Unfolding the Moments of Multiplicity distributions

Anar Rustamov

a.rustamov@cern.ch

- ✓ Outline
  - ✓ introduction to unfolding
  - ✓ a novel approach
    - ✓ test on simulation
    - ✓ application to data
- ✓ Summary

FIAS Frankfurt Institute for Advanced Studies

Unfolding - Introduction

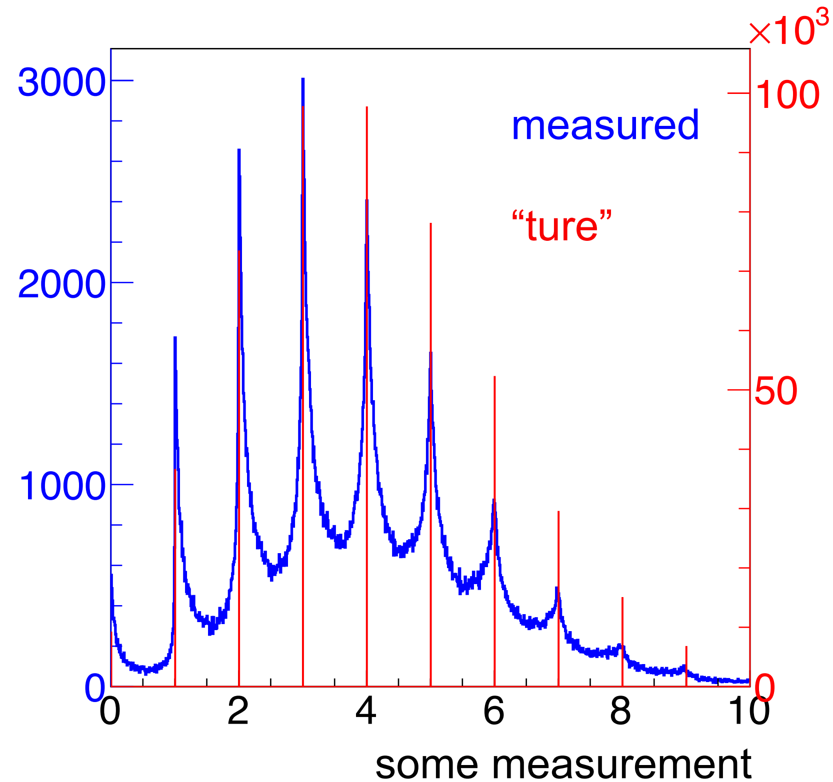
Also referred to as: deconvolutiion, unsmearing, etc.

Experimental measurements are affected by:

- ✓ Statistical fluctuations (vertical axis)
- ✓ Detector effects (migration between bins)

Unfolding:

- ✓ correct for migration effects
  in the presence of statistical fluctuations
- ✓ restore “true” distribution
- ✓ obtain covariance matrix



$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j, \quad i = 1,\ldots,N$$

measured in bin *i* — response matrix — true value in bin *j*

$$R_{ij} = P(\text{observed in } i \mid \text{ true value in } j)$$

Moments, Cumulants …

Given: Random variable Y with probability distribution p(y)

n^{th} moment about origin: $$\left\langle y^n \right\rangle = \left. \frac{d^n m(t)}{dt^n} \right|_{t=0} \qquad m(t) = \sum_y e^{ty} p(y)$$

Cumulants: $$C_n = \left. \frac{d^n g(t)}{dt^n} \right|_{t=0} \qquad g(t) = \log[m(t)]$$

$$C_1 = \langle y \rangle$$

$$C_2 = \left\langle \left(y - \langle y \rangle\right)^2 \right\rangle = \sigma^2 \qquad S\sigma = \frac{C_3}{C_2}$$

$$C_3 = \left\langle \left(y - \langle y \rangle\right)^3 \right\rangle = S\sigma^3 \qquad k\sigma^2 = \frac{C_4}{C_2}$$

$$C_4 = \left\langle \left(y - \langle y \rangle\right)^4 \right\rangle - 3\sigma^4 = k\sigma^4$$

all cumulants can be expressed in terms of moments

Selected fluctuation measures

Multiplicity fluctuations

$$\omega = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} = \frac{Var(N)}{\langle N \rangle}$$

Poisson case: $\langle N^2 \rangle = \langle N \rangle^2 + \langle N \rangle, \quad \omega = 1$

Chemical (particle composition) fluctuations

$$\nu_{dyn} = \frac{\langle N_1^2 \rangle - \langle N_1 \rangle}{\langle N_1 \rangle^2} + \frac{\langle N_2^2 \rangle - \langle N_2 \rangle}{\langle N_2 \rangle^2} - 2\frac{\langle N_1 N_2 \rangle}{\langle N_1 \rangle \langle N_2 \rangle}$$

$$\nu_{dyn} \approx \mathrm{sgn}(\sigma_{dyn})\sigma^2_{dyn}$$

Independent Poisson distributions: $\langle N_i^2 \rangle = \langle N_i \rangle^2 + \langle N_i \rangle, \quad \langle N_1 N_2 \rangle = \langle N_1 \rangle \langle N_2 \rangle \equiv \nu_{dyn} = 0$

$$\Sigma[A,B] = \frac{\langle B \rangle \omega(A) + \langle A \rangle \omega(B) - 2(\langle AB \rangle - \langle A \rangle \langle B \rangle)}{\langle B \rangle + \langle A \rangle}$$

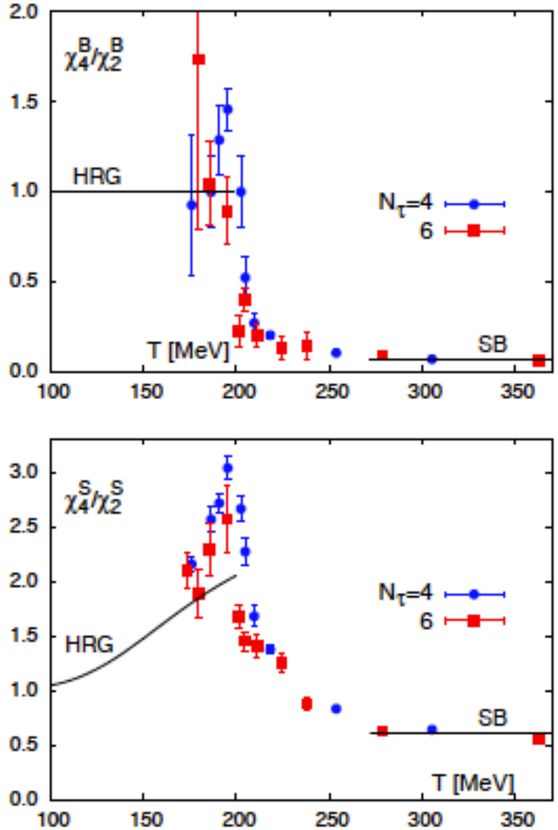
$$\Delta[A,B] = \frac{\langle B \rangle \omega(A) - \langle A \rangle \omega(B)}{\langle B \rangle - \langle A \rangle}$$

