Dear Mikhail and members of the paper preparation group,

My apologies for the late reply to the second release of this draft.

While I realize the changes made in the draft, I do share the sentiment already expressed by members of the Collaboration that the paper is not yet ready for submission and major issues still need to be addressed. The draft also needs numerous wording/style/grammar changes. As for the first release, I will not deal with these now and will wait till the major issues are properly resolved.

My comments are below listed by line nr wrt to the version of the paper attached to your email below.

Best, Itzhak

L 30: How was the luminosity determined? It is not clear what luminosity you are quoting here. The relevant quantity is the luminosity for each target. So you should have 5 different numbers and not only one. I think all the information on luminosity should be in a short paragraph in Section 2 or 3.

Answer: Luminosity is calculated as a product of the integrated flux of the argon beam to the number of nuclei in the targets on the way of the beam ion, namely as L = N(Ar)*N_A*L/A, N(Ar) – integrated flux of Ar beam, L (g/cm2) – target thickness, A – target atomic number.

Updated sentence, lines 29-30:

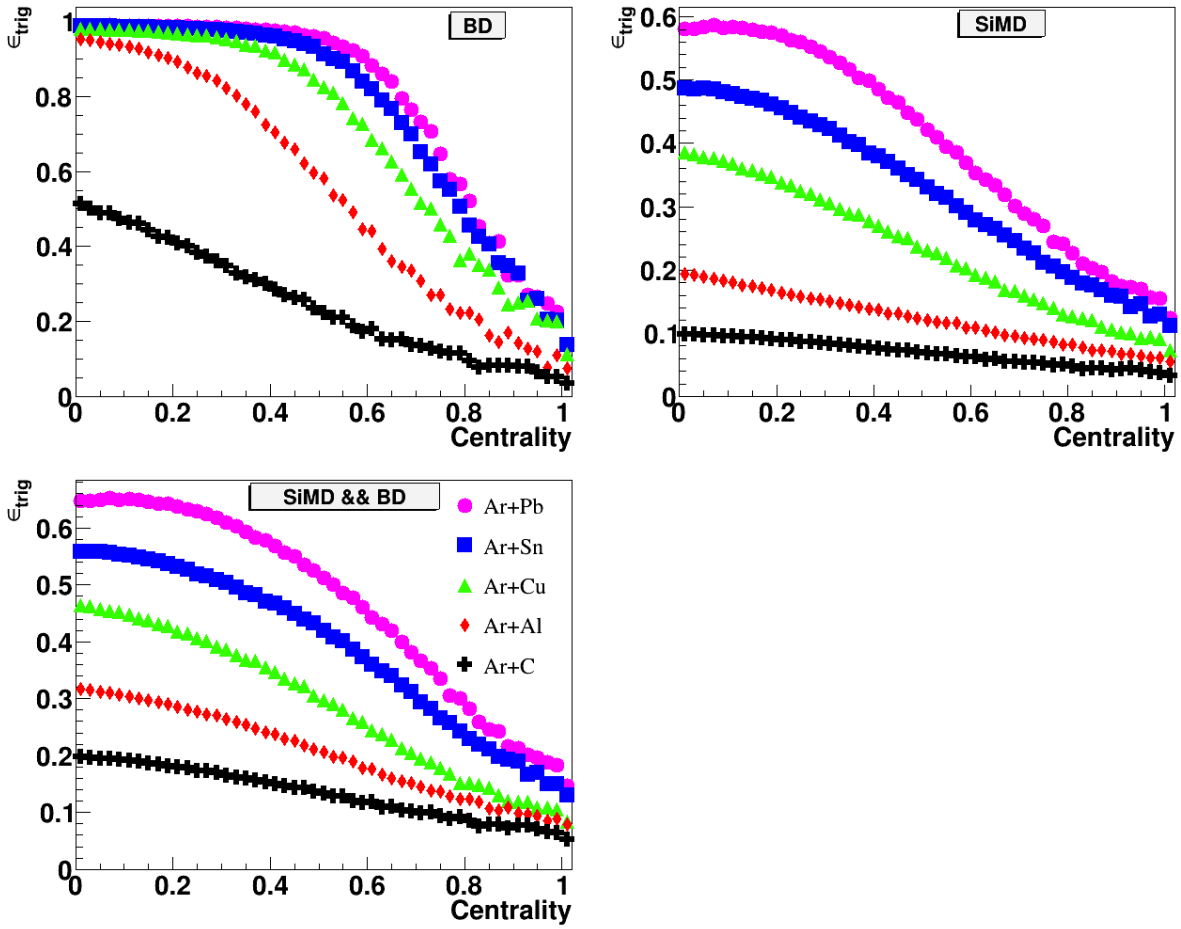
The experimental data correspond to an integrated luminosity of 7.8 μb collected with different targets: 2.1 μb (carbon), 2.3 μb (Al), 1.8 μb (Cu), 1.1 μb (Sn), 0.5 μb (Pb).

