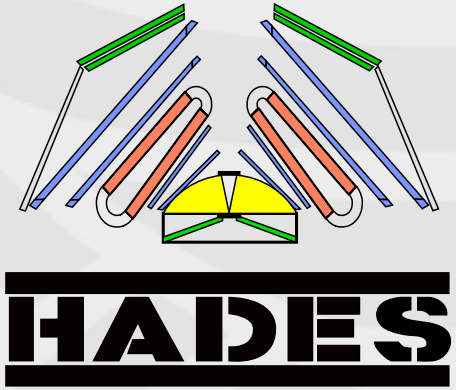


MACHINE LEARNING APPROACH TO REMOVE CARBON CONTAMINATION

Niklas Schild

(TU Darmstadt)

Analysis Meeting

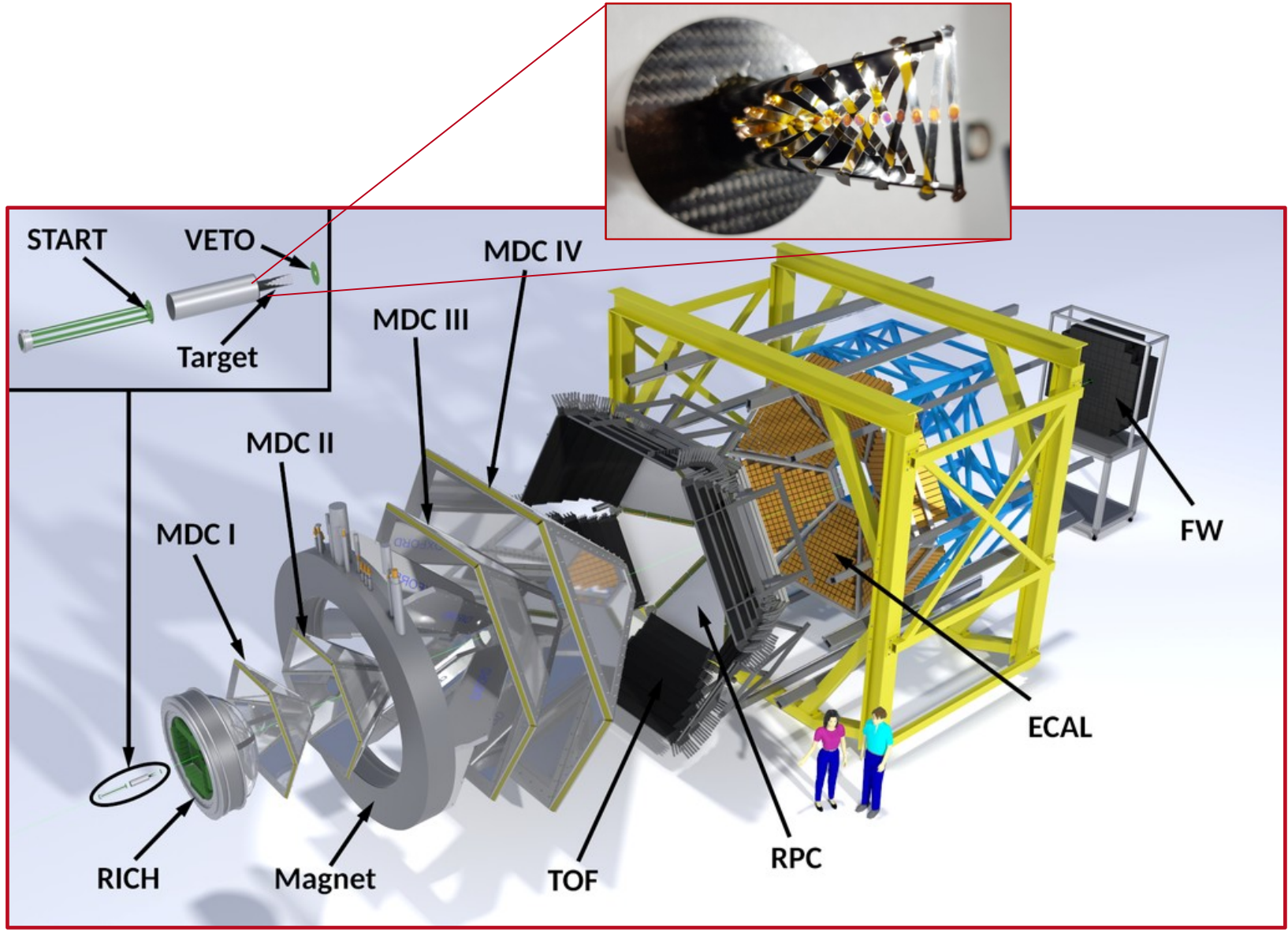
INTRODUCTION

This talk:

Focus on mar19

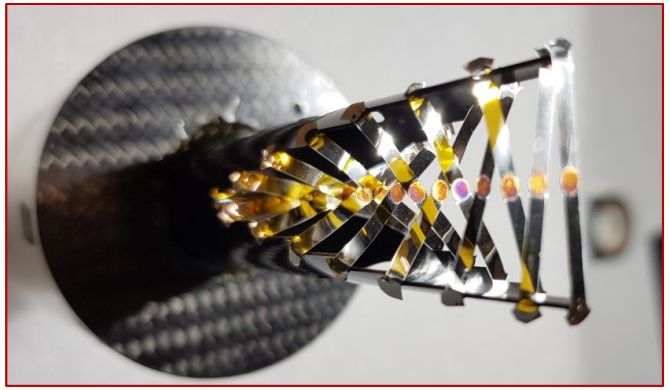
Ag+Ag $\sqrt{s_{NN}} = 2.55$ GeV

Ag+Ag $\sqrt{s_{NN}} = 2.42$ GeV



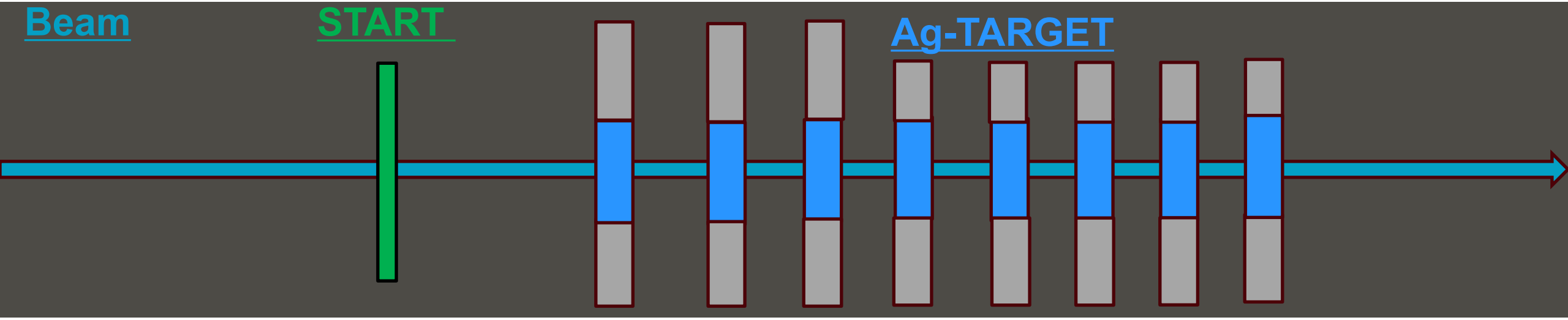
INTRODUCTION

Target consists of 15 Segments with kapton structure material and center of Ag

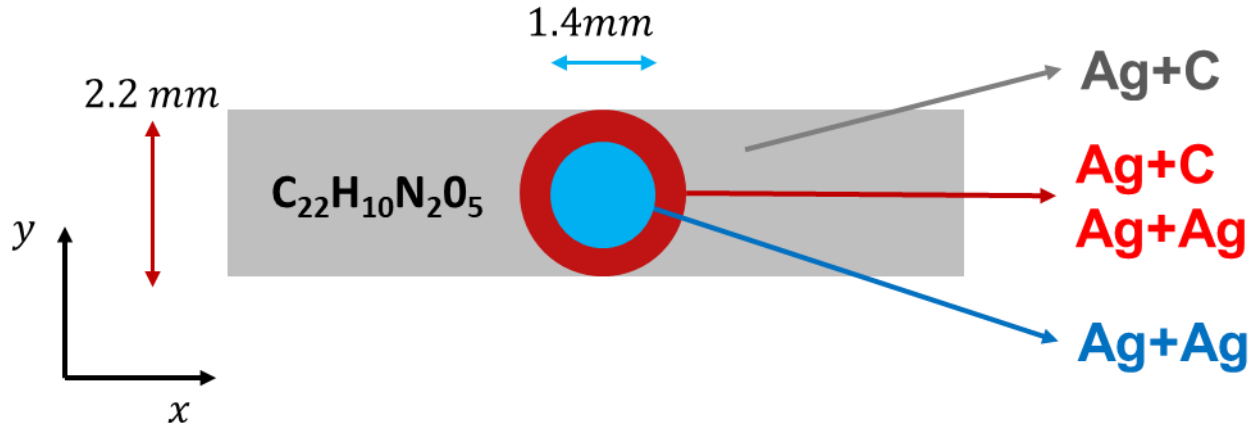


TECHNISCHE UNIVERSITÄT DARMSTADT

*not drawn to scale



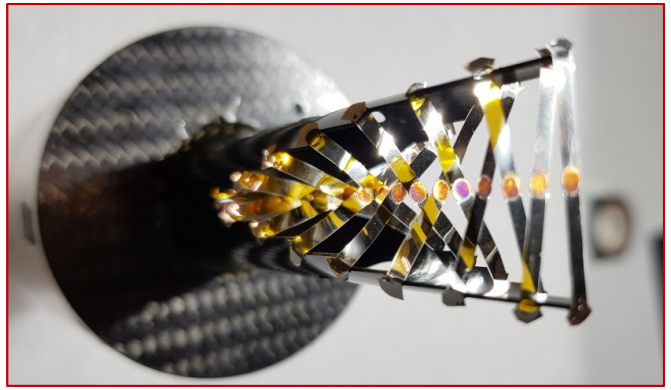
Possibilites for Ag+C collisions:



