Reconstructing particle number distributions with convoluting volume fluctuations

ShinIchi Esumi†, ‡ and Toshihiro Nonaka†, ‡

† Tomonaga Center for the History of the Universe, University of Tsukuba, Tsukuba, Ibaraki 305, Japan
‡ Key Laboratory of Quark & Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079, China

We propose the methods to reconstruct the net-particle distributions with and without excluding the initial volume fluctuations. It enables us to correct for detector effects and initial volume fluctuations simultaneously under the assumption of the independent particle production (IPP) model or any particle production mechanism. Our study suggests a tool to investigate possible two-component structure of net-proton distribution in Au+Au collisions at \( \sqrt{s_{NN}} = 7.7 \) GeV as signature of the first-order phase transition or critical end point \([1,2]\).

Currently, no complete solution has been established for these issues, but in particular, the detector efficiency has to be corrected prior to other issues from the experiment side. The efficiency correction methods have been developed by Refs. \([13,18]\). All of them are derived under the assumption that efficiencies follow binomial distributions given by

\[
B(n; \varepsilon, N) = e^n (1 - p)^{N - n} \frac{N!}{n!(N - n)!},
\]

where \( N \) represents the number of generated particles, \( n \) represents the number of measured particles and \( \varepsilon \) represents the efficiency. If the experiment measurement strictly follows the binomial distribution, those correction methods are justified. Otherwise, they lose their validity as is pointed out by Ref. \([19]\). One possible method is moment expansion proposed in Ref. \([20]\), where detector-response matrices are utilized to correct cumulants. In addition, experimental attempts are ongoing to get cumulants corrected for possible non-binomial effects by reconstructing distribution itself \([1,21]\).

Initial volume fluctuations are another major backgrounds coming into the distributions and cumulants. Various studies are done to understand and correct for the initial volume fluctuations \([22,29]\). Correction methods proposed in the literature are derived based on the independent particle production model for each source like a pair of participant nucleons. Further, it is pointed out that the enhancement of \( C_4/C_2 \) net-proton multiplicity distributions has been also measured in Au+Au collisions at \( \sqrt{s_{NN}} = 54.4 \) and 200 GeV \([9]\) to study the nature of crossover predicted by the lattice QCD calculations \([9]\). In addition, higher-order cumulants of net-particle distributions have been reported by ALICE \([10]\), HADES \([11]\) and NA61 experiments \([12]\).

There are two main advantages to measure the higher order cumulants of conserved quantities. First, higher-order the cumulant is, more sensitive it is to the correlation length:

\[
C_2 \approx \xi^2, \quad C_3 \approx \xi^4.5, \quad C_4 \approx \xi^7, \quad (1)
\]

\[
C_5 \approx \xi^9.5, \quad C_6 \approx \xi^{12}, \quad (2)
\]

where the correlation length \( \xi \) is predicted to diverge near the critical point. Second, the cumulants can be directly (in principle) compared to the susceptibilities by taking ratio between different order:

\[
S_\sigma = \frac{C_3}{C_2} = \frac{\chi_3}{\chi_2}, \quad \kappa \sigma^2 = \frac{C_4}{C_2} = \frac{\chi_4}{\chi_2}, \quad (3)
\]

where \( \chi_n \) represents the \( n \)-th order susceptibility. However, the direct comparison between experimental and theoretical results are still difficult due to following reasons:

- Baryon number conservation effects
- Initial volume fluctuation
- Detector efficiency

I. INTRODUCTION

In recent years, the higher-order cumulants of net-particle distributions have been actively measured in heavy-ion collision experiments to study the QCD phase structure. Up to the 4th-order cumulants of net-proton, net-charge and net-kaon multiplicity distributions were measured in Au+Au collisions at various beam energies from \( \sqrt{s_{NN}} = 200 \) GeV down to 7.7 GeV at the STAR experiment \([3,7]\). Non-monotonic beam energy dependence of net-proton \( C_4/C_2 \) could indicate a possible critical point in low collision energies. The 6th-order cumulant of net-proton multiplicity distributions has been also measured in Au+Au collisions at \( \sqrt{s_{NN}} = 54.4 \) and 200 GeV \([9]\) to study the nature of crossover predicted by the lattice QCD calculations \([9]\). In addition, higher-order cumulants of net-particle distributions have been reported by ALICE \([10]\), HADES \([11]\) and NA61 experiments \([12]\).
II. PARTICLE NUMBER UNFOLDING