L 53-54: the wording of this sentence is not clear. Do you mean that the vertex is measured with a resolution of 2.4 mm?

Updated sentence, lines 53-54:

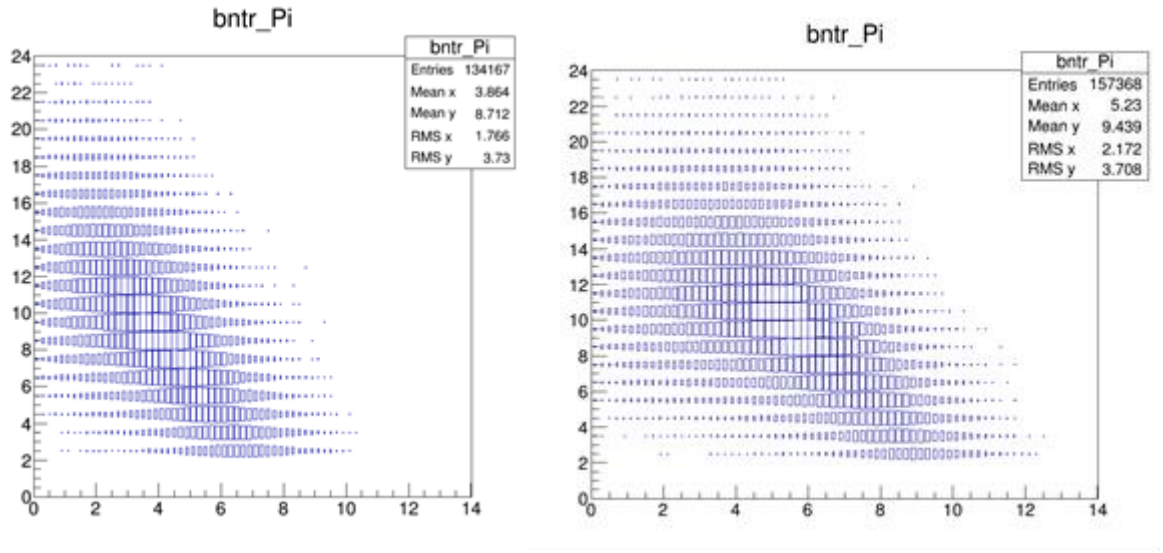
The vertex position is measured with a resolution of 2.4mm in the X-Y plane and in the Z direction at the target position. (or along the X-Y-Z coordinates).

L 62-70: the trigger description is better now. However, you are still not specifying what fraction of the total cross section you are triggering on or what is the trigger efficiency vs centrality. See below also comment on the trigger efficiency.



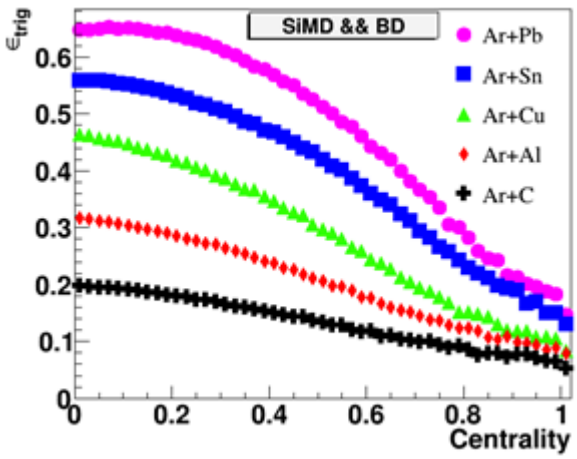
L 66-68: it seems that the values of m and n were kept constant for a given target and were changed only when the target was changed? If so, you should say that explicitly.