https://doi.org/10.1016/j.nima.2011.06.029

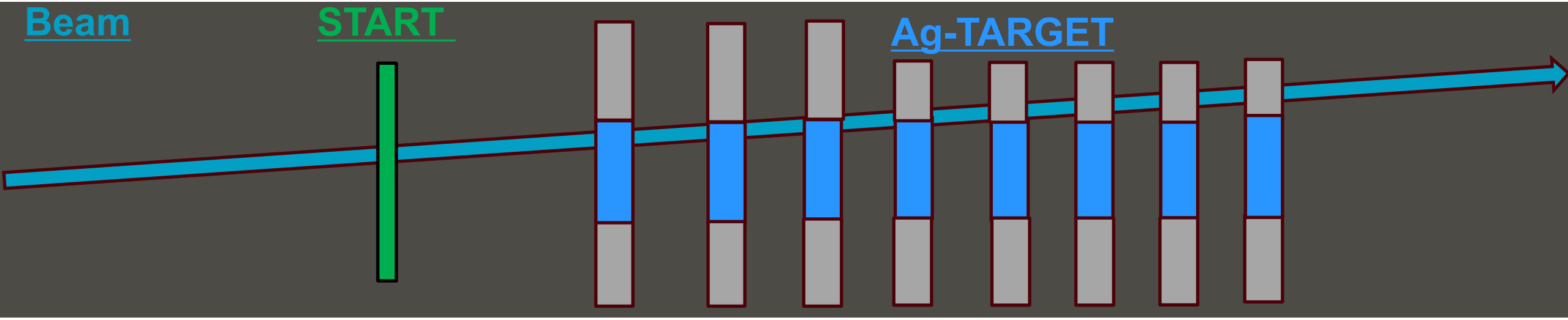
INTRODUCTION

Target consists of 15 Segments with kapton structure material and center of Ag



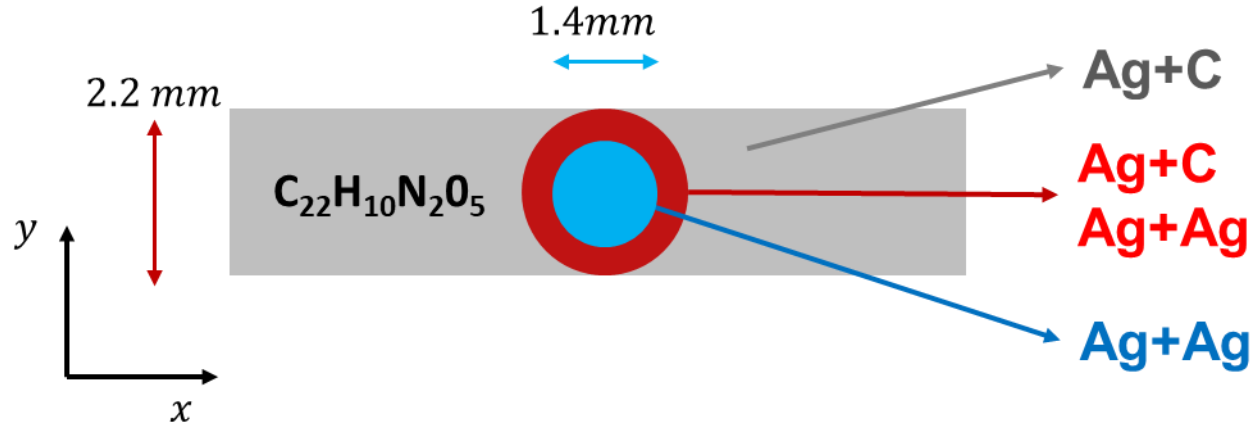
TECHNISCHE UNIVERSITÄT DARMSTADT

*not drawn to scale



Possibilites for Ag+C collisions:

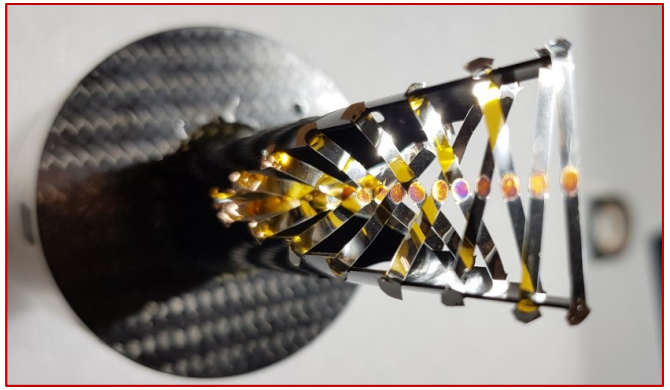
- Inclined beam



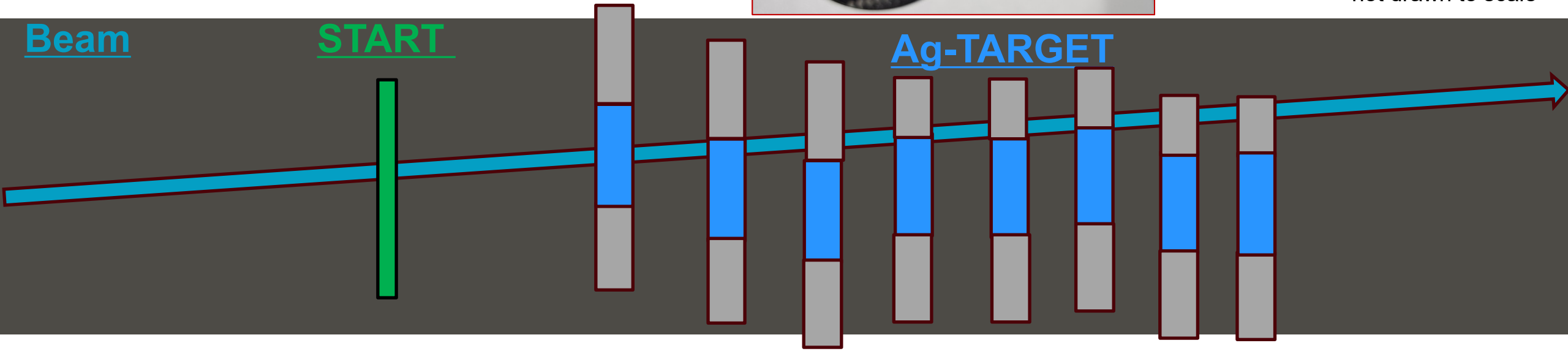
https://doi.org/10.1016/j.nima.2011.06.029

INTRODUCTION

Target consists of 15 Segments with kapton structure material and center of Ag

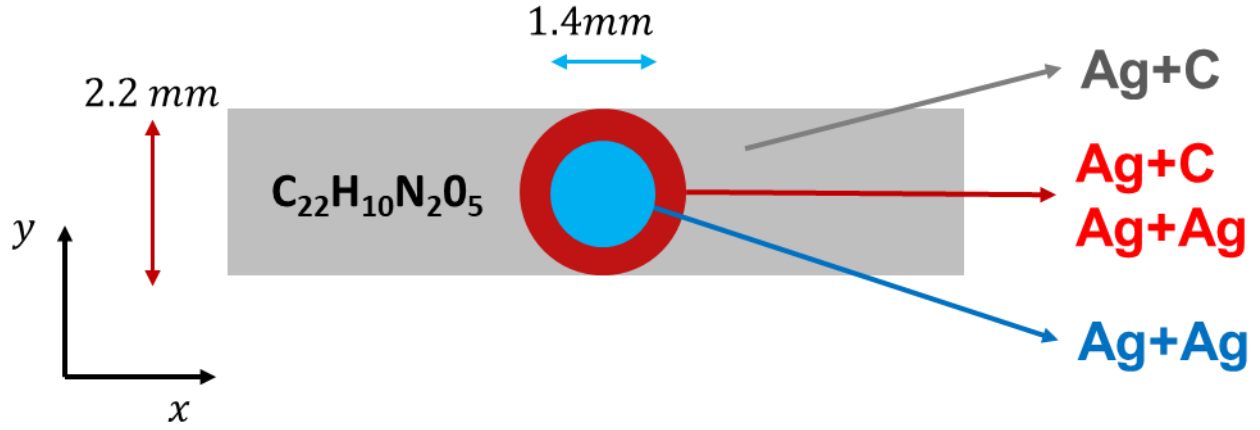


*not drawn to scale



Possibilites for Ag+C collisions:

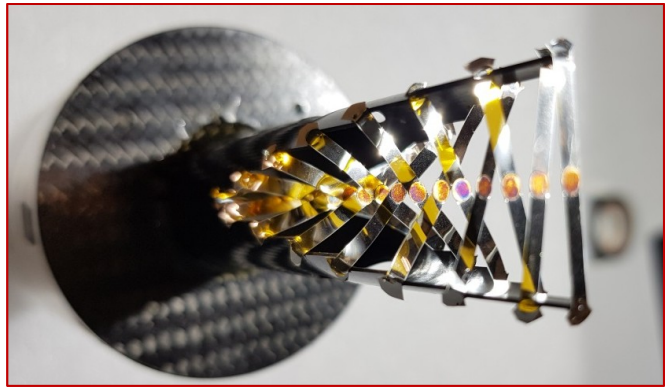
- Inclined beam
- Misaligned target segments



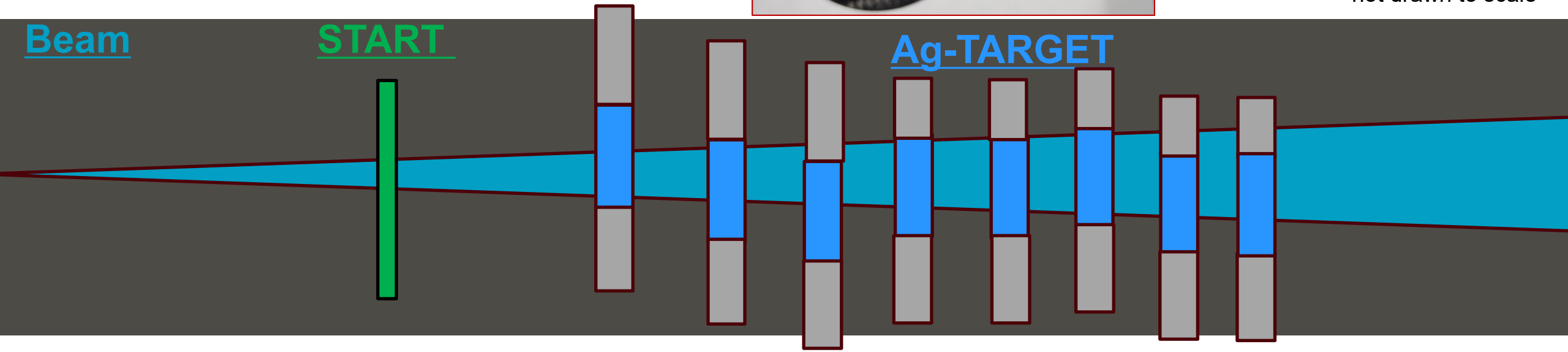
https://doi.org/10.1016/j.nima.2011.06.029