in GCE depend neither
on system volume no on its fluctuations

M. I. Gorenstein, M. Gazdzicki PRC 84, 014904 (2011)

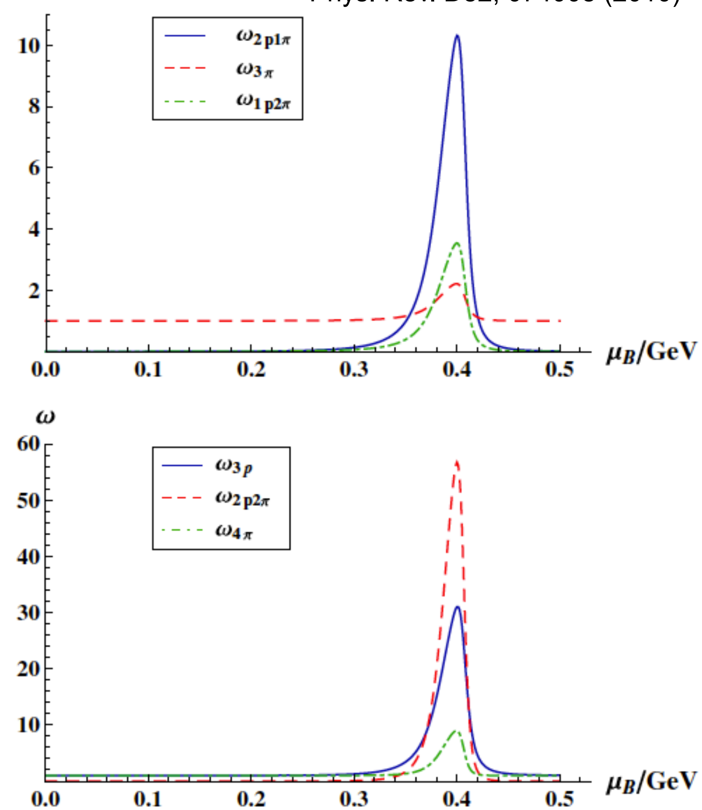
all fluctuation measures can be expressed in terms of moments (also mixed)

Theoretical predictions



Phys. Rev. D79, 074505 (2009)

Phys. Rev. D82, 074008 (2010)



$$C_{n,q} = VT^3\chi_n^q$$

$$\omega_{1x2y} = \frac{\left\langle \left(x-\langle x\rangle\right)\left(y-\langle y\rangle\right)^2 \right\rangle}{\langle x\rangle^{1/3}\langle y\rangle^{2/3}}$$

Also cf. talk of M. Gorenstein from yesterday

Particle Identification

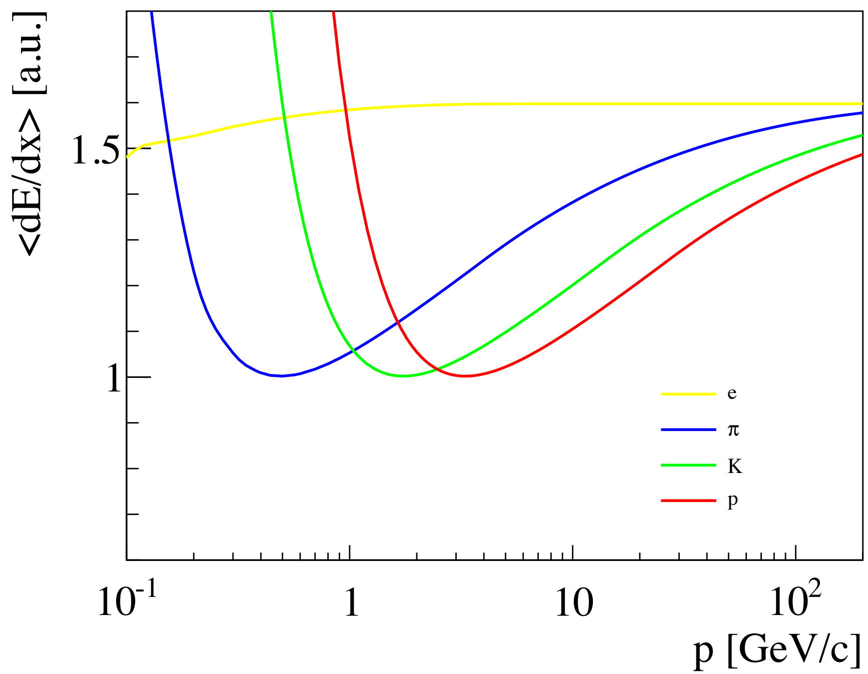
$$\vec{p} = m\vec{\beta}\gamma$$

in order to identify the particle
at least two independent measures are needed

$$-\left\langle \frac{dE}{dx} \right\rangle (\beta\gamma) \propto \frac{z^2}{\beta^2} \ln(a\beta\gamma)$$

Momentum is obtained by solving the equation of motion in a magnetic field.

- ✓ Input:
  - ✓ set of measured points in space
  - ✓ magnetic field map (typically calculated)



simultaneous measurement of momentum and dE/dx allows to identify the identity (mass) of a particle

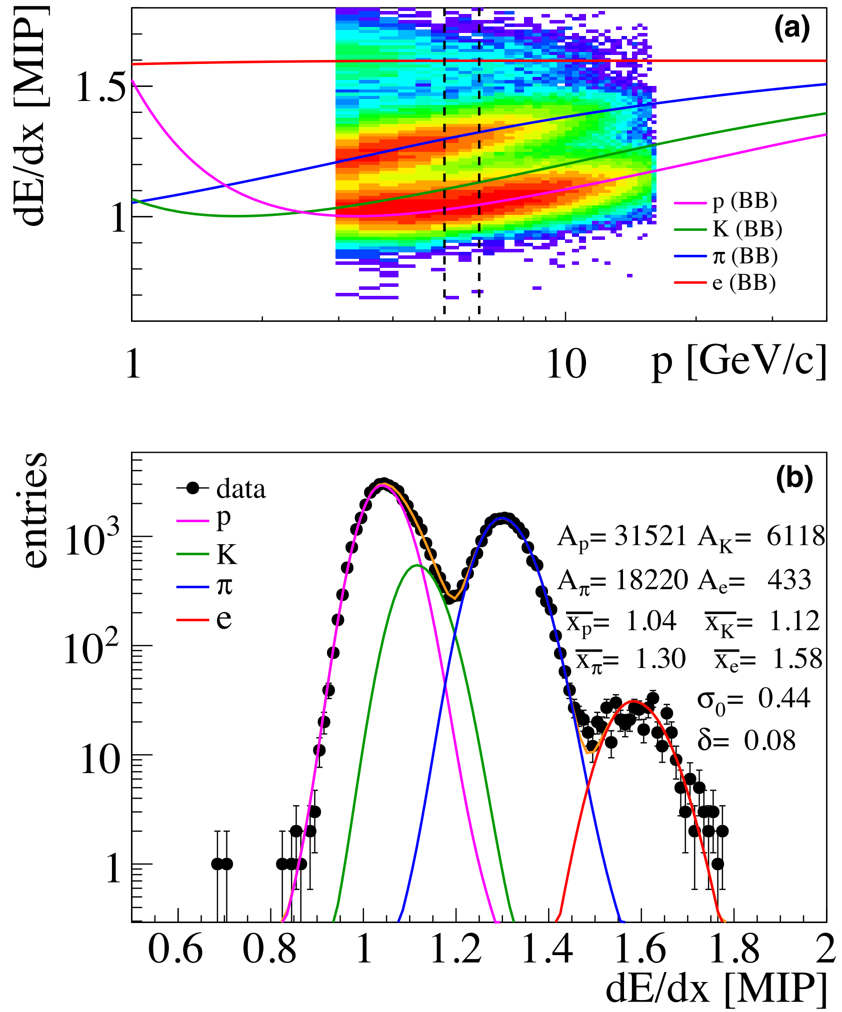
Particle Identification

measured specific energy loss is in fact distorted:

