\lambda^{(2)} \sim \lambda^{(1)}/60$. A natural source of

this odd mode is the small difference between the participant numbers of projectile and target nuclei induced by fluctuations, which creates a forward-backward asymmetry of the multiplicity [26, 27]. Since both the colliding system and the analysis window are symmetric around $\eta = 0$, principal components have definite parity, up to statistical fluctuations. Indeed, the next mode $v_0^{(3)}(\eta)$ is even, suggesting that principal components typically have alternating parities. The corresponding eigenvalue is again much smaller, $\lambda^{(3)} \sim \lambda^{(2)}/5$. $v_0^{(4)}(\eta)$ and higher modes are blurred by statistical fluctuations. Note that Eq. (7) defines principal components up to a sign. Here, we conventionally choose $v_n^{(\alpha)}(\eta) > 0$ at forward rapidity. Fig. 1 also illustrates the orthogonality of principal components, that is,

$$\sum_p V_n^{(\alpha)}(p)V_n^{(\beta)*}(p) = 0 \quad \text{if} \quad \alpha \neq \beta. \qquad (11)$$

Thus, $v_n^{(\alpha)}(p)$ typically has $\alpha - 1$ nodes.

Fig. 1 (b) and (c) display the principal components of elliptic and triangular flow fluctuations as a function of pseudorapidity. The leading modes $v_n^{(1)}(\eta)$ correspond to the usual elliptic and triangular flows, which depend weakly on η at the LHC [28, 29]. The subleading modes $v_n^{(2)}(\eta)$ are odd and of smaller amplitude ($\lambda^{(2)} \simeq \lambda^{(1)}/13$). These rapidity-odd harmonic flows, or torqued flows, can be attributed to the small relative angle between n-th harmonic participant planes defined in the projectile and target nuclei [30].

Note that the correlation matrix $V_{n\Delta}(p_1, p_2)$ is the sum of flow and nonflow correlations [31]. The nonflow correlation is significant only for small values of the relative pseudorapidity $\Delta\eta \equiv |\eta_1 - \eta_2|$. If the range in $\Delta\eta$ is smaller than the binning, it contributes to the diagonal elements, and its effect is to shift all eigenvalues by a constant. We observe in general a clear ordering of eigenvalues ($\lambda^{(3)}/\lambda^{(2)} \sim \lambda^{(4)}/\lambda^{(3)} \sim 2$) which indicates that the correlation has a long range in $\Delta\eta$ and is therefore dominated by flow. Visual inspection of correlation matrix $V_{n\Delta}(\eta_1, \eta_2)$ qualitatively confirms this reasoning.

We then carry out the analysis as a function of transverse momentum. In addition to AMPT generated events, we use experimental data for $V_{n\Delta}(p_1, p_2)$ provided by the ALICE collaboration [13] for Pb+Pb collisions in the 0-10% centrality window. ALICE uses all charged particles in the pseudorapidity window $|\eta| < 1$. The definition of $V_{n\Delta}$ is not quite the same as ours: First, it is averaged (as opposed to summed) over pairs. We correct for this difference by multiplying by the average multiplicity of pairs in each (p_1, p_2) bin, which we estimate using the statistical errors provided by ALICE ($\sigma \simeq (2N_{\text{pairs}})^{-1/2}$). Second, the analysis is done with a rapidity gap to suppress nonflow correlations: this means that particles 1 and 2 in Eq. (6) are separated by a rapidity gap of 0.8. For sake of comparison, we repeat the analysis using AMPT events (the same as in Fig. 1). In order to compensate for the lower statistics, we use a

---

[4] This can be checked by applying the PCA to purely statistical fluctuations. We generated random matrices according to the statistical error of $V_{n\Delta}(p_1, p_2)$, and found that the negative eigenvalues of $V_{n\Delta}(p_1, p_2)$ are compatible with the negative eigenvalues of these random matrices.