INTRODUCTION

Target consists of 15 Segments with kapton structure material and center of Ag

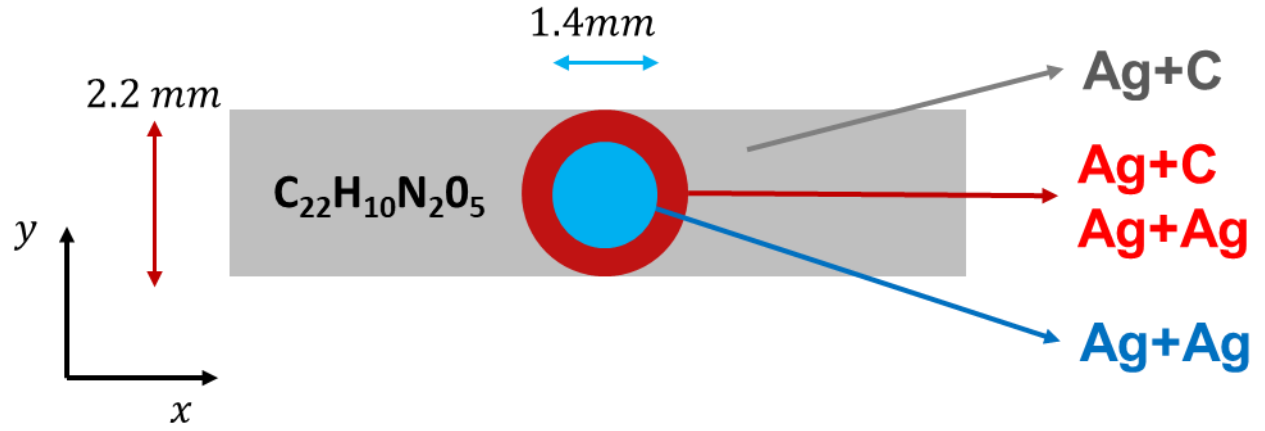


*not drawn to scale



Possibilites for Ag+C collisions:

- Inclined beam
- Misaligned target segments
- Beam divergence



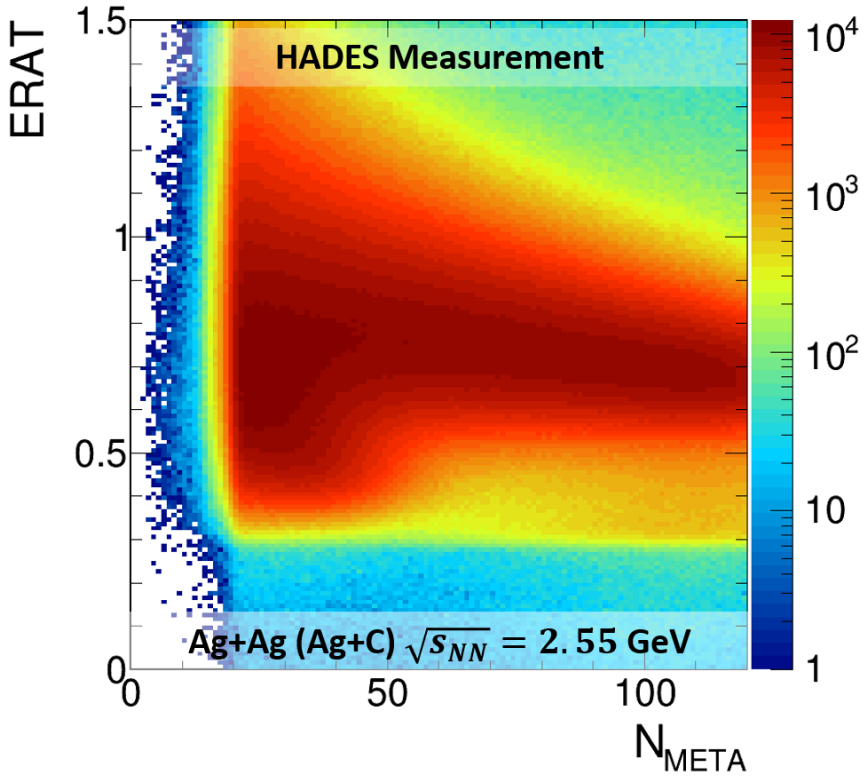
https://doi.org/10.1016/j.nima.2011.06.029

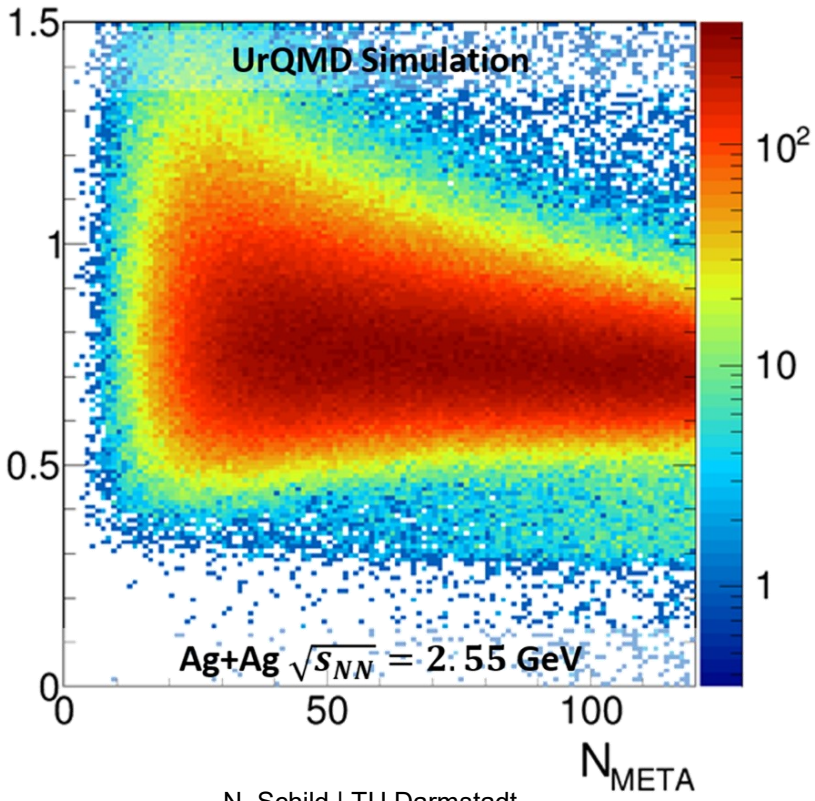
INTRODUCTION

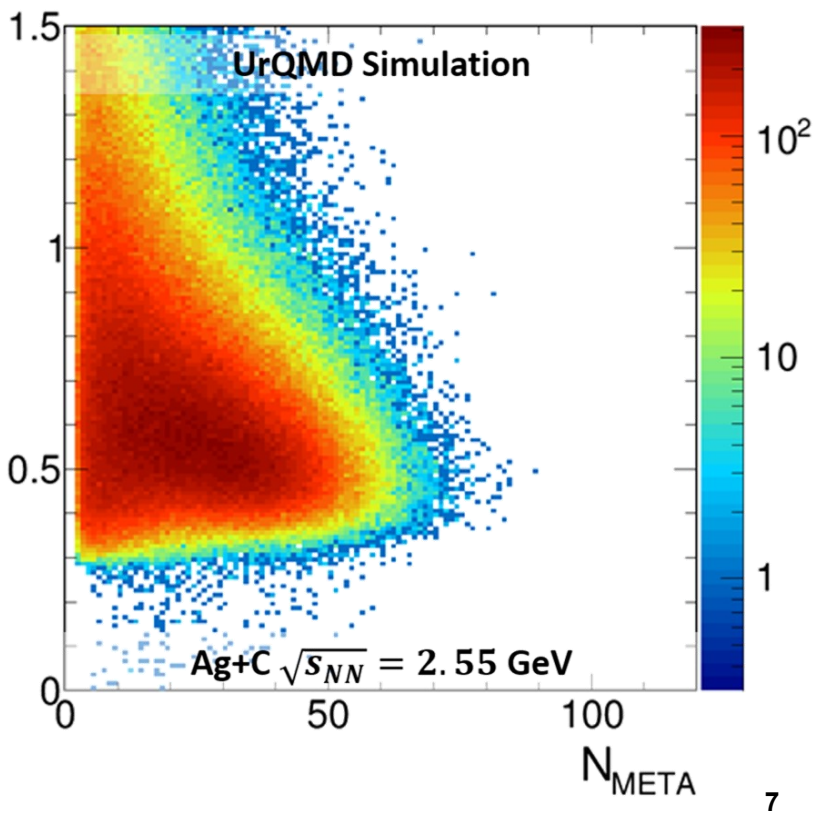
- Various observables show Ag+C contamination beyond ~ 30% centrality
- Visible in e.g. ERAT distribution
- No single observable able to sufficiently separate Ag+Ag from Ag+C

ERAT:
Ratio of transverse to longitudinal energy over all particles i

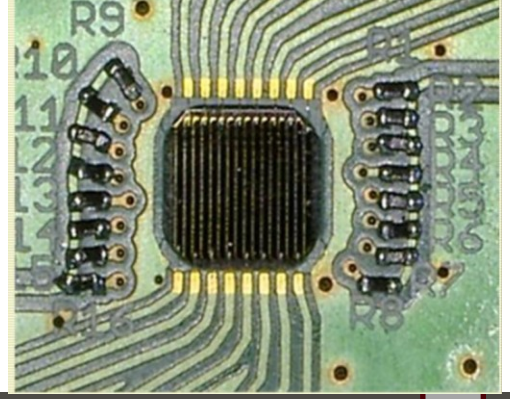
$$\mathrm{ERAT} = \frac{\sum_{i=1}^{N} E_i \ \sin\varphi}{\sum_{i=1}^{N} E_i \ \cos\varphi}$$





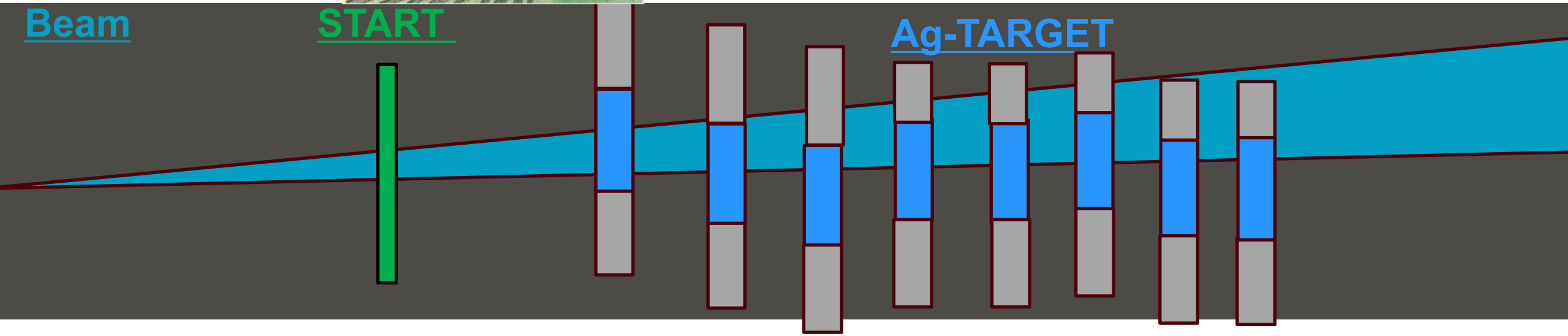


KEY IDEA



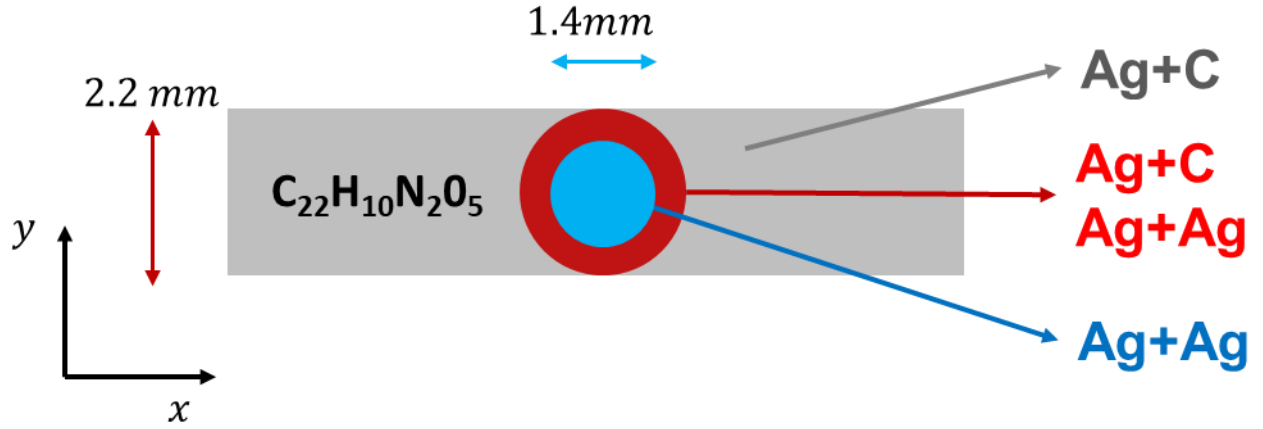
Correlate early segments with START hit position and extrapolate with machine learning