Our goal is to correct for cumulants of the net-particle distribution. We thus consider the number of measured particles and antiparticles (that we can measure in experiments) and the number of generated particles and antiparticles (that cannot be directly accessed in experiments). In this paper, those are denoted by \( N_p, N_{\bar{p}}, n_p \) and \( n_{\bar{p}} \), which are related to each other as following:

\[
P(N_p, N_{\bar{p}}) = \sum_{n_p, n_{\bar{p}}} \mathcal{R}_{\text{rev}}(N_p, N_{\bar{p}}; n_p, n_{\bar{p}}) \tilde{P}(n_p, n_{\bar{p}}),
\]

where \( \tilde{P}(n_p, n_{\bar{p}}) \) and \( P(N_p, N_{\bar{p}}) \) are two-dimensional probability distribution function for particles and antiparticles, and \( \mathcal{R}_{\text{fo}}(n_p, n_{\bar{p}}; N_p, N_{\bar{p}}) \) is the conversion matrix from \( (N_p, N_{\bar{p}}) \) (generated coordinate) to \( (n_p, n_{\bar{p}}) \) (measured coordinate). The conversion matrix is typically called "response matrix". On the other hand, the conversion matrix from measured to generated coordinate can also be defined by \( \mathcal{R}_{\text{rev}}(N_p, N_{\bar{p}}; n_p, n_{\bar{p}}) \). We explicitly call them "forward response matrix" and "reversed response matrix", respectively. We also use the word "detector filter", which represents the Monte Carlo way to determine \( n_p \) and \( n_{\bar{p}} \) from given \( N_p \) and \( N_{\bar{p}} \).

In the rest of this section, let us demonstrate an unfolding approach by using the toy model. The model utilizes two sets of data. One corresponds to the distributions in the real experiments, which is the "answer" (experiment) in the model. The other one is virtual distribution (simulation) that can be prepared arbitrary. The procedures we mention below is to modify the simulation to be close to the experiment.

Figure 1 shows the flowchart of the procedures. First, we prepare the experimental true distribution in panel (a) with the critical shape. Poisson distributions are generated for \( P(N_p) \) \( (N_p < 10) \), and the Poisson distributions are generated for \( P(N_{\bar{p}}) \) \( (N_{\bar{p}} > 10) \). The detector filter is then applied to panel (a) to get the experimental measured distribution in panel (b). Second, we define a virtual distribution according to the Poisson distributions in panel (c). Panel (d) shows the distribution after detector filter is applied to (c). During this process we compute \( \mathcal{R}_{\text{fo}} \) and \( \mathcal{R}_{\text{rev}} \). The correction function in panel (f) is determined by subtracting (d) from (b). Since panel (f) represents the difference between experiments and simulations in measured coordinate, we multiply \( \mathcal{R}_{\text{rev}} \) to (f) to get the difference between experiments and simulations in generated coordinate, as shown in panel (e). The virtual distribution is then modified to be close to (a) by adding (e) to (c) with a scaling factor \( \alpha < 1 \). The same procedures are repeated until cumulants of net-particle distributions are converged.

Figure 2 shows three kinds of distributions from top to bottom. The most right panels in the middle and bottom rows show the experimental true distributions, and the 1st to 5th panels from left-hand side represent the virtual distribution at 0th, 1st, 5th, 10th and 100th iterations. The top row in Fig. 2 shows the correction functions and virtual distributions in the generated coordinate for different iterations. It is found that the correction function is getting flattened with iterations, which indicates that the virtual distribution is modified to be close to the experimentally generated true distribution. The middle row shows the virtual distribution in the generated coordinate for different iterations. The critical shape of the experimental distribution is found to be reconstructed in the virtual distribution starting from the Poisson distribution. The bottom row shows the net-particle distribution for virtual one for different iterations. We find that the two-peak critical shape in experiments are successfully reproduced for the virtual distributions.