- Statistical nature of process (Landau)
  - Perform several (n) measurements along the track

    $\sigma_{\langle dE/dx \rangle} \propto 1/\sqrt{n}$

- Detector specific effects:
  - pad size
  - gas pressure
  - …



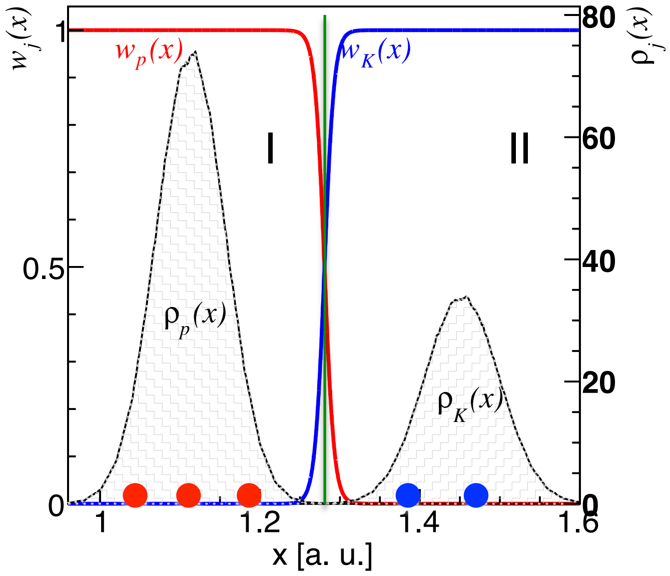
NA49: PRC 89 (2014) 054902

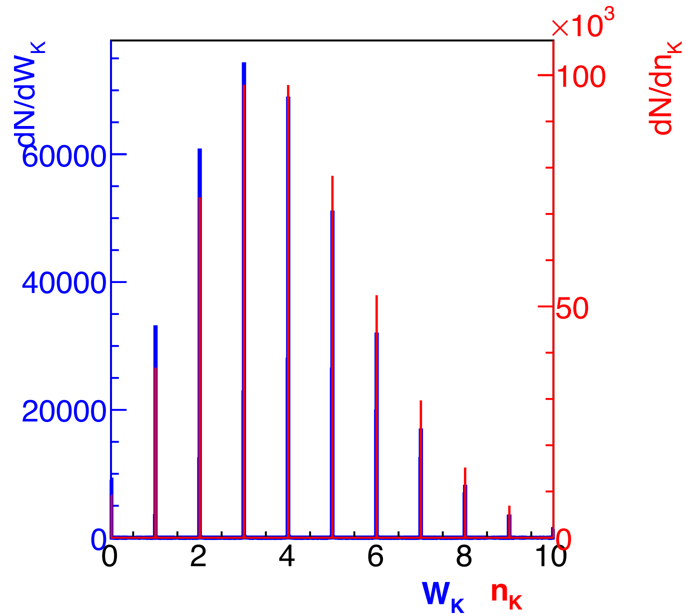
Identity method "in a nutshell"

(developed in close collaboration with NA49 and NA61/SHINE)

M. Gazdzicki et al., PRC 83, 054907 (2011)
M. I. Gorenstein, PRC 84, 024902 (2011), second moments
A. Rustamov, M. I. Gorenstein, PRC 86, 044906 (2012)

New way of particle counting





single event: 3 protons, 2 kaons

standard approach

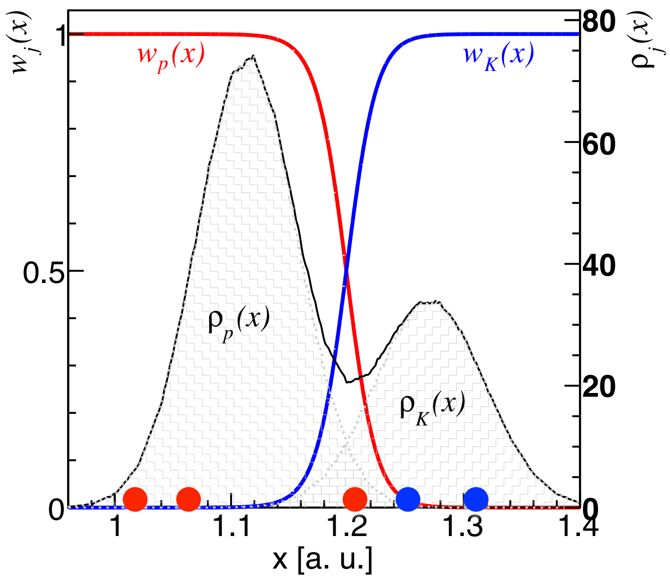
measurement in region I
count as proton
measurement in region II
count as kaon

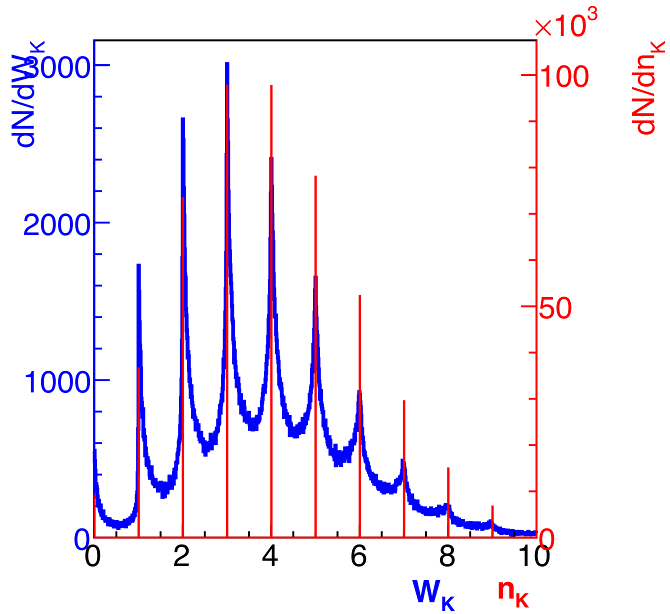
new approach

$$w_p(x_i) = \frac{\rho_p(x_i)}{\rho_p(x_i)+\rho_K(x_i)} \qquad W_p = \sum_{i=1}^{5} w_p(x_i)$$

$$w_K(x_i) = \frac{\rho_K(x_i)}{\rho_p(x_i)+\rho_K(x_i)} \qquad W_K = \sum_{i=1}^{5} w_K(x_i)$$

Folded multiplicity distribution





standard approach

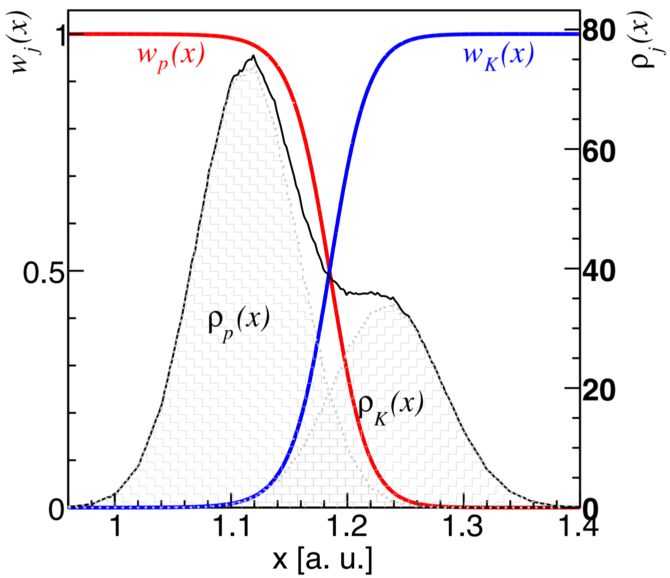
use additional detector information
or reject a given phase space bin