*not drawn to scale



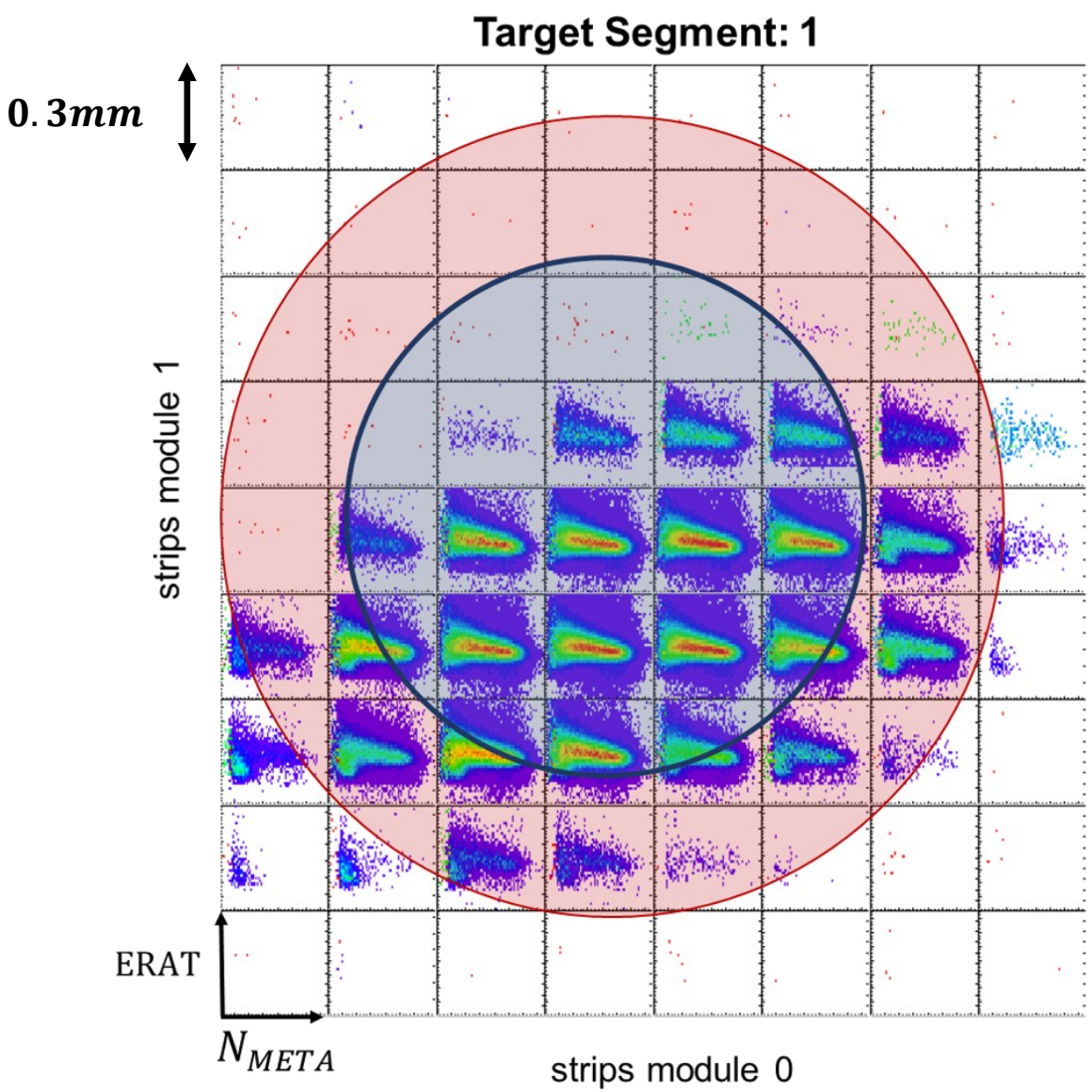
Possibilites for Ag+C collisions:

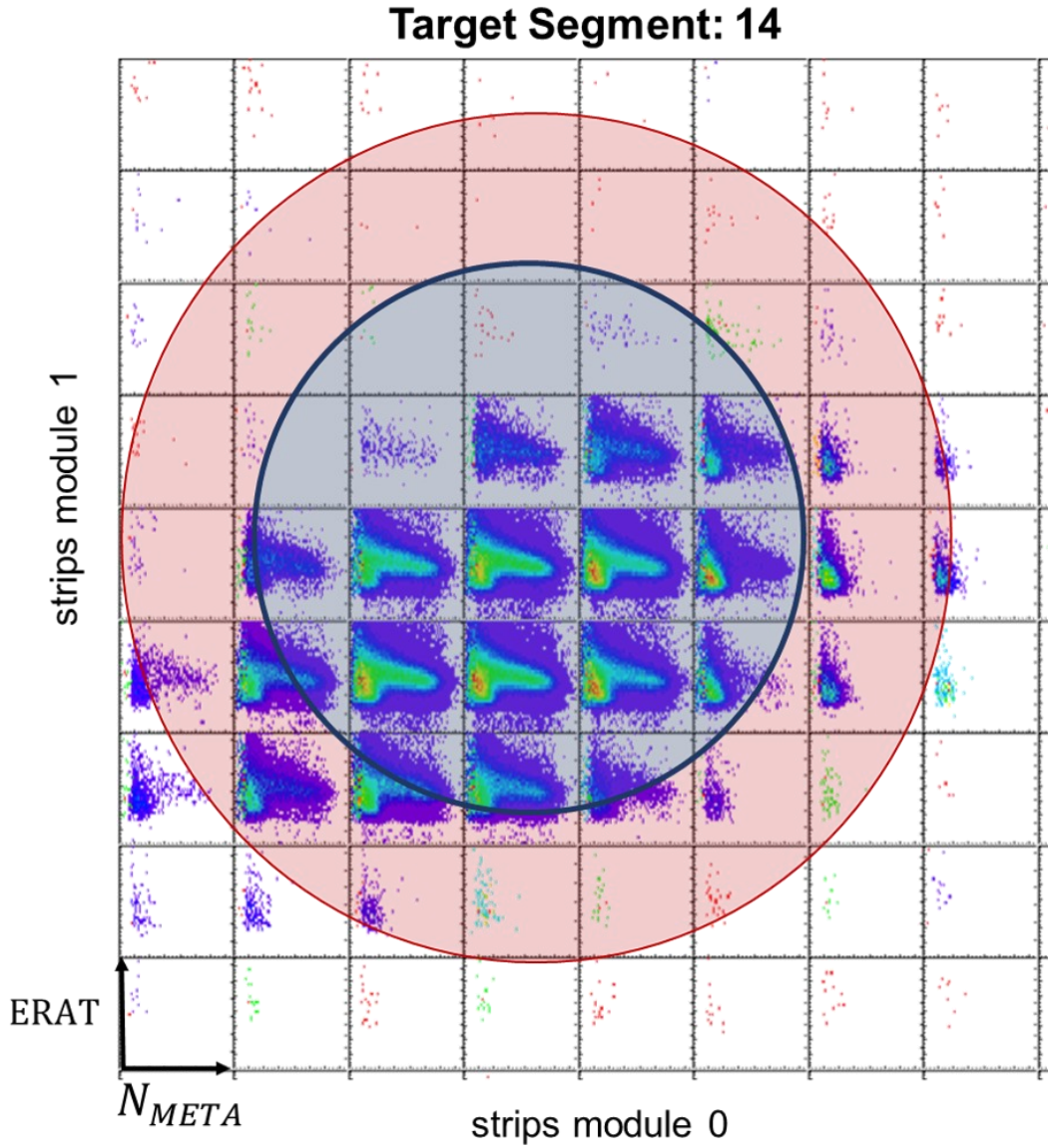
- Inclined beam
- Misaligned target segments
- Beam divergence