Answer: The centrality manifest itself by the track multiplicity. The average track multiplicity per event is between 8 and 11 measured for the Ar+C and Ar+Pb interactions. The trigger efficiency is measured for each target as a function of the track multiplicity and applied to the experimental data as an event weight. Using the correlation plot between the track multiplicity and the event centrality (centrality is a function of the collision impact parameter) taken from the DCM-SMM simulation, the dependence of the trigger efficiency on the centrality is evaluated. Three figures above show the trigger efficiency as functions of the centrality for the BD>3 ("BD") , SiMD>3 ("SiMD") trigger conditions and for the combination of all BD>3, SiMD>3 and BD>1&FD>2 ("SiMD && BD") trigger conditions. Two figures below show the correlation of the reconstructed track multiplicity (vertical axis) vs the collision impact parameter b (horizontal axis, fm) taken from simulated events of Ar+Al (left plot) and Ar+Sn (right plot) interactions.



Updated text, lines 63-70:

To form a trigger signal, the following logic conditions were applied: 1) BT⊗(BD≥ 4); 2) BT⊗(SiMD≥ 4); 3) BT⊗(BD≥ 2)⊗(SiMD≥ 3). The trigger conditions were varied to find the optimal ratio between the event rate and the trigger efficiency for each target. The BD trigger condition 1) was applied for 60% of data collected with the carbon target and continuously reduced with the atomic weight of the target down to 26% for the Pb target. The fraction of data collected with the SiMD trigger condition 2) was rising from 6% for the carbon target up to 34% for the Pb target. The rest of data were collected with the combined BD&SiMD trigger condition 3). The measurements cover the whole range of event centralities, but the trigger efficiency was lower for peripheral interactions than that for central and semi-central collisions. The trigger efficiency averaged over all data collected with trigger conditions 1), 2), 3) is presented in Fig.\ref{TrigEff}.



Figure{TrigEff}. Trigger efficiency averaged over data collected with trigger conditions 1), 2), 3) and evaluated as a function of the event centrality.

L 92-93: It is not clear what do you mean by "CSC (DCH)" or by "TOF-400 (TOF-700)". If you mean that you require a matching hit in either the CSC or the DCH and in either TOF-400 or TOF-700, then say that explicitly.

Updated text, lines 91-93:

Candidates to π+ and K+ should originate from the primary event vertex and match hits in CSC and ToF400 or in the DCH and ToF-700 detectors.

L 100: Why the vertex cut was placed at -3.4 cm? from the plot in Fig. 7e it seems that you are triggering on background collisions. Did you have an empty target run to justify this cut??

Answer: We do not have an empty target run. There are no construction elements or detectors down to -80 cm relative the Z target position. The Monte Carlo simulation of interactions in the target describes the lower tail of the Z vertex distribution. We took a wide range to be on safe side - do not lose the statistic.

L 113-5: should be moved to L 53 so as to have the information on the primary vertex in one place.

Answer: done

Fig. 4 and the background under the pi and K peaks: the procedure as described is not convincing. Are you mixing with events of the same centrality class? If not (and I assume that this is the case) I would expect the background to be considerably underestimated. So the red spectra in Fig.4 are quite surprising. Is this an absolute comparison or were the red spectra normalized to the real event spectra?

Answer: Mixed event spectra are normalized to the statistics in the M2 ranges between pi+ and K+ peaks and above the K+ peak in every bin of (y, pt). The mixed event spectra were evaluated and subtracted in data and as well as in Monte Carlo events. Below the figure 10e from the analysis note illustrates the normalization of the mixed event spectra in data and Monte Carlo in different y bins used in the pi+ analysis. It was also checked that the shape of the mixed event spectra are consistent for all mixed events and for mixed events with selected number of tracks (i.e. centrality class).