III. VOLUME FLUCTUATION CONVOLUTED UNFOLDING

A. Volume fluctuation

In heavy ion collisions, the volume of the colliding system is characterized by the impact parameter \( b \), which is defined by a distance between center of two nuclei. However, the impact parameter is not directly measured experimentally, we often consider participant nucleons \( (N_{\text{part}}) \), spectator nucleons and binary collisions. They are still not accessed directly in experiments, so determined by comparing experimentally measured multiplicity distribution (or energy loss of particles) with the model. Those variables are useful to characterize and understand the experimental quantities like collision centrality, nuclear modification factor and so on. For higher-order cumulants of net-particle distributions, the volume fluctuation is typically defined as participant fluctuations [30]. Assuming that particles are produced from the independent source of participant nucleons, the true cumulants are expressed by superposition of cumulants for each source [21, 25].

\[
C_r(\Delta N) = \sum_{N_{\text{source}}} \kappa_r(\Delta m)
\]

with

\[
\Delta N = N_p - N_{\bar{p}}, \quad \Delta m = m_p - m_{\bar{p}}.
\]
B. Methodology

We explain how to implement the volume fluctuations in the unfolding approach discussed in Sec. II. The flowchart is shown in Fig. 3. First, let us define the first set of data corresponding to the experiment. We consider two dimensional histogram for particle and antiparticle number produced for one source (a), which is generated according to the Poisson distribution for both particles and antiparticles. Statistically independent distributions for (a) are generated with $N_{\text{source}}$ times, and their superposition is shown in (b). During this procedure, on
the other hand, $N_{\text{source}}$ value can be determined event-
by-event according to the Gauss distribution to implement the
volume fluctuation. The resulting distribution is shown in (c). The
detector filter is applied to (c) to get the distribution in the measured coordinate (d). Second,
with the same procedures using quite different dis-
tributions, another set of data is generated for the initial
virtual distribution as shown in (e)–(h). The correction
function (i) is then determined by subtracting (h) from
(d) for the measured coordinate. In order to modify the
virtual distribution for each source (e), the following rela-
tion needs to be computed:

$$P(m_p, m_p) = \sum_{n_p, n_p} R_{\text{rev}}(m_p, m_p; n_p, n_p) \tilde{P}(n_p, n_p), \quad \text{(8)}$$

where $n_p(p)$ represents measured (anti)particles and $m_p(p)$
are particles for each source, with $R_{\text{rev}}(m_p, m_p; n_p, n_p)$
being the response matrix from (h) and (e). By multiply-
ing $R_{\text{rev}}(m_p, m_p; n_p, n_p)$ to (i), the correction function
is converted to that for each source (j). Adding (j) to (e)
with scaling factor $\alpha$ gives a modified distribution $(e')$
which should be slightly closer to (a). The superposition of $(e')$
with and without the volume fluctuation are shown in $(g')$
and $(f')$ which should be also closer to $(c)$ and $(b)$
than (g) and (f).

In this way, the virtual distribution for each source is
modified with iterations. Their superposition with and
without the volume fluctuation are modified accordingly.
The top row in Fig. 4 shows cumulants up to the fourth
order as a function of iteration. Black and red lines rep-
resent the cumulants with and without volume fluctua-
tions. Results from statistically independent 100 samples
are plotted as well. The observed gap between two results
are due to the volume fluctuations. Bottom panels show
the cumulants of superimposed sources. The bottom row
in Fig. 4 shows the correlation between input and output
cumulants. Output cumulants are taken from the 1000th
iteration. The consistency between x and y axis indicates
that our unfolding approach with and without the volume
fluctuation works well.