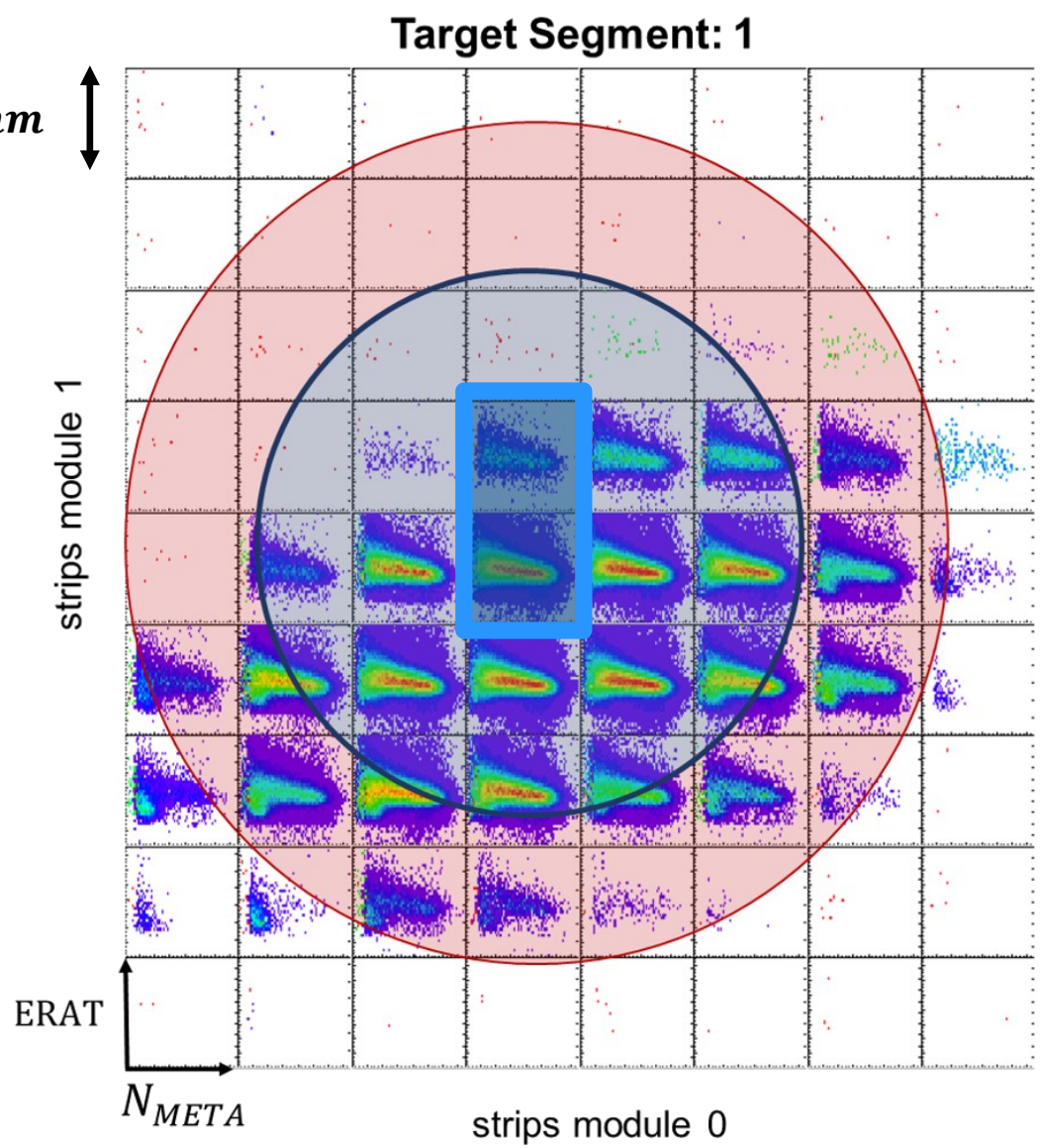
https://doi.org/10.1016/j.nima.2011.06.029

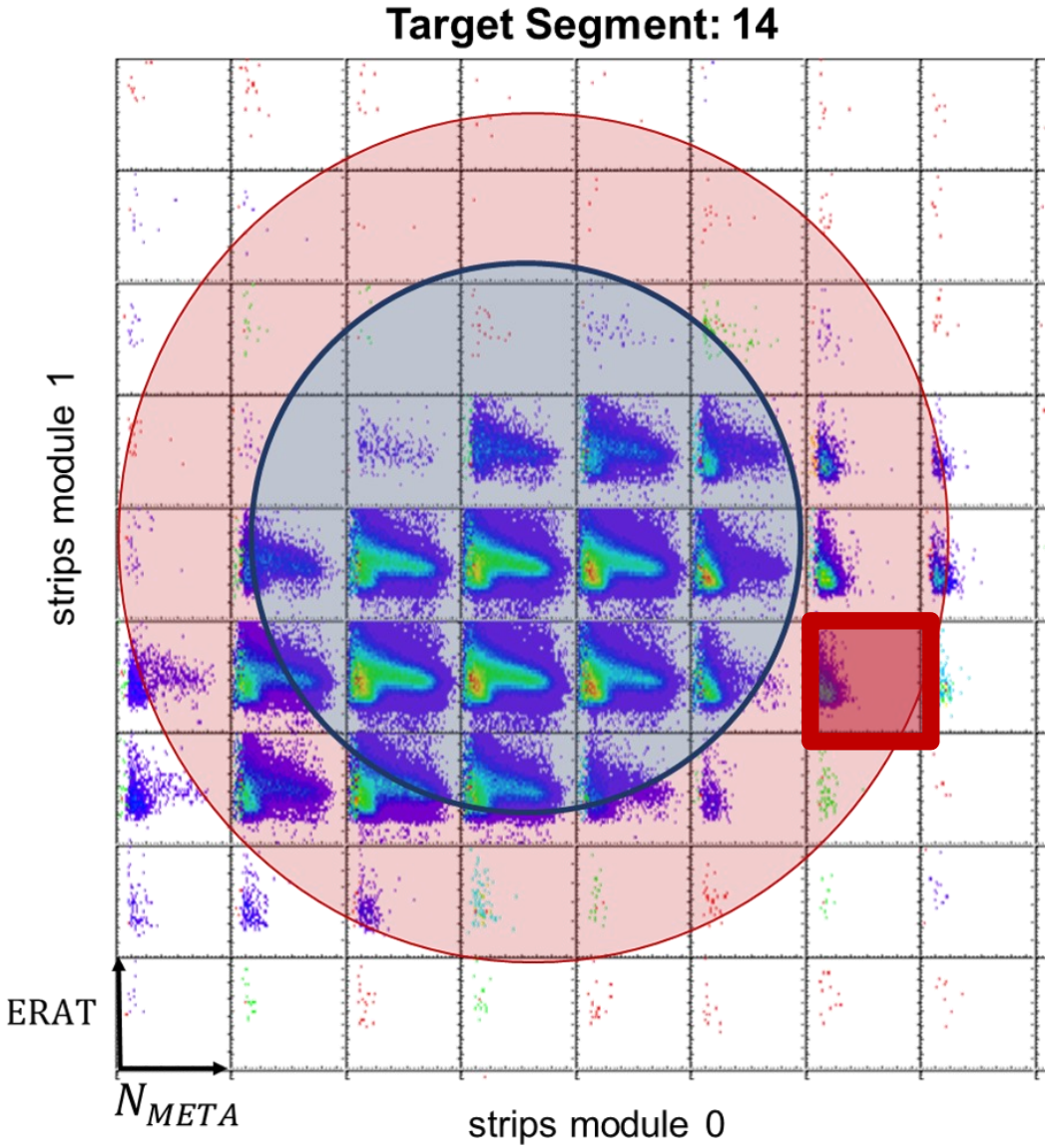
1. SELECTION OF TRAINING DATA





1. SELECTION OF TRAINING DATA





1. SELECTION OF TRAINING DATA

- **Software version:** /cvmfs/hades.gsi.de/install/5.34.34/hydra2-5.6/defall.sh
- **Generation 5 of march 19:** /lustre/hades/dst/mar19/gen5/
- **Days 088-089 in Ag123Ag**
- **Days 078-080 in Ag158Ag**

} Stable beam position on START !

- **Min. bias event selection:**
  - kGoodVertexClust
  - kGoodSTART

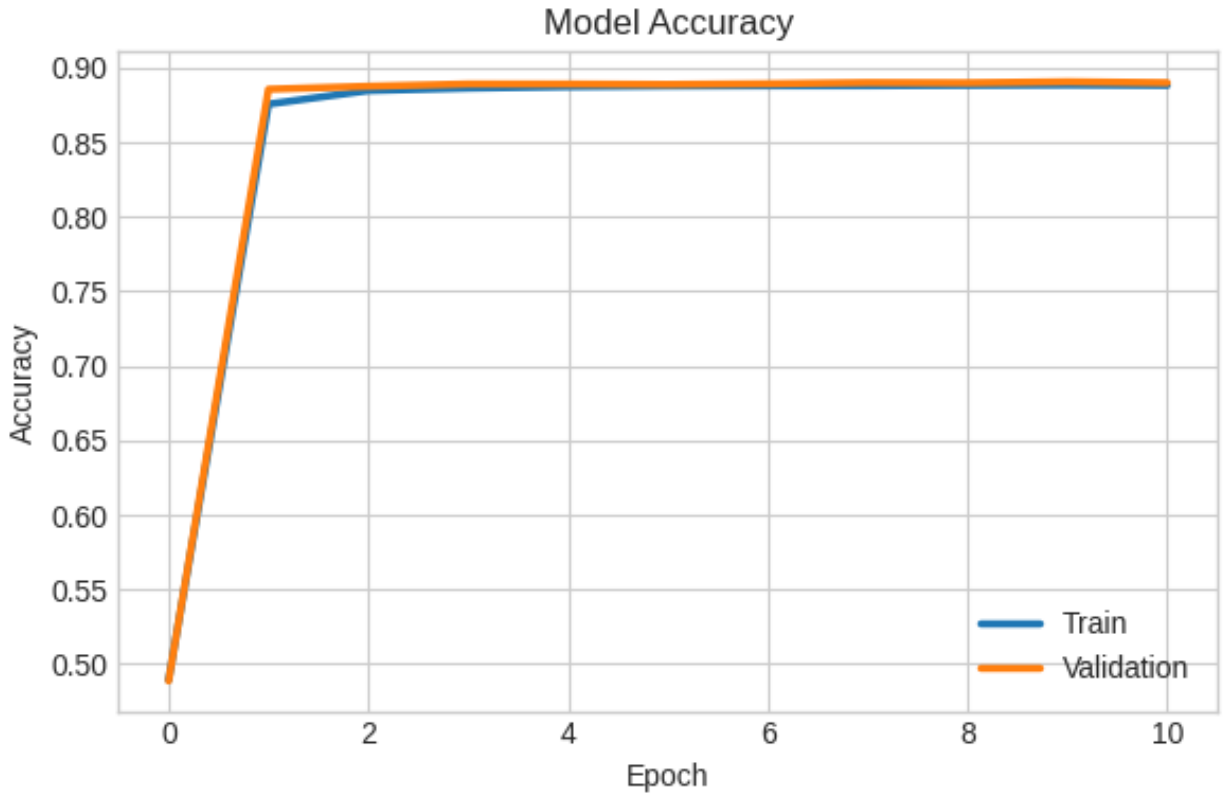
- **START – Target Matching:**
- **Additional cut for Ag+C:**