wider pseudorapidity window, from -2 to 2, with a 0.8 rapidity gap between -0.4 and 0.4.
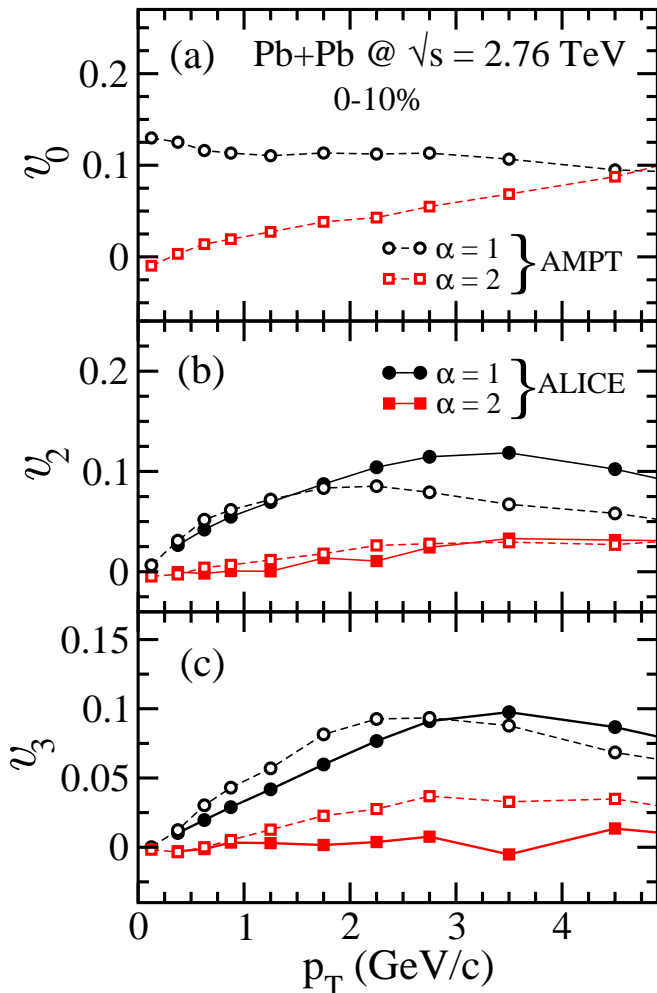

FIG. 2. Principal component analysis as a function of transverse momentum for Pb+Pb collisions at $\sqrt{s} = 2.76$ TeV in the 0-10% centrality window, using ALICE data [13] (full symbols) and AMPT events (open symbols). (a) Multiplicity fluctuations; (b) Elliptic flow fluctuations; (c) Triangular flow fluctuations.

Fig. 2 (a) displays the two principal modes of multiplicity fluctuations ($n = 0$) as a function of transverse momentum. Since no experimental data are available for this analysis, we only use AMPT events. As in Fig. 1 (a), the leading mode is essentially constant and corresponds to the 12% fluctuation in the total multiplicity. The subleading mode increases linearly as a function of p_t for large p_t, which can be interpreted as the result of a radial flow fluctuation. Specifically, in a hydrodynamic model the number of particles at high p_t decreases as, $\exp{(p_t(u-u_0)/T)}$, where u is the maximum fluid 4-velocity, $u_0 = \sqrt{1+u^2}$ and T is the temperature [32]. A small variation in u therefore produces a relative variation in the yield increasing linearly with p_t.

Fig. 2 (b) and (c) display the two leading principal

components for elliptic and triangular flows. ALICE data show a very strong ordering between the two leading eigenvalues ($\lambda^{(1)}/\lambda^{(2)} \sim 400$ for $n = 2$, 300 for $n = 3$). The leading components for $n = 2$ and $n = 3$ are very close to the value of v_2 and v_3 obtained using the same data [13]. The subleading mode is of much smaller magnitude and becomes significant only at large transverse momentum. In a hydrodynamical model, such a behavior is expected. Indeed, because of fluctuations the direction of high-momentum particles typically has a slight angle relative to the direction of low-momentum particles [15]. Thus, the flow of high-p_t particles has a small component independent of the flow of low-p_t particles. The subleading mode determines the magnitude of this additional component. As we shall see below, it is also responsible for the factorization breaking of azimuthal correlations observed in these data [13, 23].