### IV. DISCUSSIONS

#### A. Systematic uncertainty

Possible source of systematic uncertainties are listed
below.

- Iterations
- Smoothing
- Scaling

As were demonstrated in the models, one needs to care-
fully check the convergence with respect to the iterations.
If the cumulants don't converge, statistics for the simu-
lated virtual data may not be enough. It is important to
use substantial statistics for the virtual data compared to
the experimental data.

Smoothing parameter is necessary to require that the
resulting distribution should be smooth. Figure 8 shows
the smoothing distributions, the correction functions, vir-
tual 2D distributions, and virtual net-particle distributions
after substantial iterations for different smoothing parameters. It is found that the distributions are smeared
with larger smoothing parameters. Figure 7 shows the
cumulants up to the 4th order as a function of iteration with different smoothing parameters. Although the
convergence behaviour seems different, the cumulants are
consistent for all the cases within statistical uncertainties
after the 1000 iterations. Figure 5 shows the cumulants
up to the 4th order as a function of iteration with different
scaling parameters. It is found that the scaling parameters
controls the convergence speed with the same value of
cumulants after substantial iterations. The scaling factor
is necessary to minimize the negative bin content (after
applying the correction functions) which could lead to
oscillating iterations.
FIG. 4. Particle and antiparticle number correlations in experiments for (a) each source (b) superposition for fixed $N_{\text{part}}$ (c) superposition for fluctuating $N_{\text{part}}$ (d) superposition for fluctuating $N_{\text{part}}$ with the detector filter. The same plots are shown for simulations in panel (e)–(h). The panel (i) shows the correction function. The panel (h) shows the correction function per source. The panels (c'), (f') and (g') are the distributions after the 1st correction (iteration). Blank bins in panels (i) and (j) represent the negative value.

FIG. 5. (Top) Lines show the cumulants as a function of iteration. Results from 100 independent trials are superimposed. Barcodes for each bin show true value of cumulants for each trial. (Bottom) Correlation between input and output cumulants. Results without excluding the volume fluctuation are shown in black, and those with excluding the volume fluctuations are shown in red.
\( \alpha \approx 1.5 \) at the 80th iteration of incremental unfolding. Different style of bands are results from different smoothing parameter \( \alpha \) at the 80th iteration of incremental unfolding. (4th column) Net-particle distributions in generated coordinate for simulation sample representing the negative value. (3rd column) Particle number distributions in generated coordinate for simulation sample at the 80th iteration of incremental unfolding. (2nd column) Correction functions in generated coordinate with different smoothing parameter. Blank bins represent the negative value. (1st column) Weight distribution for different smoothing parameter \( \alpha_{sm} \). The bin content at \((N_p, N_{\bar{p}}) = (0, 0)\) is normalized to unity. (2nd column) Correction functions in generated coordinate with different smoothing parameter. Blank bins represent the negative value. (3rd column) Particle number distributions in generated coordinate for simulation sample at the 80th iteration of incremental unfolding. (4th column) Net-particle distributions in generated coordinate for simulation sample at the 80th iteration of incremental unfolding. The most left panels show the distributions without smoothing \((\alpha_{sm} = 0)\).

FIG. 6. (1st column) Weight distribution for different smoothing parameter \( \alpha_{sm} \). The bin content at \((N_p, N_{\bar{p}}) = (0, 0)\) is normalized to unity. (2nd column) Correction functions in generated coordinate with different smoothing parameter. Blank bins represent the negative value. (3rd column) Particle number distributions in generated coordinate for simulation sample at the 80th iteration of incremental unfolding. (4th column) Net-particle distributions in generated coordinate for simulation sample at the 80th iteration of incremental unfolding. The most left panels show the distributions without smoothing \((\alpha_{sm} = 0)\).