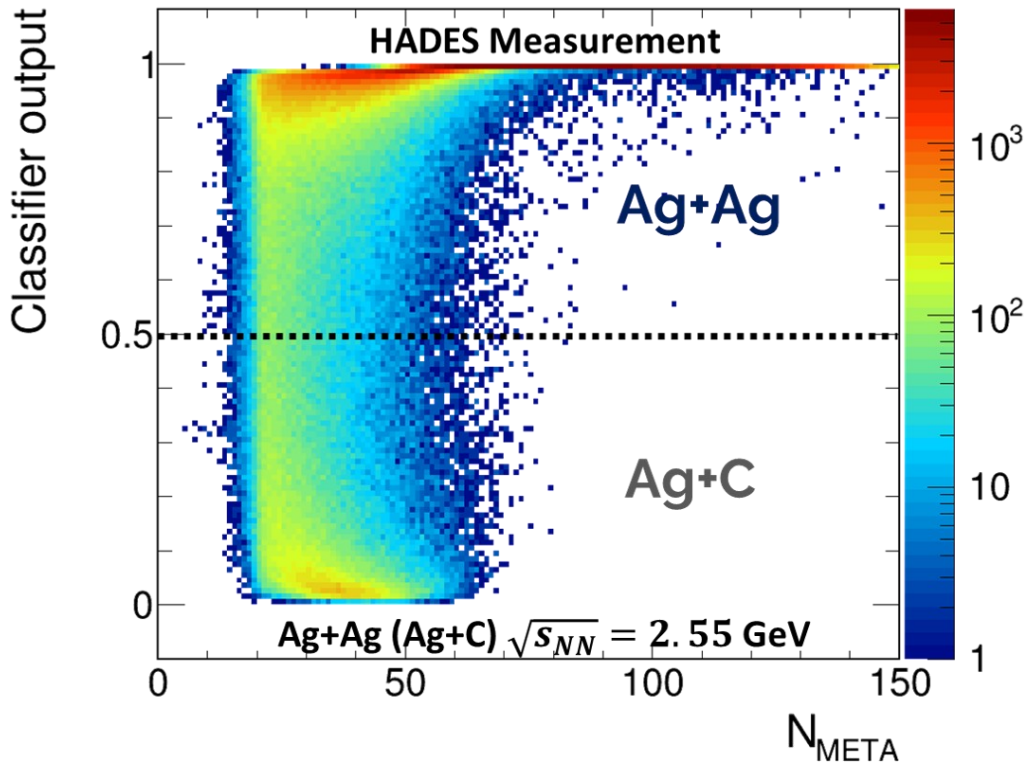
$$N_{META} < 60$$

Energy	System	Target Segment t	START hit (module 0, module 1)
1.23A GeV	Ag+C	1, 2, 11, 13, 14	(7, 4), (7, 3), (7, 2), (6, 1)
		3	(7, 3), (7, 2)
		4, 5, 7, 8, 9, 10, 12	(7, 4), (7, 3), (7, 2)
		6	(7, 2)
	Ag+Ag	0, 1	(3, 4), (3, 3)
1.58A GeV	Ag+C	2, 4, 6, 7, 8, 9, 10, 12	(0, 4), (1, 4), (7, 7), (7, 6), (6, 5)
		1, 3, 5, 11	(0, 4), (1, 4), (7, 7), (7, 6), (6, 5), (0, 5)
		13, 14	(0, 4), (1, 4), (7, 7), (7, 6), (6, 5), (6, 7), (6, 6), (5, 5)
	Ag+Ag	0, 1	(3, 8), (3, 7)

2. MODEL TRAINING

- Neural network successfully trained (~89% accuracy on training data)
- Model output between 0 and 1, with standard threshold at $p = 0.5$





3. MODEL EVALUATION - SIMULATION

- Full event simulation with transport models: **UrQMD**, **DCM-QGSM**
- Here: Even mix of 30 000 Ag+C and 30 000 Ag+Ag accepted events

→ Simulations suggest slight favour towards pure Ag+Ag… but overall:

Succesful separation between Ag+C and Ag+Ag

Accuracy
What overall proportion of predictions was correct?

Recall
What proportion of Ag+C was identified correctly?

Precision
What proportion of predicted Ag+C was really Ag+C?

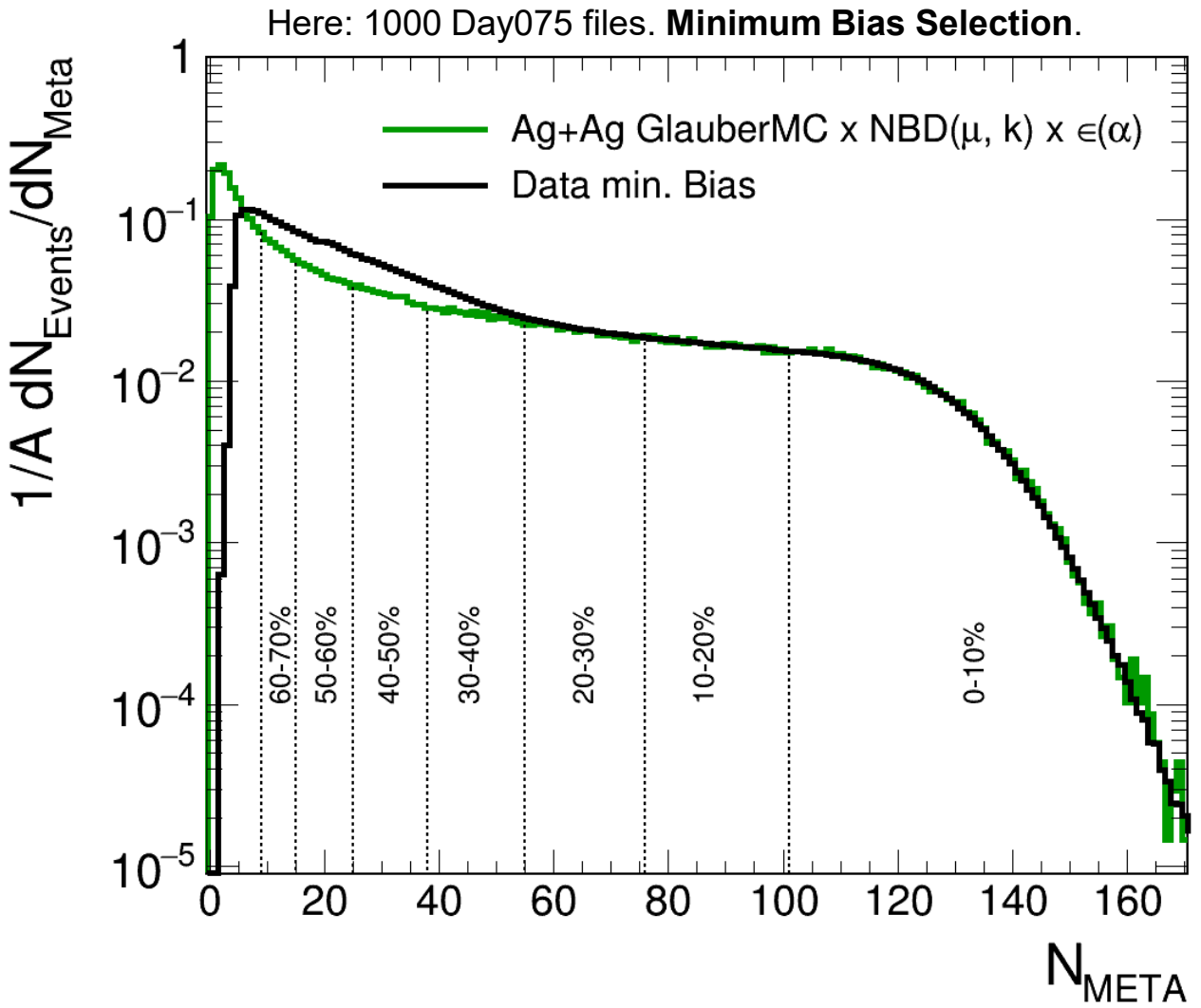
Centrality	Model	Accuracy (%)	Recall (%)	Precision (%)
30–40%	UrQMD	90.6	95.1	83.5
	DCM-QGSM	89.3	96.9	81.8
40–50%	UrQMD	87.2	96.1	85.5
	DCM-QGSM	82.7	97.4	80.3

3. MODEL EVALUATION - SIMULATION

- Compare minimum bias TOF+RPC distribution with Glauber Model*

μ	k	α
0.81	41.9	4.1 E-6

$$A = \int_{70}^{160} \frac{dN}{dN_{Meta}} \, dN_{Meta}$$



*https://github.com/bkardano/TGlauberMC

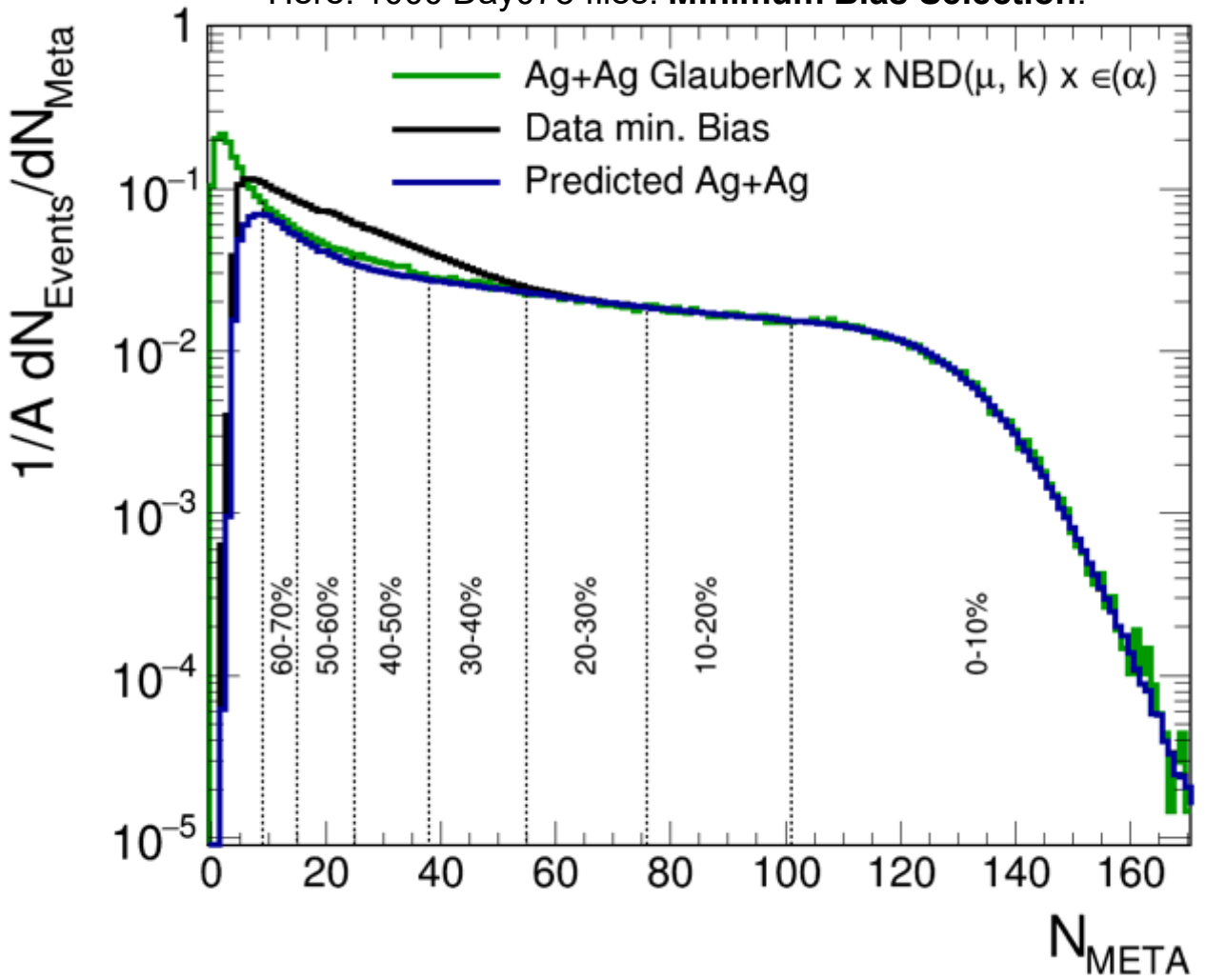
3. MODEL EVALUATION - SIMULATION

- Compare minimum bias TOF+RPC distribution with Glauber Model*

μ	k	α
0.81	41.9	4.1 E-6

$$A = \int_{70}^{160} \frac{dN}{dN_{Meta}} \, dN_{Meta}$$

Here: 1000 Day075 files. **Minimum Bias Selection**.