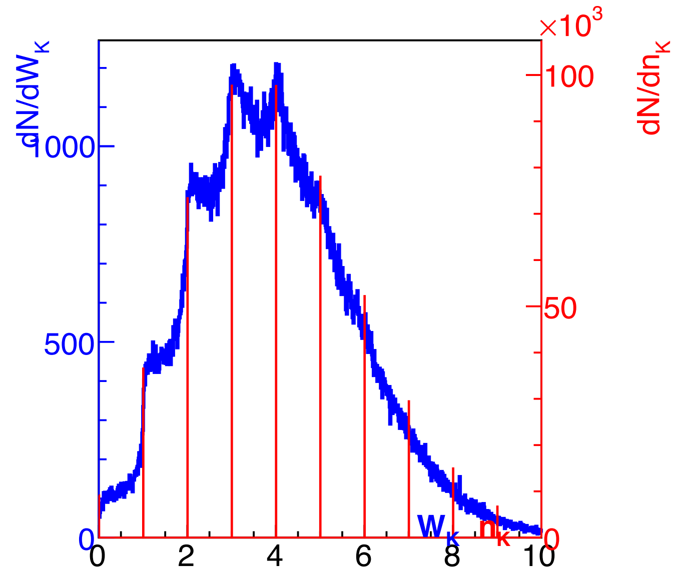
new approach

gives folded multiplicity distribution

(standard unfolding problem)

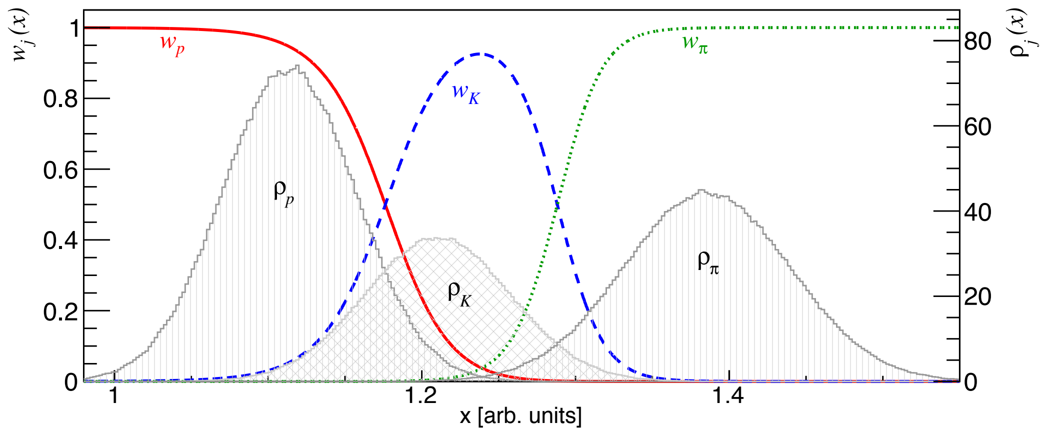
Folded multiplicity distribution

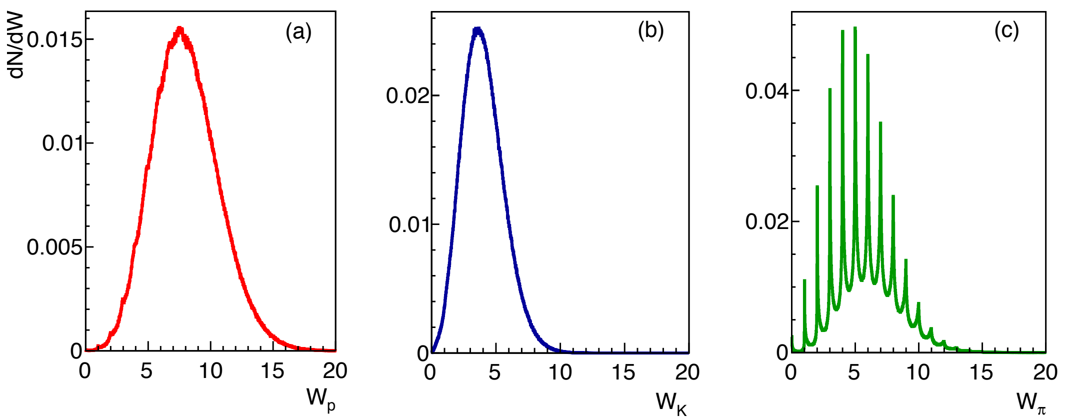




More overlap -> Stronger folding

3 – particle example





Unfolding

instead of unfolding the multiplicity distribution we unfold its moments

$$\left\langle \mathrm{W}_p \right\rangle = \frac{1}{\mathrm{N}_{\mathrm{events}}} \sum_{i=1}^{\mathrm{N}_{\mathrm{events}}} \mathrm{W}_p^i \equiv \left\langle \mathrm{N}_p \right\rangle \quad \left\langle \mathrm{W}_p^n \right\rangle \neq \left\langle N_p^N \right\rangle \quad \left\langle W_p \cdot \ldots \cdot W_q \right\rangle \neq \left\langle N_p \cdot \ldots \cdot N_q \right\rangle$$

$$\langle \vec{W} \rangle = \mathrm{A} \times \langle \vec{N} \rangle$$

vector of folded moments = response matrix x vector of true moments

$$\langle \vec{N} \rangle = A^{-1} \langle \vec{W} \rangle$$

$$V\left(\langle \vec{N} \rangle \right) = A^{-1} V\left(\langle \vec{W} \rangle \right) \left(A^{-1} \right)^T$$

Working example for 2 particles

The Identity Method relates corresponding moments of *W* and multiplicity distributions through a set of linear equations. An example for the second moments:

$$\begin{pmatrix} \langle N_p^2 \rangle \\ \langle N_k^2 \rangle \\ \langle N_p N_k \rangle \end{pmatrix} = \begin{pmatrix} \overline{w}_{pp}^2 & \overline{w}_{pk}^2 & 2\overline{w}_{pp}\overline{w}_{pk} \\ \overline{w}_{kp}^2 & \overline{w}_{kk}^2 & 2\overline{w}_{kp}\overline{w}_{kk} \\ \overline{w}_{pp}\overline{w}_{kp} & \overline{w}_{pk}\overline{w}_{kk} & \overline{w}_{pp}\overline{w}_{kk} + \overline{w}_{pk}\overline{w}_{kp} \end{pmatrix}^{-1} \begin{pmatrix} \langle W_p^2 \rangle - b_p \\ \langle W_k^2 \rangle - b_k \\ \langle W_p W_k \rangle - b_{pk} \end{pmatrix}$$

3 equations, 3 unknowns
(unique solution)

$$b_i = \sum_{j=p,k} \langle N_j \rangle (\overline{w_{ij}^2} - \overline{w}_{ij}^2), \quad b_{pk} = \sum_{j=p,k} \langle N_j \rangle (\overline{w}_{pkj} - \overline{w}_{pj}\overline{w}_{kj})$$