FIG. 7. Cumulants up to the 4th-order as a function of iteration. Solid lines and bands show the averaged value and \( \pm 1 \sigma \) for 100 independent trials. Different style of bands are results from different smoothing parameter \( \alpha_{sm} \). The box drawn at \( x \approx 1500 \) is the true value with \( \pm 1 \sigma \) for the experimental samples in generated coordinate.
FIG. 8. Cumulants up to the 4th-order as a function of iteration. Solid lines and bands show the averaged value and ±1σ for 100 independent trials. Different style of bands are results from different scaling parameter α_{sc}. The box drawn at x ≈ 1500 is the true value with ±1σ for the experimental samples in generated coordinate.
B. Application to experimental data

The distribution measured in the real experiment corresponds to (d) in Fig. 4. One can start from Poisson distribution with mean parameter taken from efficiency corrected protons and antiprotons divided by averaged number of participant nucleons for the centrality. The source distribution can be superimposed according to \( N_{\text{source}} \) distribution obtained by Glauber model used in the centrality determination in the experiment. One can try using the binomial efficiency for the detector filter. If one wants to correct for possible non-binomial efficiencies, that needs to be understood by GEANT based full detector Monte Carlo simulations including event and track reconstructions. The detector response including the reconstruction would also need to be parametrized somehow, since lots of computing power will be necessary and repeating it many times would not be realistic.

C. Independent particle production model

It is pointed out by Ref. [29] that the independent particle production model would be broken in the framework of UrQMD, as well as in the real experiment where we expect the strongly interacting hot and dense matter to be formed. Methodology in this paper would allow us to use any other model numerically, even if the model cannot be formalized. For example, the correlated production among more than two participant nucleons can be easily implemented during the superposition procedure.

V. ACKNOWLEDGEMENT

We thank X. Luo, B. Mohanty and N. Xu for fruitful discussions. This work was supported by the National Key Research and Development Program of China (2018YFE0205201), the National Natural Science Foundation of China (No.11828501, 11575069, 11909711, 1186131009 and 11950410505), China Postdoctoral Science Foundation funded project 2018M642878, Ito Science Foundation (2017) and JSPS KAKENHI Grant No. 25105504 and 19H05598.

Appendix A: Response matrix

Figures 9 and 10 are forward and reversed four-dimensional response matrices, \( R_{\text{for}}(p_{np}, n_{np}, N_{p}, N_{\bar{p}}) \) and \( R_{\text{rev}}(p_{np}, n_{np}, N_{p}, N_{\bar{p}}) \), in Sec. II. Each two-dimensional histogram is normalized to one. The correction functions in the measured coordinate can be converted to those in the generated coordinate by multiplying the reversed response matrices.