IV. DISCUSSION

This new method, unlike traditional analysis methods, makes use of *all* the information contained in two-particle azimuthal correlations. Specifically, it uses the detailed information on how they depend on the momenta of *both* particles, as opposed to traditional analyses which are integrated over one of the momenta.

Previously, this double-differential structure has been used to test the factorization of azimuthal correlations. Small factorization breaking effects have been seen experimentally [13, 33] and in event-by-event hydrodynamic calculations [23, 34, 35]. They have been characterized by the Pearson correlation coefficient between two different momenta:

$$r = \frac{V_{n\Delta}(p_1, p_2)}{\sqrt{V_{n\Delta}(p_1, p_1) V_{n\Delta}(p_2, p_2)}}. \quad (12)$$

$r = \pm 1$ if correlations factorize, and $|r| \leq 1$ in general if correlations are due to flow.

These results are easily recovered in the language of principal components, in a physically transparent way. Factorization is the limiting case of just one principal component ($k = 1$ in Eq. (7)). In the more general case $k > 1$, Eq. (7) guarantees $|r| \leq 1$: Cauchy-Schwarz inequalities [23] are equivalent to the condition that all eigenvalues of the matrix $V_{n\Delta}$ are positive. Flow fluctuations are typically dominated by a single subleading mode, i.e., $k = 2$, with $|V_n^{(2)}(p)| \ll |V_n^{(1)}(p)|$. In this limit, Eq. (12) gives

$$1 - r \simeq \frac{1}{2} \left| \frac{V_n^{(2)}(p_1)}{V_n^{(1)}(p_1)} - \frac{V_n^{(2)}(p_2)}{V_n^{(1)}(p_2)} \right|^2. \quad (13)$$

We see that $r \leq 1$ as required by the Cauchy-Schwarz inequality. Further, we see that the breaking of factorization is induced by the relative difference between the subleading mode and the leading mode.

Thus principal components express the detailed information contained in correlations in a convenient way, which can be directly compared with model calculations. We anticipate a rich experimental and theoretical program studying the dynamics of these subleading flow vectors, which can be used to further constrain the plasma response to the geometry.

[1] U. Heinz and R. Snellings, Ann. Rev. Nucl. Part. Sci. **63**, 123 (2013) [arXiv:1301.2826 [nucl-th]].

[2] C. Gale, S. Jeon and B. Schenke, Int. J. Mod. Phys. A **28**, 1340011 (2013) [arXiv:1301.5893 [nucl-th]].

[3] B. Alver *et al.* [PHOBOS Collaboration], Phys. Rev. Lett. **98**, 242302 (2007) [nucl-ex/0610037].

[4] B. Alver and G. Roland, Phys. Rev. C **81**, 054905 (2010) [Erratum-ibid. C **82**, 039903 (2010)] [arXiv:1003.0194 [nucl-th]].

[5] A. M. Poskanzer and S. A. Voloshin, Phys. Rev. C **58**, 1671 (1998) [nucl-ex/9805001].

[6] N. Borghini, P. M. Dinh and J. Y. Ollitrault, Phys. Rev. C **64**, 054901 (2001) [nucl-th/0105040].

[7] C. Adler *et al.* [STAR Collaboration], Phys. Rev. C **66**, 034904 (2002) [nucl-ex/0206001].

[8] J. Adams *et al.* [STAR Collaboration], Phys. Rev. C **72**, 014904 (2005) [nucl-ex/0409033].

[9] B. Alver, B. B. Back, M. D. Baker, M. Ballintijn, D. S. Barton, R. R. Betts, R. Bindel and W. Busza *et al.*, Phys. Rev. C **77**, 014906 (2008) [arXiv:0711.3724 [nucl-ex]].

[10] J. Y. Ollitrault, A. M. Poskanzer and S. A. Voloshin, Phys. Rev. C **80**, 014904 (2009) [arXiv:0904.2315 [nucl-ex]].