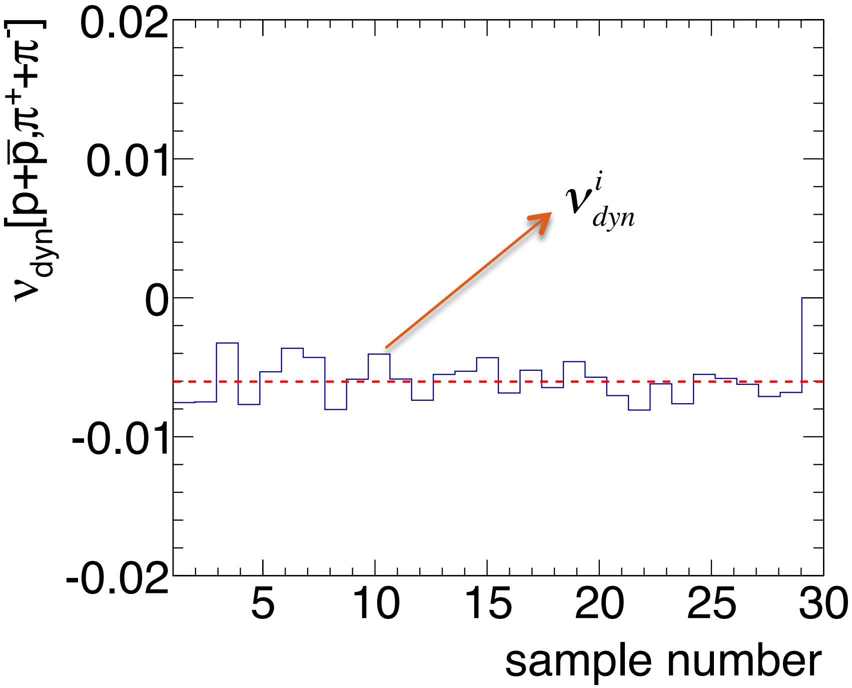
$$\overline{w}_{ij} = \frac{\int w_i(m)\rho_j(m)dm}{\int \rho_j(m)dm} \qquad \overline{w_{ij}^2} = \frac{\int w_i^2(m)\rho_j(m)dm}{\int \rho_j(m)dm} \qquad \overline{w}_{ikj} = \frac{\int w_i(m)w_k(m)\rho_j(m)dm}{\int \rho_j(m)dm}$$

- Advantages:
  - Event-by-Event fits of PID variable is not needed
  - Also no need for event mixing
  - Mathematically proven

M. Gazdzicki et al., PRC 83, 054907 (2011)
M. I. Gorenstein, PRC 84, 024902 (2011), second moments
A. Rustamov, M. I. Gorenstein, PRC 86, 044906 (2012) (third and higher moments)

Error Estimation

Instead of using analytical error propagation for each fluctuation measure we use 2 methods:



1. Sub-sampling method

- Data is subdivided into N (30) samples
- For each sample v_{dyn} is calculated

$$\langle \nu_{dyn} \rangle = \frac{1}{N}\sum_{i=1}^{N} \nu^{i}_{dyn} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(\nu^{i}_{dyn} - \langle \nu_{dyn} \rangle\right)^{2}}{N-1}}$$

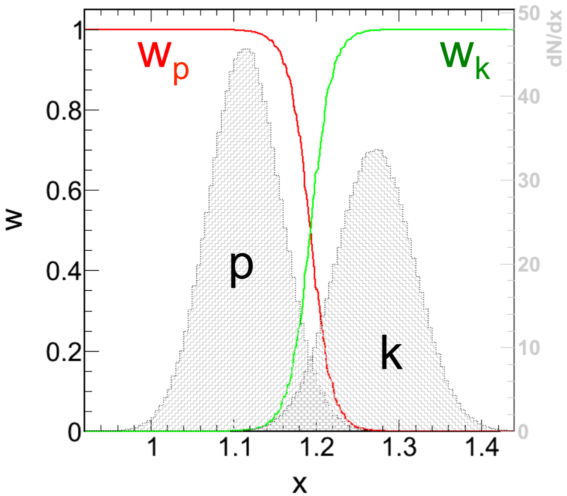
$$\sigma_{\langle \nu \rangle} = \frac{\sigma}{\sqrt{N}}$$

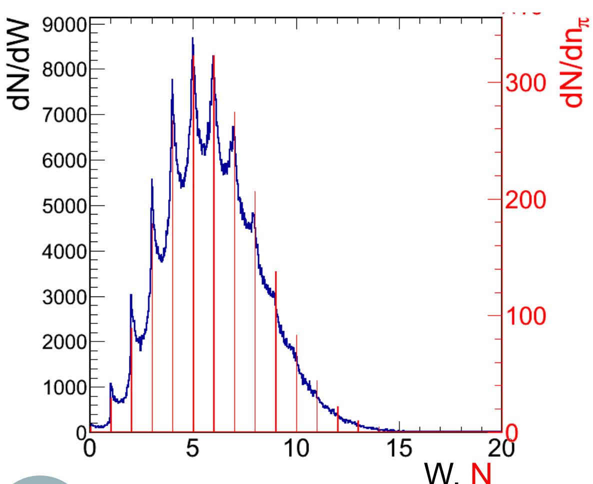
$$\nu_{dyn} = \langle \sigma_{dyn} \rangle \pm \sigma_{\langle \nu \rangle}$$

2. Statistical bootstrapping
(not discussed in this talk)

The strategy

2 particle example





- Available information:
  - inclusive distribution of PID variable, $\rho_j(x)$
- The Problem:
  - how to find the moments of multiplicity distributions?
- The strategy:
  - for each measurement x and particle j in an event one defines

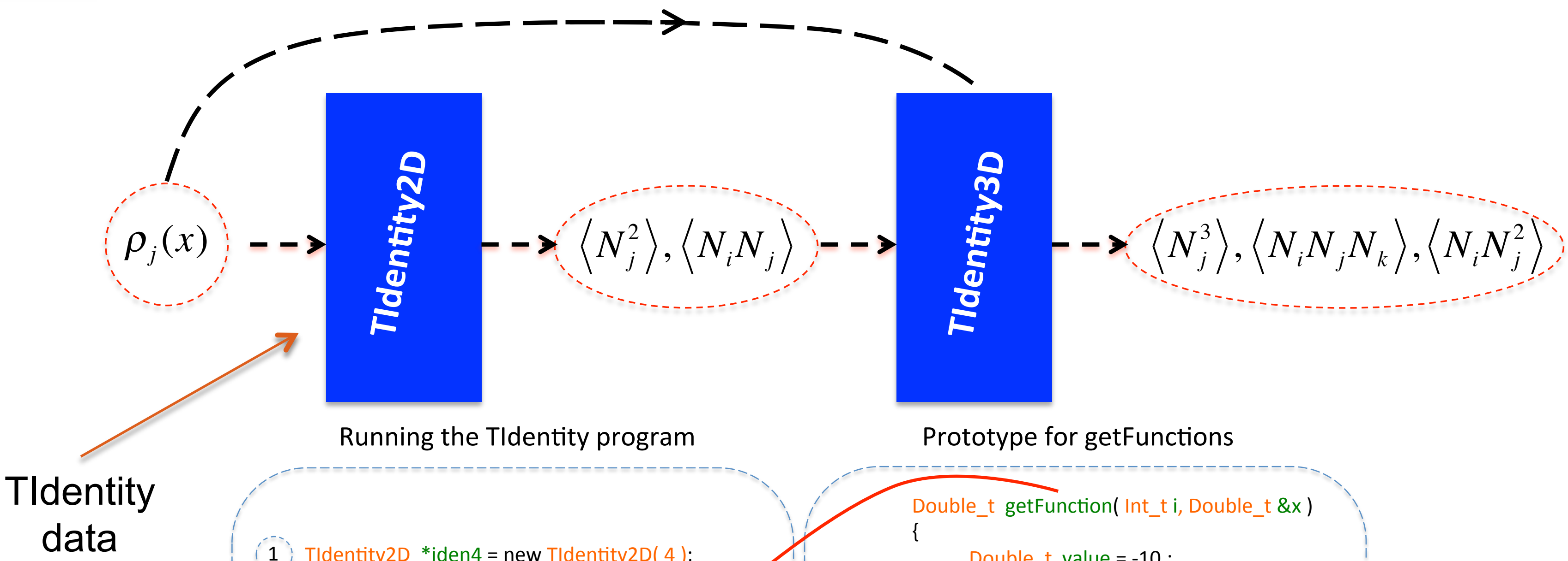
$$w_j(x) = \frac{\rho_j(x)}{\sum_j \rho_j(x)}$$