References

[1] A. Bzdak, V. Koch, D. Oliinychenko, and J. Steinheimer, Phys. Rev. C98, 054901 (2018), arXiv:1804.04463 [nucl-th].
[2] A. Bzdak and V. Koch, (2018), arXiv:1811.04456 [nucl-th].
[3] M. Aggarwal et al. (STAR), Phys. Rev. Lett. 105, 022302 (2010) arXiv:1004.4959 [nucl-ex].
[4] L. Adamczyk et al. (STAR), Phys. Rev. Lett. 112, 032302 (2014) arXiv:1309.5681 [nucl-ex].
[5] L. Adamczyk et al. (STAR), Phys. Rev. Lett. 113, 092301 (2014) arXiv:1402.1558 [nucl-ex].
[6] L. Adamczyk et al. (STAR), (2017), arXiv:1709.00773 [nucl-ex].
[7] J. Adam et al. (STAR), (2020), arXiv:2001.02852 [nucl-ex].
[8] T. Nonaka (STAR) (2020) arXiv:2002.12505 [nucl-ex].
[9] Y. Aoki, G. Endrodi, Z. Fodor, S. D. Katz, and K. K. Szabo, Nature 433, 675 (2006) arXiv:hep-lat/0611014 [hep-lat].
[10] M. Arslanbek, in 28th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (Quark Matter 2019) Wuhan, China, November 4-9, 2019 (2020) arXiv:2002.03906 [nucl-ex].
[11] J. Adamczyk-Musch et al. (HADES), (2020), arXiv:2002.08701 [nucl-ex].
[12] M. Mackowiak-Pawłowska (NA61/SHINE), in 28th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (Quark Matter 2019) Wuhan, China, November 4-9, 2019 (2020) arXiv:2002.04847 [nucl-ex].
[13] M. Kitazawa and M. Asakawa, Phys. Rev. C86, 024904 (2012) [Erratum: Phys. Rev.C86,069902(2012)], arXiv:1205.3292 [nucl-th].
[14] A. Bzdak and V. Koch, Phys. Rev. C86, 044904 (2012) arXiv:1206.4286 [nucl-th].
[15] A. Bzdak and V. Koch, Phys. Rev. C91, 027901 (2015) arXiv:1312.4574 [nucl-th].
[16] X. Luo, Phys. Rev. C91, 034907 (2015) arXiv:1410.3914 [physics.data-an].
[17] T. Nonaka, M. Kitazawa, and S. Esumi, Phys. Rev. C95, 064912 (2017), arXiv:1702.07106 [physics.data-an].
[18] X. Luo and T. Nonaka, Phys. Rev. C99, 044917 (2019) arXiv:1812.10303 [physics.data-an].
[19] A. Bzdak, R. Holzmann, and V. Koch, Phys. Rev. C94, 064907 (2016) [arXiv:1603.09057 [nucl-th]].
[20] T. Nonaka, M. Kitazawa, and S. Esumi, Nucl. Instrum. Meth. A906, 10 (2015), arXiv:1805.00279 [physics.data-an].
[21] T. Nonaka (STAR), Proceedings, 27th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (Quark Matter 2018): Venice, Italy, May 14-19, 2018, Nucl. Phys. A982, 863 (2019).
[22] X. Luo, J. Xu, B. Mohanty, and N. Xu, J. Phys. G40, 105104 (2013) arXiv:1302.2332 [nucl-ex].
[23] M. I. Gorenstein and M. Gazdzicki, Phys. Rev. C84, 014904 (2011) arXiv:1101.4865 [nucl-th].
[24] V. Skokov, B. Friman, and K. Redlich, Phys. Rev. C88, 034911 (2013) arXiv:1205.4756 [hep-ph].
[25] P. Braun-Munzinger, A. Rüstamov, and J. Stachel, Nucl. Phys. A960, 114 (2017) arXiv:1612.00702 [nucl-th].
[26] S. Sombun, J. Steinheimer, C. Herold, A. Limphirat, Y. Yan, and M. Bleicher, J. Phys. G45, 025101 (2018) arXiv:1709.00879 [nucl-th].
[27] R. Rogly, G. Gacaloni, and J.-Y. Ollitrault, Phys. Rev. C98, 024902 (2018) arXiv:1804.03031 [nucl-th].
[28] W. Broniowski and A. Olszewski, Phys. Rev. C95, 064910 (2017) arXiv:1704.01532 [nucl-th].
[29] T. Sugihara, T. Nonaka, and S. Esumi, Phys. Rev. C100, 044904 (2019) arXiv:1903.02314 [nucl-th].
FIG. 9. Forward response matrices with respect to $n_p$ and $n_{\bar{p}}$ for fixed $(N_p, N_{\bar{p}})$ ($R_{\text{foo}}(n_p, n_{\bar{p}}; N_p, N_{\bar{p}})$), for the 1st iteration in Sec. III The area is normalized to one.

[30] X. Luo and N. Xu, Nucl. Sci. Tech. 28, 112 (2017) arXiv:1701.02105 [nucl-ex]
FIG. 10. Reversed response matrices with respect to $N_p$ and $N_{\bar{p}}$ for fixed $(n_p, n_{\bar{p}})$ ($R_{\text{rev}}(N_p, N_{\bar{p}}; n_p, n_{\bar{p}})$), for the 1st iteration in Sec. III. The area is normalized to one.