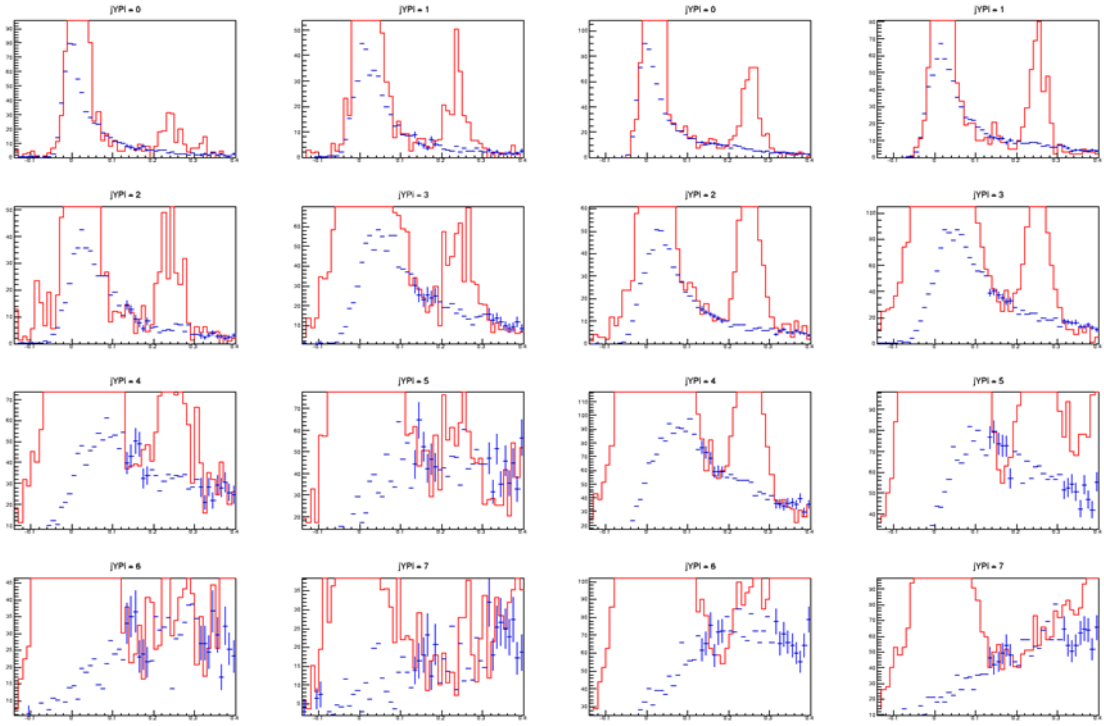


Fig. 10e. Spectrum of M2 in bins of y of π+ identified in ToF-400 in Ar+Sn interactions at 3.2 AGeV argon beam energy: left) experimental events, right) simulated events. Background (blue histogram) is taken from mixed events and normalized to the red signal histogram in the mass squared range between the π+ and K+ peaks (blue points with the errors).

L 134-136: the fraction of fake combinations should have a strong centrality dependence and this is even not mentioned. Even if you do not have the tools for a precise centrality determination, you could see the effect by comparing events with similar total number of hits in one of the detectors.

Answer: The fraction of fake combinations of tracks and hits in the ToF detectors is estimated from mixed events and subtracted from mass squared M2 spectra of identified particles. It was checked that "fake" hits are taken from mixed events with the number of tracks similar to original non-mixed events. The number of tracks from the event vertex correlated with the event centrality.

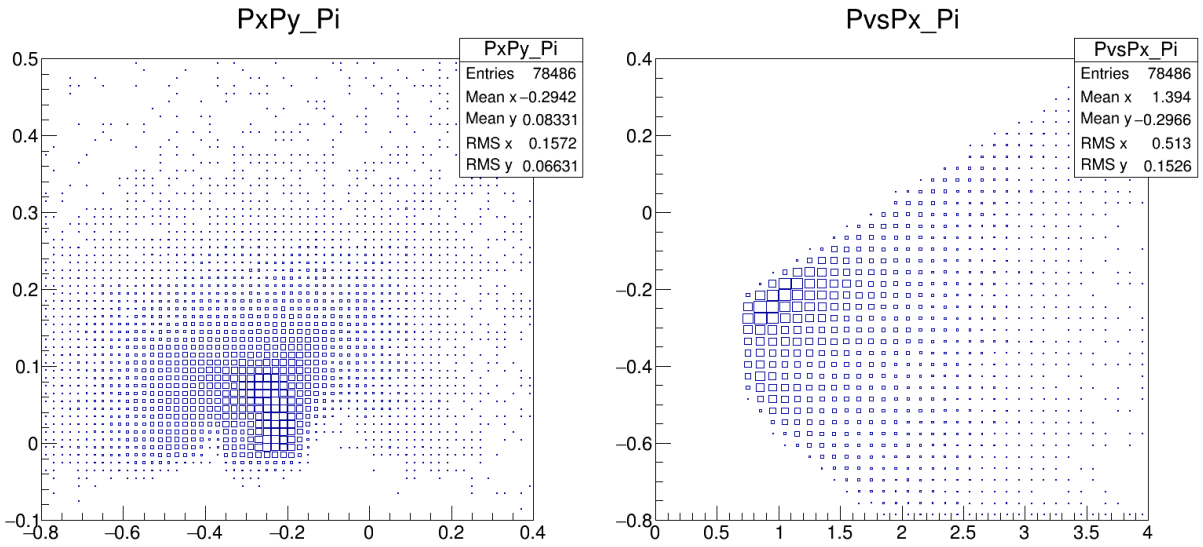
Fig. 5 and L 137-139 should be moved to Section 2, after L 75.