[11] C. Gombeaud and J. Y. Ollitrault, Phys. Rev. C **81**, 014901 (2010) [arXiv:0907.4664 [nucl-th]].

[12] M. Luzum and J. Y. Ollitrault, Phys. Rev. C **87**, no. 4, 044907 (2013) [arXiv:1209.2323 [nucl-ex]].

[13] K. Aamodt *et al.* [ALICE Collaboration], Phys. Lett. B **708**, 249 (2012) [arXiv:1109.2501 [nucl-ex]].

[14] M. Luzum, J. Phys. G **38**, 124026 (2011) [arXiv:1107.0592 [nucl-th]].

[15] J. Y. Ollitrault and F. G. Gardim, Nucl. Phys. A **904-905**, 75c (2013) [arXiv:1210.8345 [nucl-th]].

[16] I. Jolliffe, Principal component analysis, *in* Encyclopedia of statistics in behavioral science, Wiley Online Library (2005).

[17] Z. W. Lin, C. M. Ko, B. A. Li, B. Zhang and S. Pal, Phys. Rev. C **72**, 064901 (2005) [nucl-th/0411110].

[18] P. Danielewicz and G. Odyniec, Phys. Lett. B **157**, 146 (1985).

[19] N. Borghini, P. M. Dinh and J. Y. Ollitrault, Phys. Rev. C **63**, 054906 (2001) [nucl-th/0007063].

[20] R. S. Bhalerao, N. Borghini and J. Y. Ollitrault, Nucl. Phys. A **727**, 373 (2003) [nucl-th/0310016].

[21] A. Bilandzic, R. Snellings and S. Voloshin, Phys. Rev. C **83**, 044913 (2011) [arXiv:1010.0233 [nucl-ex]].

[22] A. Bilandzic, C. H. Christensen, K. Gulbrandsen, A. Hansen and Y. Zhou, Phys. Rev. C **89**, 064904 (2014) [arXiv:1312.3572 [nucl-ex]].

[23] F. G. Gardim, F. Grassi, M. Luzum and J. Y. Ollitrault, Phys. Rev. C **87**, no. 3, 031901 (2013) [arXiv:1211.0989 [nucl-th]].

[24] W. T. Deng, X. N. Wang and R. Xu, Phys. Rev. C **83**, 014915 (2011) [arXiv:1008.1841 [hep-ph]].

[25] S. Pal and M. Bleicher, Phys. Lett. B **709**, 82 (2012) [arXiv:1201.2546 [nucl-th]].

[26] A. Bzdak and D. Teaney, Phys. Rev. C **87**, no. 2, 024906 (2013) [arXiv:1210.1965 [nucl-th]].

[27] V. Vovchenko, D. Anchishkin and L. P. Csernai, Phys. Rev. C **88**, no. 1, 014901 (2013) [arXiv:1306.5208 [nucl-th]].

[28] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **707**, 330 (2012) [arXiv:1108.6018 [hep-ex]].

[29] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Rev. C **89**, 044906 (2014) [arXiv:1310.8651 [nucl-ex]].

[30] P. Bozek, W. Broniowski and J. Moreira, Phys. Rev. C **83**, 034911 (2011) [arXiv:1011.3354 [nucl-th]].

[31] B. Alver *et al.* [PHOBOS Collaboration], Phys. Rev. C **81**, 034915 (2010) [arXiv:1002.0534 [nucl-ex]].

[32] N. Borghini and J. Y. Ollitrault, Phys. Lett. B **642**, 227 (2006) [nucl-th/0506045].

[33] CMS Collaboration [CMS Collaboration], CMS-PAS-HIN-14-012.

[34] U. Heinz, Z. Qiu and C. Shen, Phys. Rev. C **87**, no. 3, 034913 (2013) [arXiv:1302.3535 [nucl-th]].

[35] I. Kozlov, M. Luzum, G. Denicol, S. Jeon and C. Gale, arXiv:1405.3976 [nucl-th].