➡ Predicted Ag+Ag align significantly better with GlauberMC

*https://github.com/bkardano/TGlauberMC

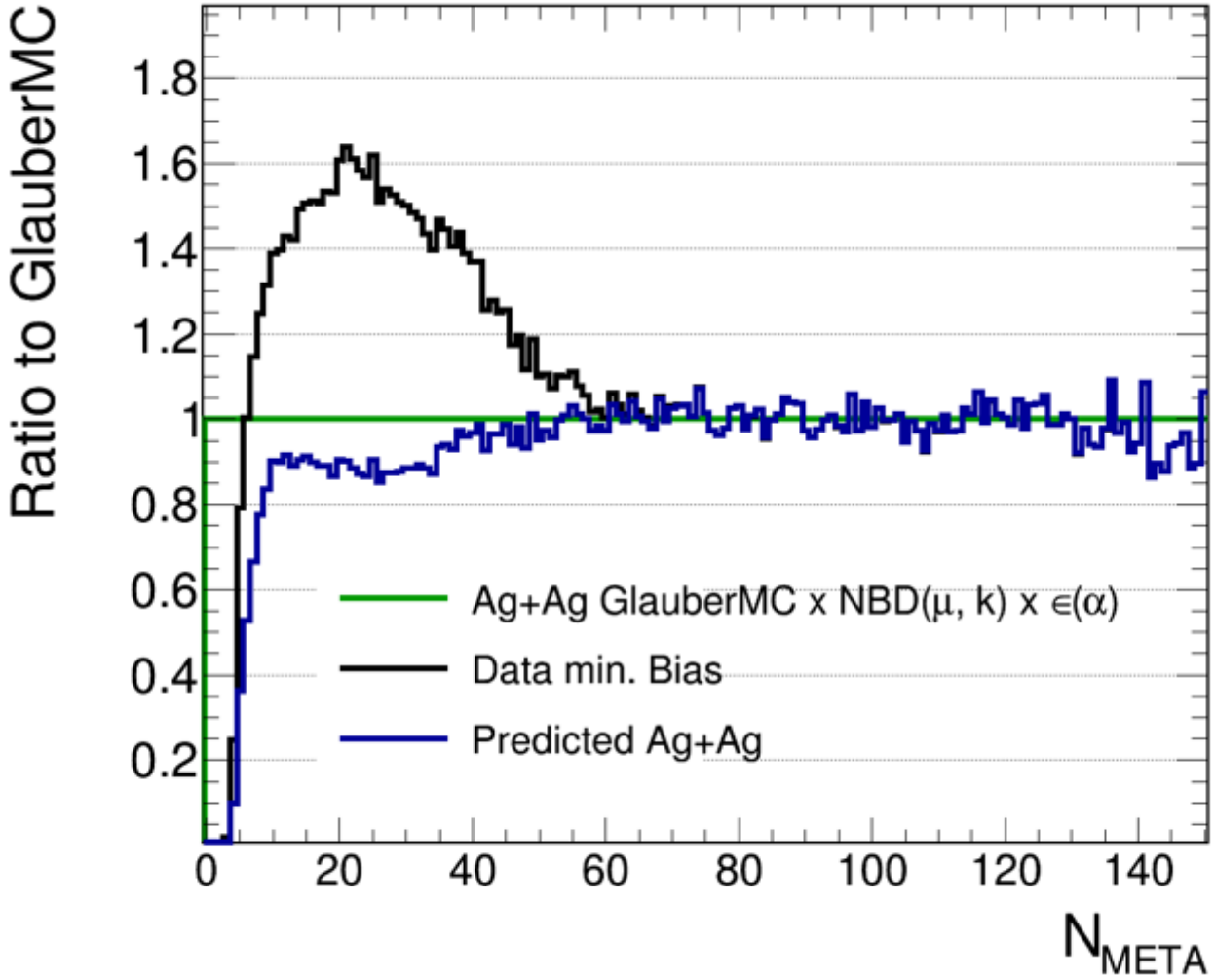
3. MODEL EVALUATION - SIMULATION

- Compare minimum bias TOF+RPC distribution with Glauber Model*

μ	k	α
0.81	41.9	4.1 E-6

$$A = \int_{70}^{160} \frac{dN}{dN_{Meta}} \, dN_{Meta}$$

Here: 1000 Day075 files. **Minimum Bias Selection**.



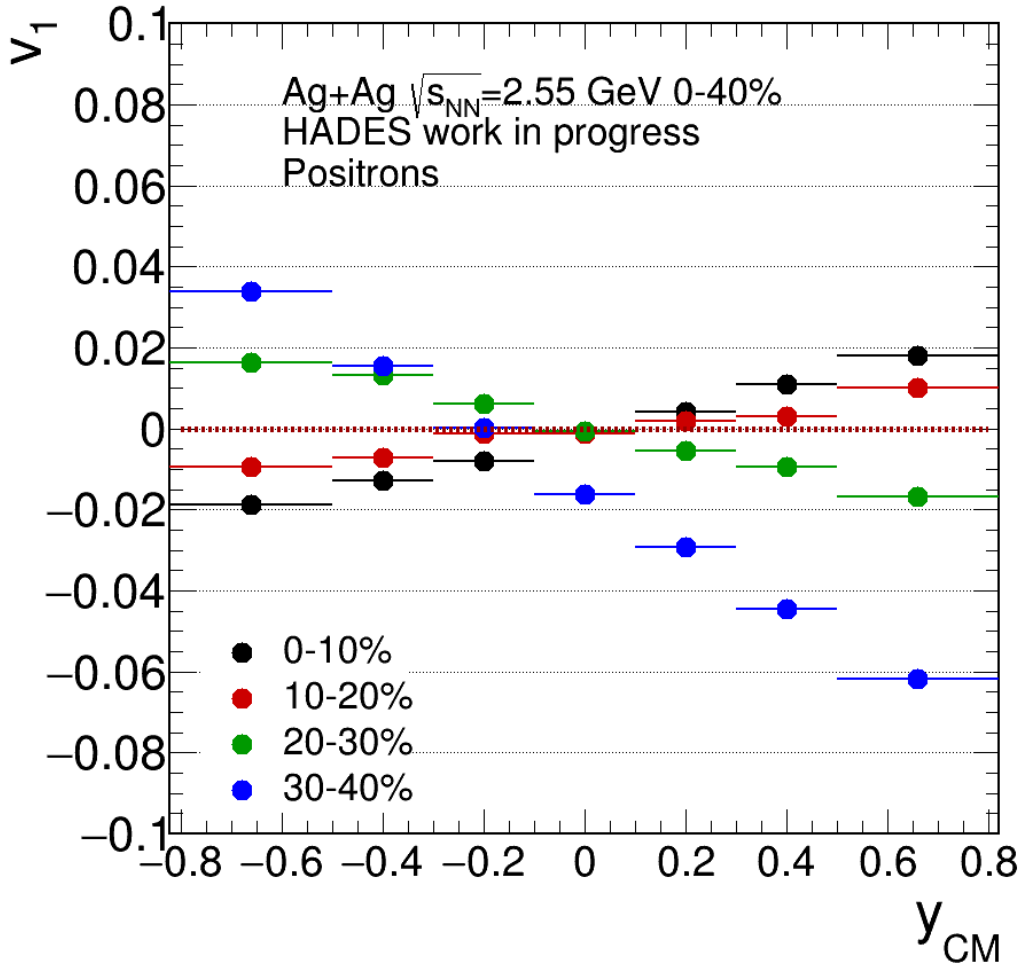
➡ Predicted Ag+Ag align significantly better with GlauberMC

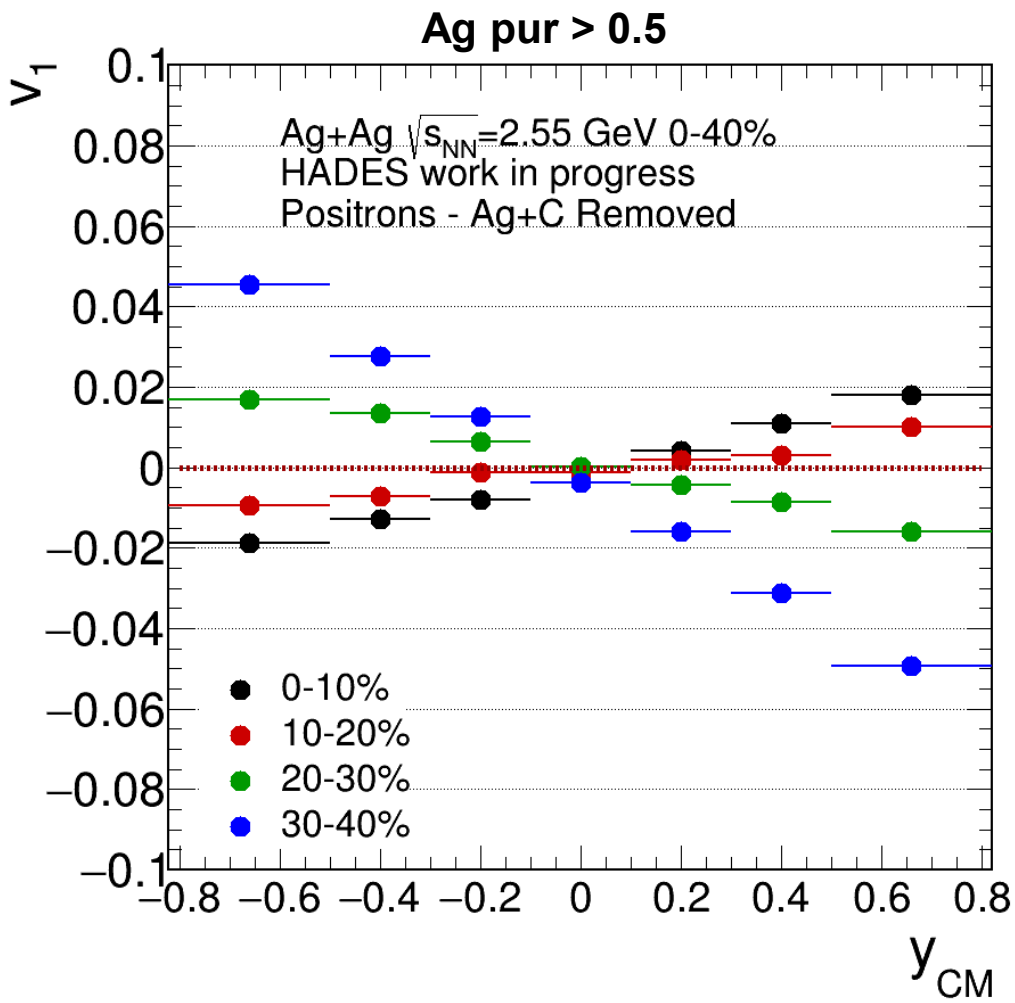
➡ Model appears to show similiar tendency to favour Ag+Ag purity