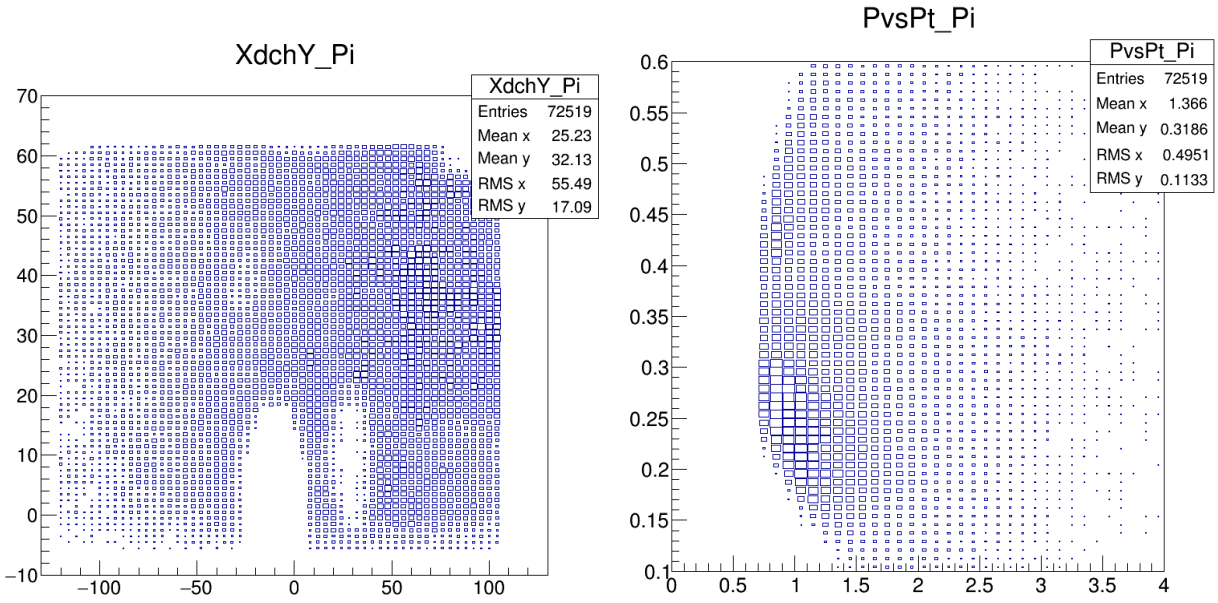
Answer: Section 2 is a short experimental setup description while Fig. 5 and comments are related to measurement procedure description. It is logical to keep the text as it is.

Section 4 should be labeled Reconstruction and Trigger Efficiency.

Answer: done

Fig. 8: There are no error bars on the points? If the statistics is so good, can you explain the jump in efficiency for pT = 0.25 GeV?





Answer: Jump around pT~0.25 GeV in the efficiency of the DCH - ToF-700 system is due to the empty holes designed for the beam in the DCH and ToF-700 detectors visible in the XdchY coordinate plot and in the PxPy transverse momentum plot (affects pT<0.25 GeV) and due to the second “blind” spot in these distributions where the real Ar beam crossed the sensitive areas of these detectors. The beam tracks were not reconstructed because they crossed the GEM detectors through the GEM beam hole. Plots PvsPx and PvsPt show limitation of the pT below 0.25 GeV due to the acceptance rising with the total momentum P.

L 161-169: the trigger efficiency was calculated using "events with reconstructed pi+ and K+ mesons". Why? This is not the experimental trigger used online. Why don't you use the full sample of events recorded with the special trigger conditions based on BD alone or SiMD alone? Your procedure introduces a bias that is even not mentioned in the draft. I think that this needs to be addressed in order to convince the reader that the procedure used to determine the absolute cross section is correct.

Answer: We follow a standard procedure of the efficiency evaluation. In case of signal in form of narrow peak in the invariant mass distribution the efficiency is evaluated in the narrow mass region limited by the peak width. In our case we use the M2 regions of the pi+ and K+ signals for the efficiency evaluation. The first aim of the analysis is to calculate the cross section for processes of pi+ and K+ production. To do this we need to know the efficiency of the trigger for these types of processes (pi+ and K+ production). Indeed, we use events recorded with the special trigger conditions based on BD alone or SiMD alone, apply additional trigger conditions, respectively, in SiMD and BD, and calculate the SiMD and BD efficiencies. We cannot use the full events sample for the trigger efficiency calculation because in many triggered events pi+ and K+ mesons are not produced. As a result, the trigger efficiency will be calculated not for events of pi+ and K+ production, but for the mixture of all recorded events which we do not use in the analysis. We also found that the trigger efficiency for events with K+ is somewhat higher than that for pi+ events due to higher particle multiplicity in K+ events. Finally, to get pi+ and K+ multiplicities, we normalize the pi+ and K+ cross sections to the absolute inelastic cross section, taken from the parametrization of nucleus-nucleus data.