  - for each event one constructs:

$$W_j = \sum_i w_j(x_i)$$

- finally one calculates moments of W distribution
- Use the Identity method
  - to find moments of the multiplicity distributions from known moments of W quantities

TIdentity Module



Running the TIdentity program

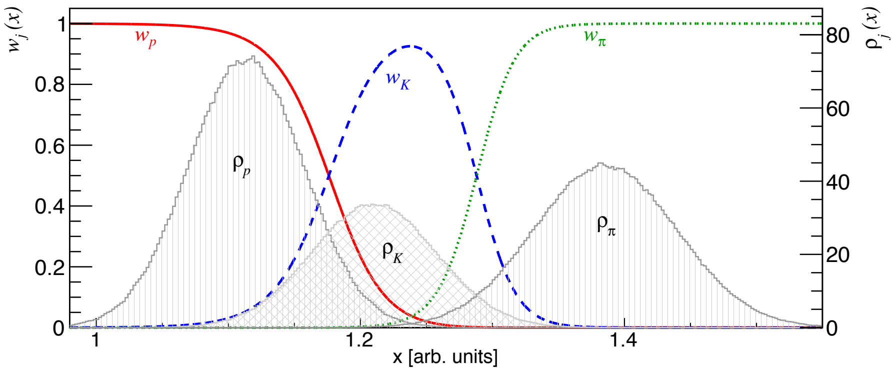
```
1  TIdentity2D  *iden4 = new TIdentity2D( 4 );
2  Iden4 -> SetInputDirectory();
3  Iden4 -> SetFunctionPointers( getFunction);
4  Iden4 -> Process();
5  Double_t  Np2   =  Iden4 -> GetMoments( proton, 2 );
6  Double_t  Npi2  =  Iden4 -> GetMoments( pion, 2 );
   ...
```

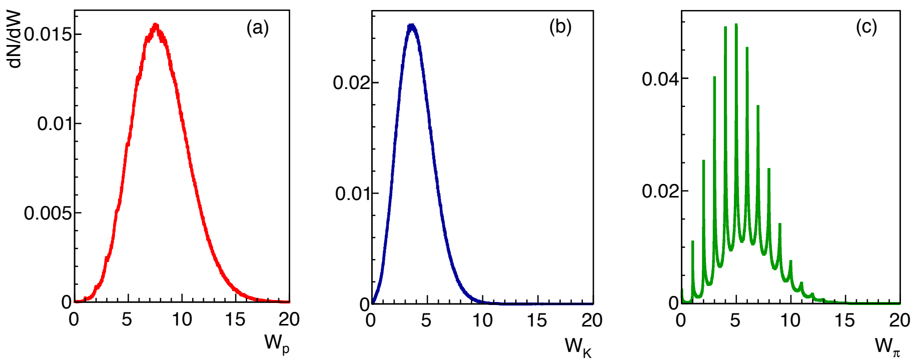
Prototype for getFunctions

```
Double_t  getFunction( Int_t i, Double_t &x )
{
        Double_t  value = -10.;
        switch ( i )
        {
        case proton :
           value  =  rhop(x);
           break;
        case kaon :
           value  =  rhoK(x);
           break;
         ...
         }
        return value;
}
```