L 192-220: The discussion of systematic uncertainties is not clear to me or not very rigorous. In particular, why are you grouping them into two categories sys1-3 and then a bunch of other factors? From eq. 1, one sees 4 sources of systematic uncertainties: systematic uncertainty in the

yield n, in the rec, eff., in the trigger eff. and in the luminosity. I think you should have a discussion on these 4 sources and analyze the various factors that contribute to their systematic uncertainties. While doing that you should pay attention at possible correlations. For example the yield n is affected by pid cuts, background subtraction, matching cuts to outside detectors, variations of the tracking algorithm....but also the rec. eff. is affected by those same factors.

Additional text, lines after 203:

This uncertainty affects the number of reconstructed π+ and K+ in (y,pT) bins in data as well as in simulated events. As a result its effect is smaller for the pi+ and K+ yields calculated by formula (1).

Additional text, lines after 206:

In addition to uncertainties Sys1-Sys3, uncertainties of the normalization of the π+ and K+ yields were evaluated. These uncertainties are applied for all bins of y and pT because they are treated as fully correlated.

Update text, lines 211-219:

The luminosity uncertainty is estimated to be within 2%. It corresponds to the fraction of the beam which can miss the target, evaluated from the vertex positions of the data. The statistical uncertainty of the trigger efficiency is 28% for K+ detection in Ar+C interactions and between 7.5% (Ar+Al) and 4% (Ar+Pb) for K+ detection in interactions of argon ions with heavier targets. The statistical uncertainty of the trigger efficiency for π+ detection ranges between 4.5% (Ar+C) and 0.9% (Ar+Pb). The uncertainty of the central tracking detector efficiency is estimated to be within 3%. Its effect is evaluated from the remaining difference in the number of detector hits per track in the simulation relative to the experimental data (see Fig.7d). The combined uncertainty of matching of extrapolated tracks to the CSC (DCH) hits and ToF-400 (ToF-700) hits is within 5%. It is estimated from the remaining difference in the matching efficiency in the simulation relative to the experimental data.

Update of table 1 (format taken from the analysis note):

Target / Systematics		π+				Target		K+			
	C sys %	***Al*** sys %	***Cu*** sys %	***Sn*** sys %	*Pb* sys%		***C*** sys%	***Al*** sys %	***Cu*** sys %	***Sn*** sys %	***Pb*** sys %
Sys1-Sys3	14	11	12	9	9		28	26	14	12	16
Sys1	7	6	6	4	4		15	14	7	7	9
Sys2	11	8	9	7	7		19	18	11	9	11
Sys3	5	5	5	4	4		14	12	5	5	7
Norm (trigger + tracking + luminosity)	7.8	6.3	6.2	6.2	6.2		29	10	8.4	7.6	7.4

L 227-237: I don't quite follow all these comparison statements between data and models and sometimes I don't see that Figs. 9 and 10 support these statements.

Answer: We have the results of our studies and results of model calculation. We presented the description of the comparison in our way. If you have more details to add or modify this part of text - please present your opinion a bit more explicit.

L 261-262: Fig, 15 does not support this statement.

Answer: The distributions are flat and UrQMD predicts twice bigger values. Please clarify where is a problem.

L 278-280: But within the experimental uncertainties, the measured ratios K to pion are fully consistent with PHSD.

Update of the text, lines 278-280:

They show no dependence on the number of nucleons-participants. The PHSD prediction is compatible with this result whereas the DCM-SMM and UrQMD models predict smooth rising of the K+ to pi+ ratio with A_part .

L 293: I don't quite see the factor of 3.5 for K:

For HADES you quote 2.8 10^-2. Table 3 Ar+Cu has 13.9 10^-2. That gives almost a factor of 5.

Update of the text, lines 292-293: are higher by factors of 5 and 1.3 relative to the results for kaons and pions measured by HADES.

L 296: could you quote the kinematic threshold for the production of pions and kaons?

Update of the text: The difference in the K+ multiplicities could be explained by the energy dependence of the K+ cross section close to the kinematical threshold for K+ production (Ethr(NN) \sim 1.58 GeV).

Fig. 17: what is the solid line in this figure?

Answer: parametrization of the pi-meson multiplicities in nucleon-nucleon reaction

Update of the text, line 316-317:

The reference [36] contains compilation of the pion data for nucleon-nucleon (N+N) [38], Mg+Mg [39], …

Professor Itzhak Tserruya

Dear Mikhail,

As I could not participate in the last collaboration meeting since the dates were in conflict with a previous commitment, I am listing below my comments on the draft. I am limiting the list to major comments on the content of the paper. In addition, the text needs considerable wording/style changes, but I am not dealing with them here.

With best regards,

Itzhak

-Two general comments first:

- Why results are presented only for the positively charged mesons pi+ and K+. Why the negatively charged ones pi- and K- are not shown?

Answer: The Time of Flight calibration of the ToF-400 arm detecting negative particles is not ready yet (but in good progress). We plan an additional paper comparing results for pi- and pi+ (as well as comparing a small signal of K- with K+).

- Why centrality is not at all mentioned in the paper? I shall elaborate further on this below.

Answer: An analysis to separate pi+, K+ data to centrality classes based on the number of tracks in the events vertex and on the number of hits in the trigger barrel detector is in progress but not ready yet (we need proper evaluation of systematic uncertainties). We plan to present results for two ranges of centrality (<40%, >40%) in the next paper comparing pi+ / pi- production.

Text added to the paper: **The measurements cover the whole range of event centralities but the trigger efficiency was lower for peripheral interactions.**

Abstract: the first two sentences do not belong to an abstract. They are better suited for the Introduction.=

1. Introduction:

I find this section rather chaotic: the energy range is mentioned in two different places. The interest in strangeness is mentioned in two different places. Reshuffling of text and better streamlining of the ideas are needed. I also suggest to add a short paragraph at the end of the Introduction specifying how the paper is organized.

Text to the paper: **The paper is organized as follows. Section 2 describes the experimental set-up and section 3 is devoted to details of the event reconstruction. Section 4 describes the evaluation of the p+, K+ reconstruction efficiency. Experimental results on transverse momentum and spectra and multiplicities of pi+ and K+ mesons are given in section 5. The BM@N measurements are compared with predictions of theoretical models and with the experimental data on middle-sized nucleus-nucleus interactions measured at lower energies. Finally, the results are summarized in section 6.**

2. Experimental set-up:

General comment: Should give a short description of all the detectors used in the analysis and when appropriate refer to published papers for more details.

L 49: specify the location of the Si detectors

Answer: we added references to the documents for details. Plot added to Fig.1 showing BD, SiMD, BC2, Veto trigger detectors.

L 67: elaborate more on the TOF detectors. Why do you need two measurements of the time of flight? Quote their time resolution.

Text added to the paper: **The acceptances of the ToF-400 [reference] and ToF-700 detectors [reference] cover different kinematical ranges of the rapidity and transverse momentum of identifies particles. The time resolutions of the ToF-400 and ToF-700 systems are 84 ps and 115 ps, respectively [reference].**

L 68-71: should explain better the trigger used, What and why "different conditions" were used? Is this a minimum bias or close to a minimum bias trigger? What is the relation of the trigger to centrality or to the total cross section? Or what fraction of the total cross section is the trigger selecting?

Answer: more details of the trigger logics are given.

Corrections to the text of the paper: **… ranging from 2 to 4. The trigger conditions were varied to find the optimal ratio between the event rate and the trigger efficiency for each individual target. The measurements cover the whole range of event centralities but the trigger efficiency for peripheral interactions is lower.**

L 70: first time the BD and FD are mentioned. What are these detectors? Where are they located? Why they are not shown in Figure 1?

Answer: Plot added to Fig.1 showing BD, SiMD, BC2, Veto trigger detectors.

3. Event reconstruction:

L 84-86: you are not using the Si hits in the track definition? Why?

Text added to the paper: **Hits in the forward silicon detectors were used for the track reconstruction but no requirement on the number of hits was applied.**

L 89: 7 sigma seems to be an over killer. Why such a high value?

Answer: The Z(vertex) distribution has tails wider than in the Gaussian distribution. The cut separates interactions in the target and the trigger counter.

Text added to the paper: **… cuts off interactions with the trigger detector situated at 3 cm behind the target.**

There is no discussion in this section (and not in the following one) on fake tracks. What is the fraction of fake tracks as function of pT? y?