Ready for publication

Test on simulated data



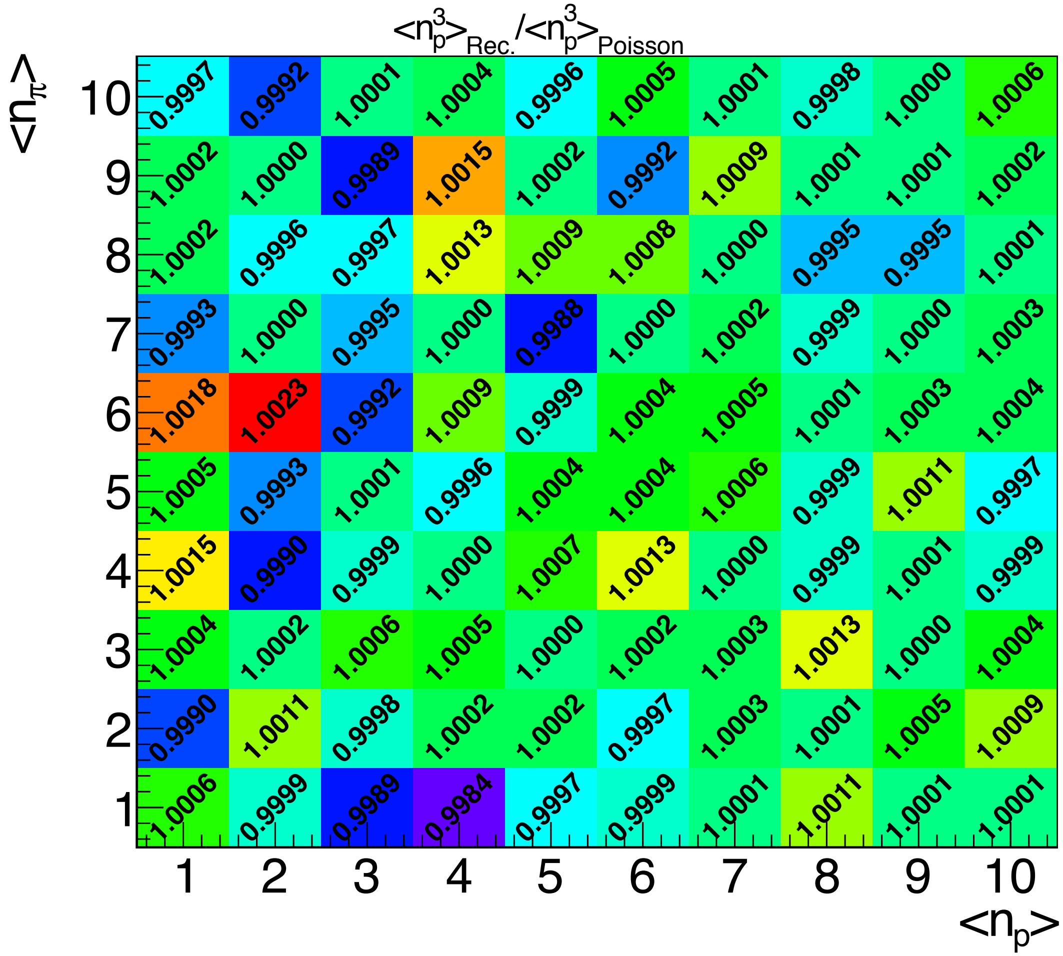


$$< p >= 8, < \pi >= 6, < K >= 4$$

A. Rustamov, M. I. Gorenstein, PRC 86, 044906 (2012)

Moments	Reconstructed (Identity method)	Generated	Poisson
$<p^2>$	72.0429 ± 0.0122	72.0452	72
$<\pi^2>$	42.0033 ± .0073	42.0011	42
$<K^2>$	20.0066 ± 0.0052	20.0057	20
$<p\pi>$	48.0178 ± 0.007	48.0184	48
$<pK>$	32.0147 ± 0.0057	32.0154	32
$<\pi K>$	24.0026 ± 0.0033	24.002	24
$<p^3>$	712.6391 ± 0.1925	712.6718	712
$<\pi^3>$	330.0547 ± 0.0878	330.0316	330
$<K^3>$	116.045 ± 0.0501	116.0372	116
$<p^2\pi>$	432.307 ± 0.0954	432.3185	432
$<p^2K>$	288.1968 ± 0.0710	288.2081	288
$<\pi^2K>$	168.0083 ± 0.033	167.9971	168
$<\pi^2p>$	336.155 ± 0.0786	336.1525	336
$<K^2p>$	160.0986 ± 0.0518	160.0875	160
$<K^2\pi>$	120.0275 ± 0.0320	120.0174	120
$<p\pi K>$	192.0765 ± 0.0379	192.0845	192

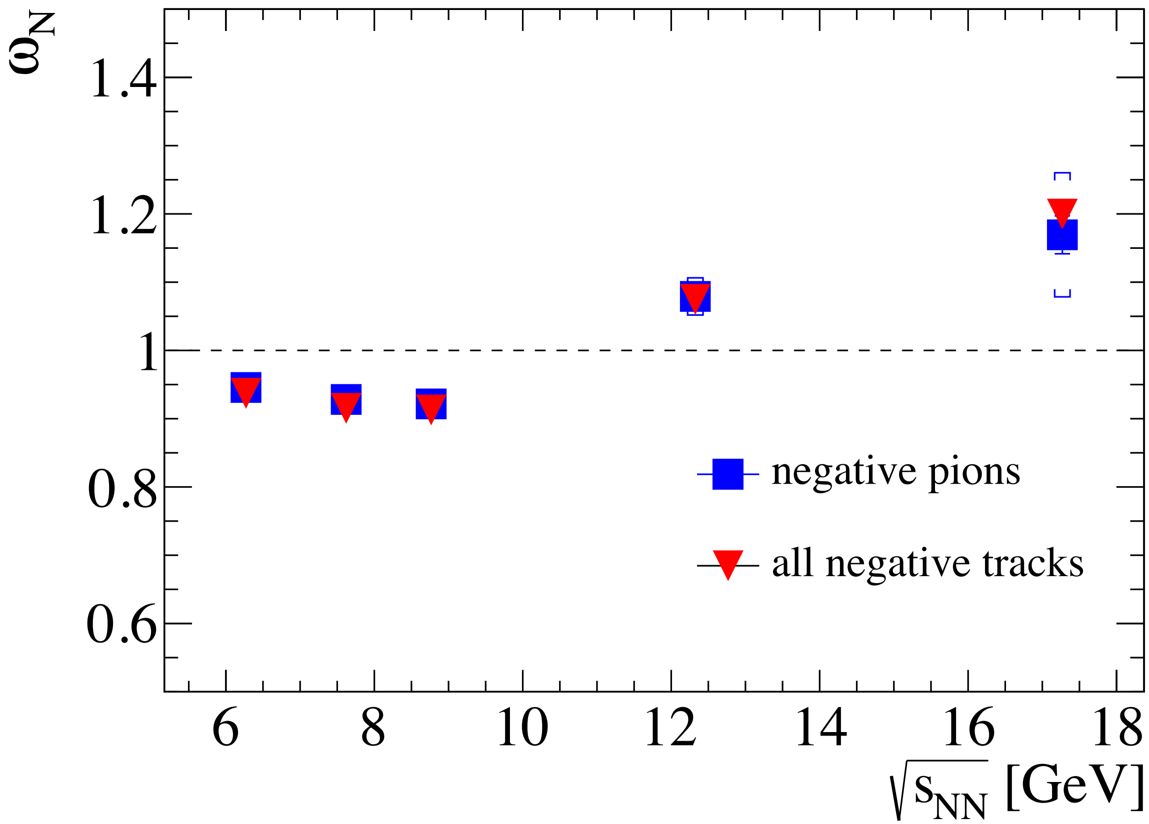
Test on simulated data



Each box contains $6*10^6$ events
$\langle n_K \rangle = \langle n_p \rangle$ to test the method in case of strong correlations

A. Rustamov, M. I. Gorenstein, PRC 86, 044906 (2012)

Test on real data of NA49



NA49: PRC 89 (2014) 054902

Some experimental results

Results from Identity Method

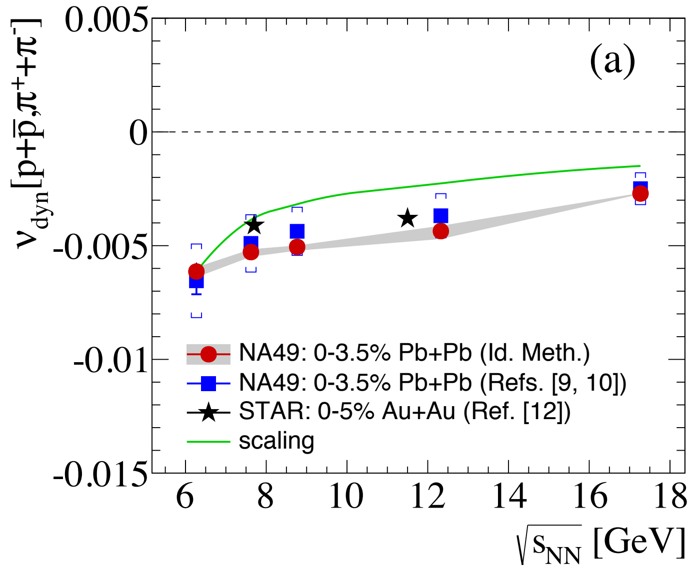
second moments for central (0-3.5%) Pb+Pb data

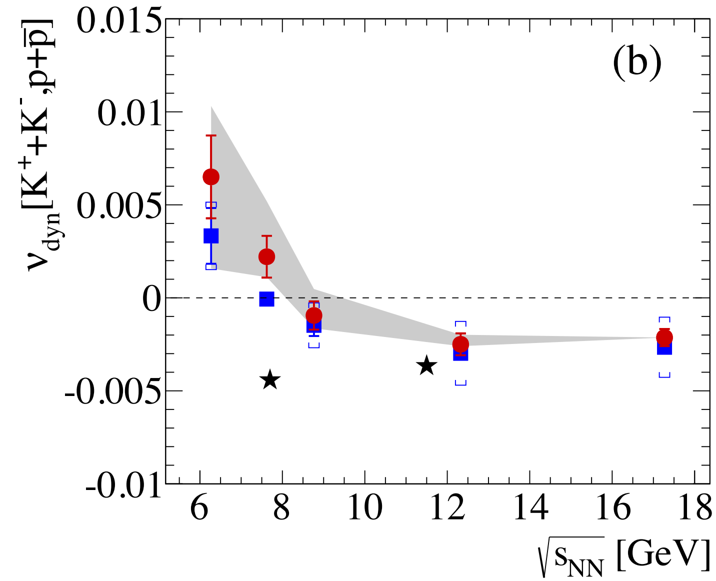
		20A GeV/c	30A GeV/c	40A GeV/c	80A GeV/c	158A GeV/c
input	$\langle N_p \rangle$	27.1786	34.876	38.186	47.5179	70.1685
	$\langle N_\pi \rangle$	30.5385	66.4564	103.046	226.819	413.295
	$\langle N_K \rangle$	4.5723	9.2489	13.6526	31.042	56.8712
output	$\langle N_p^2 \rangle$	764.277	1248.27	1493.64	2304.68	4969.01
	$\langle N_\pi^2 \rangle$	964.311	4487.12	10737.9	51850.7	172014
	$\langle N_K^2 \rangle$	25.395	94.9134	200.563	997.228	3312.25
	$Cov[N_p N_\pi]$	2.12232	4.29659	9.15544	39.03744	32.00979
	$Cov[N_p N_K]$	-0.73635	-0.62464	0.41682	3.48935	7.5732
	$Cov[N_K N_\pi]$	-1.01266	-1.2876	0.30418	15.6246	110.6174

$$Cov[N_1,N_2] = \langle N_1 N_2 \rangle - \langle N_1 \rangle * \langle N_2 \rangle$$

NA49: PRC 89 (2014) 054902

Results from Identity Method

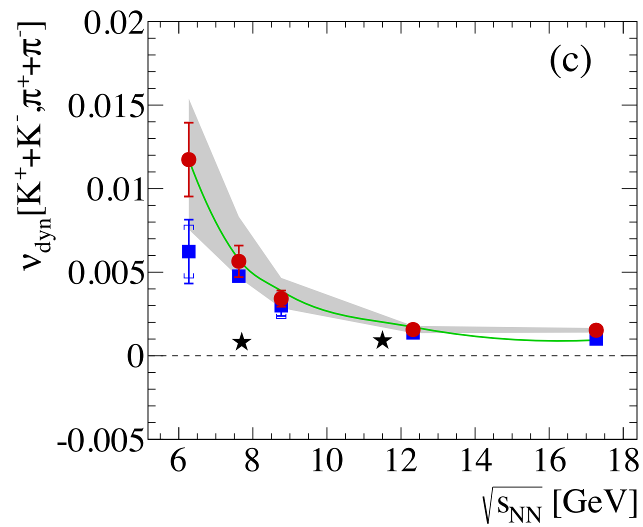




[p,π]: agreement with both, published results of NA49, and STAR

[K,π]: increasing trend at low energy published by NA49 is reproduced. Difference with STAR remains!

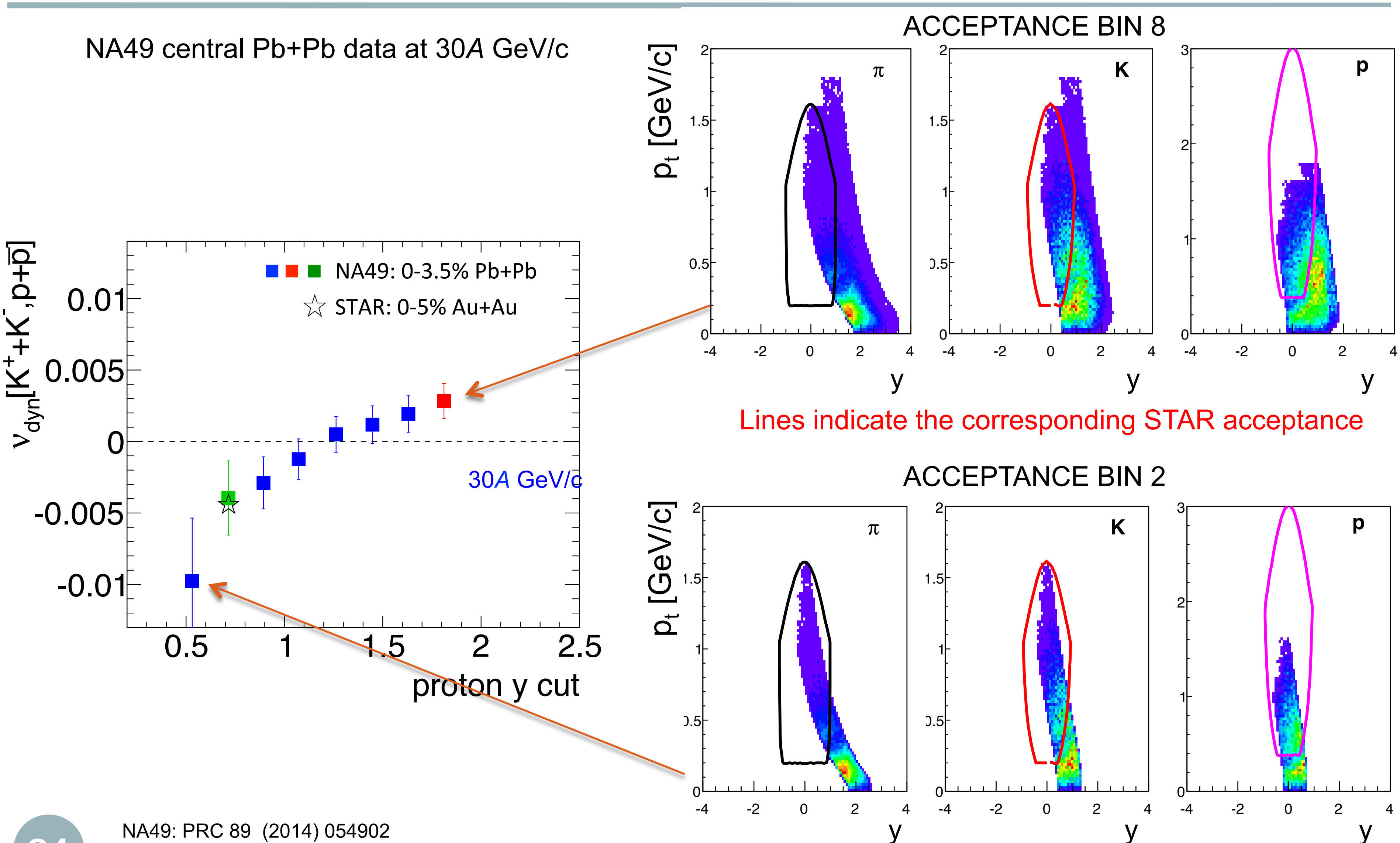
[K,p]: increasing trend at low energy published by NA49 is reproduced. Difference with STAR remains!



NA49: PRC 89 (2014) 054902

Dependence on acceptance

NA49 central Pb+Pb data at 30*A* GeV/c



NA49: PRC 89 (2014) 054902

Thank you for your attention