Answer: Fake Si+GEM tracks are eliminated by matching them with the CSC (DCH) hits and signals in the ToF-400 (ToF-700) detectors. The fraction of fake combinations of tracks and hits in the ToF detectors was evaluated using “mixed events”, i.e. by matching tracks to hits in the ToF-400 and ToF-700 detectors originated from independent events.

Text added to the paper: **The fraction of fake combinations of tracks and hits in the ToF detectors was evaluated by the “mixed event” method described above. The “mixed even” fraction was found to differ for interactions of the beam with light and heavy targets and for different bins of the transverse momentum and rapidity.**

4. Reconstruction efficiency

L 132-140: You write that the detector efficiencies in the simulations were adjusted to the measured ones. But how were the detector efficiencies measured? Then you write in L 134: "The resulting... efficiency" resulting from what? The whole paragraph is not clear. Please try to explain better the procedure used.

Text in the paper: **The pi+ and K+ reconstruction efficiency is evaluated in intervals of the rapidity y and transverse momentum pT. It takes into account the geometrical acceptance, the detector efficiency, the efficiency of kinematic, spatial cuts and the losses of pi+ and K+ due to decays on flight.**

Text with details of the detector efficiency calculation was added to the analysis note: To measure the detector efficiency tracks were reconstructed with the detector under study excluded from the reconstruction. Requirements were applied to ensure the track quality: at least 4 hits out of 6 planes of the GEM detectors, 2 hits out of 3 planes of the forward silicon detectors .The efficiency was evaluated for the remaining 2 GEM planes and 1 silicon plane. All possible combinations of the GEM and silicon planes were analyzed. To ensure the quality of tracks for the CSC (DCH) efficiency calculation, only tracks matched to the ToF-400 (ToF-700) hits were selected. Alternately, the requirements of matching tracks to the CSC (DCH) hits were applied to select tracks for the ToF-400 (ToF-700) efficiency calculation.

L 142-3: why these different conditions were used?

Text to the paper: **The trigger conditions were varied to find the optimal ratio between the event rate and the trigger efficiency for each individual target.**

Could you show in a Figure the results of the trigger efficiencies?

Answer: The trigger efficiencies are given in the analysis note (Figures 13a, 13 b, tables 5a,5b).

5. Results

L 155: According to the formula used, what you call yields should be better called multiplicity.

Answer: OK

L 158: this is a differential cross section and should therefore be written as d^2sigma / dy.dpt

Answer: OK

L 174-184 and the subsequent paragraph:

- It is not clear what the difference is between these two.
- The trigger triggers on the event, it is not clear what do you mean by trigger efficiency for K+ or trigger efficiency for pi+.
- A table listing the various systematic uncertainties would be very useful.

Answer: We found that the trigger efficiency for events with identified K+ is somewhat higher that that for events with identified pi+ (tables 5a and 5b of the analysis note). To take this effect into account the trigger efficiency was evaluated separately for for events with reconstructed K+ and pi+.mesons.

Table 1 with the systematic uncertainties is added to the text..

Figs. 5 and 6: what is exactly plotted in these figures? The text says: "differential y spectra corrected for detector acceptance and efficiency" (They are not corrected for the trigger efficiency?). The vertical axis says: d^2N/dy.dpt. The formula in L 158 says N(y,pt). Please clarify and use consistent notation in the text and the figures.

Similar comment for Figures 7 and 8.

Text is changed in the paper: **The differential y spectra of the pi+ and K+ meson yields are calculated in the pT bins using formulae (1) and presented in figure …**

No discussion from the comparison of the experimental results with the various models in Figs. 5-8? The claim later in the text, in L 222, that the models describe the shapes of the distributions is generally not supported by the comparisons shown in these figures.

Answer: Text describing the data to model comparison is added to the paper .

L 231-235: How sensitive are the extrapolation factors to the model used to derive them? Could you quote the extrapolation factors obtained with the DCM-SMM model and compare them to the extrapolation factors obtained with UrQMD and PHSD?

Answer: extrapolation factors are calculated as an averaged value from predictions of DCM-SMM, UrQMD and PHSD models. They are given in table 2. The uncertainty is RMS between predictions of the models.

L 244-259: Do I understand it properly, that in the comparison of the BM@N results with both HADES and FOPI, the BM@N K+ results are a factor of 3-3.5 higher.

Answer: pi+ yields are comparable with HADES, FOPI for similar number of participants. K+ yields are higher by a factor of 3.5 due to a higher kinetic energy of the beam.

Text added to the paper: **It should be taken into account that the beam kinetic energy of the FOPI experiment (1.93 AGeV) is lower than that of the BM@N experiment.**