*https://github.com/bkardano/TGlauberMC

3. MODEL EVALUATION - DATA

Positrons



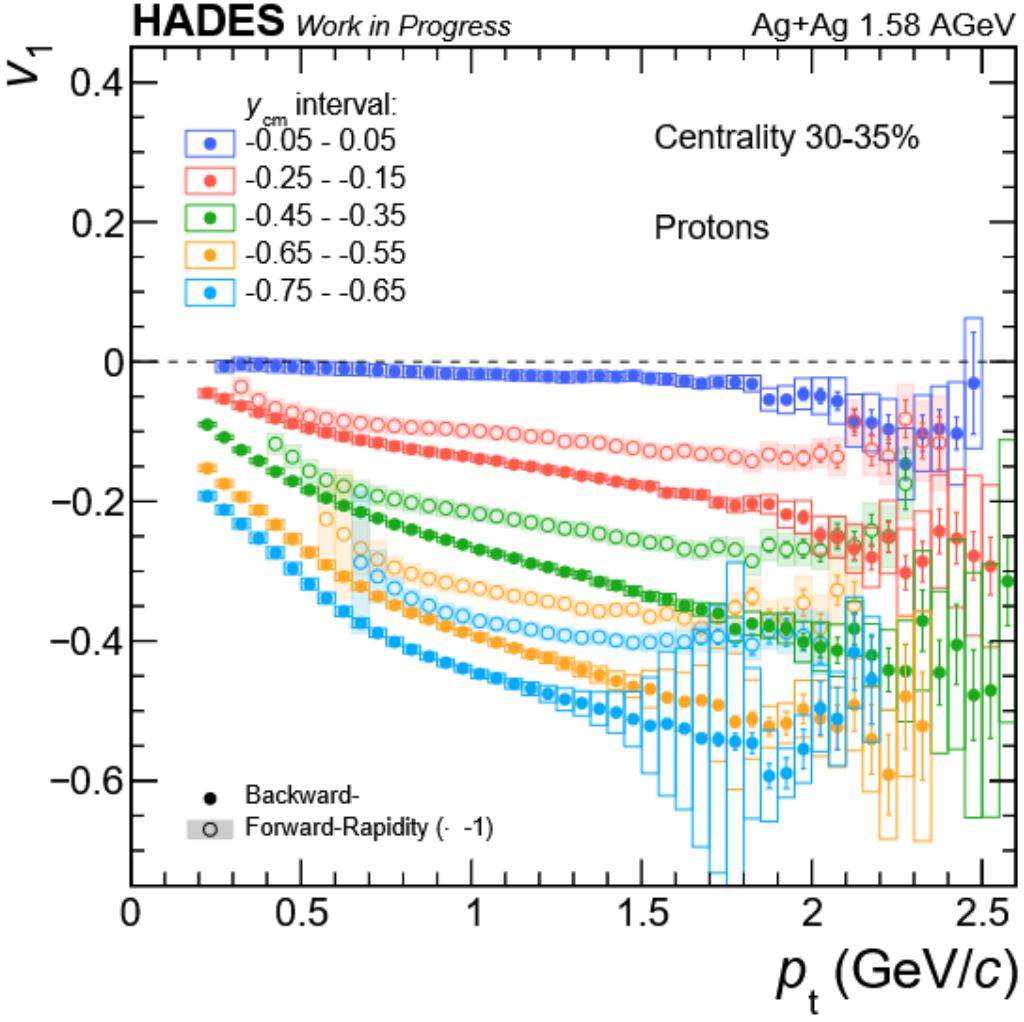


3. MODEL EVALUATION - DATA

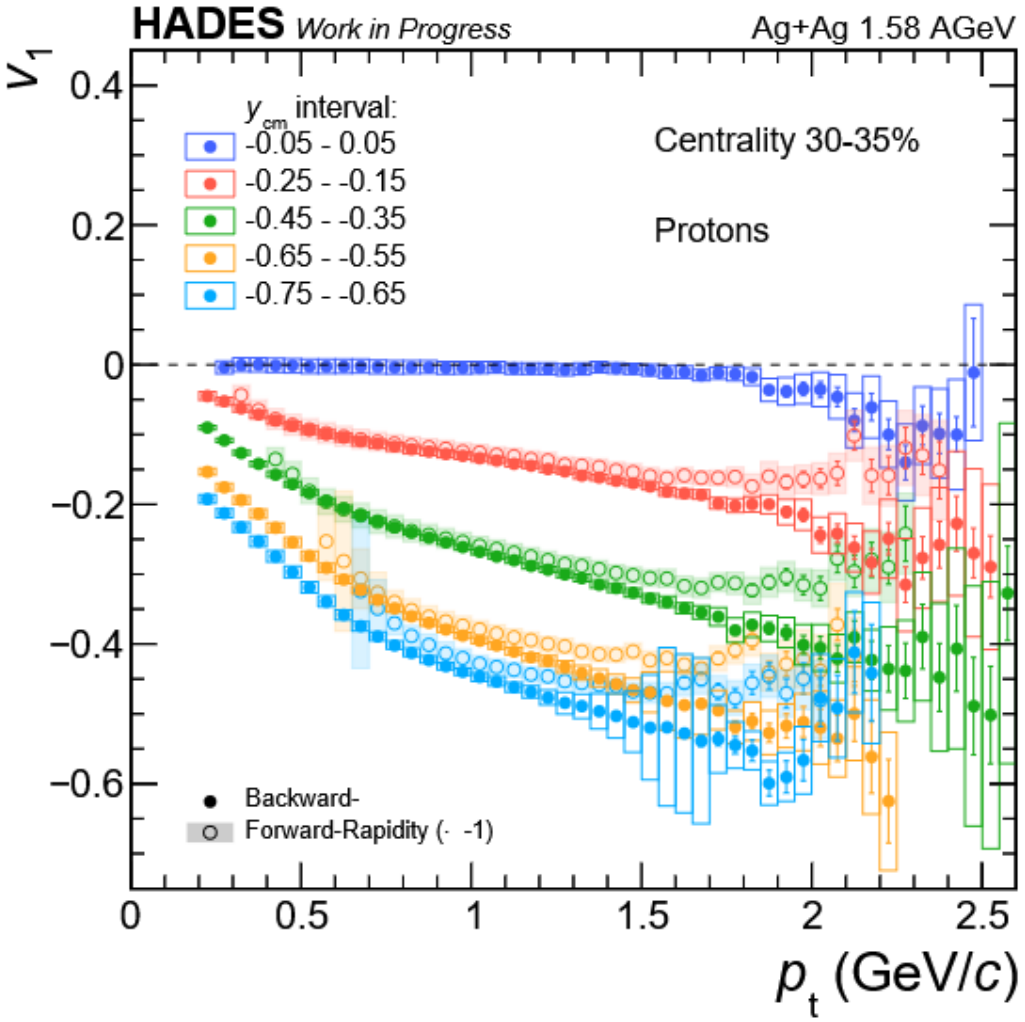
p

HADES

TECHNISCHE UNIVERSITÄT DARMSTADT

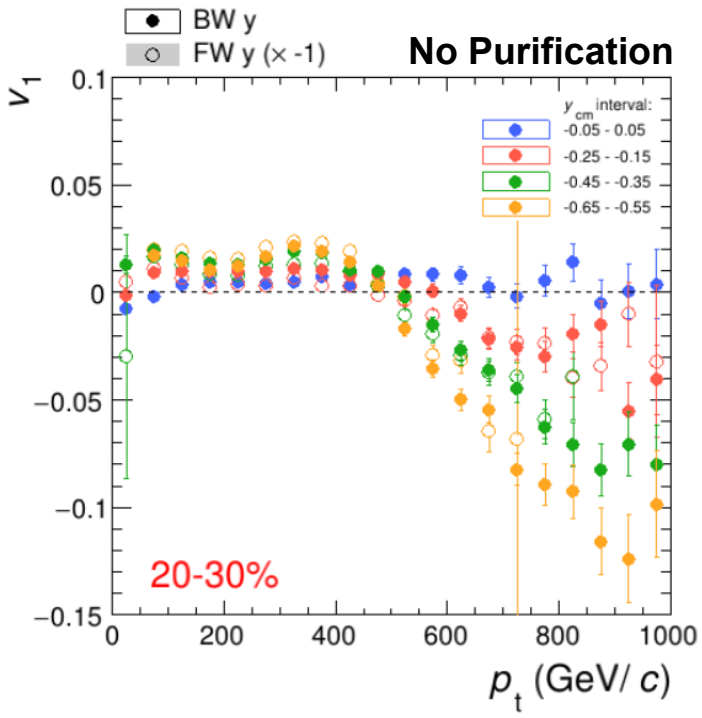


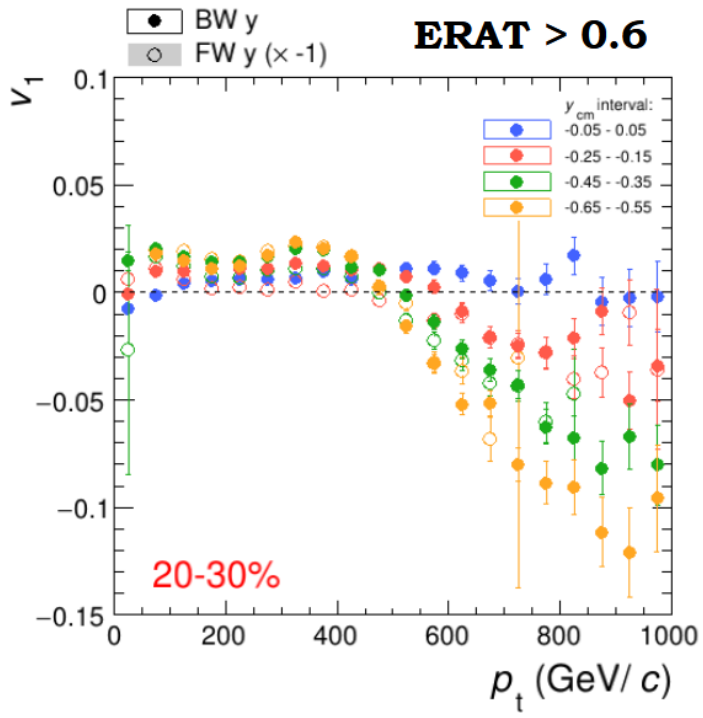
Ag pur > 0.5

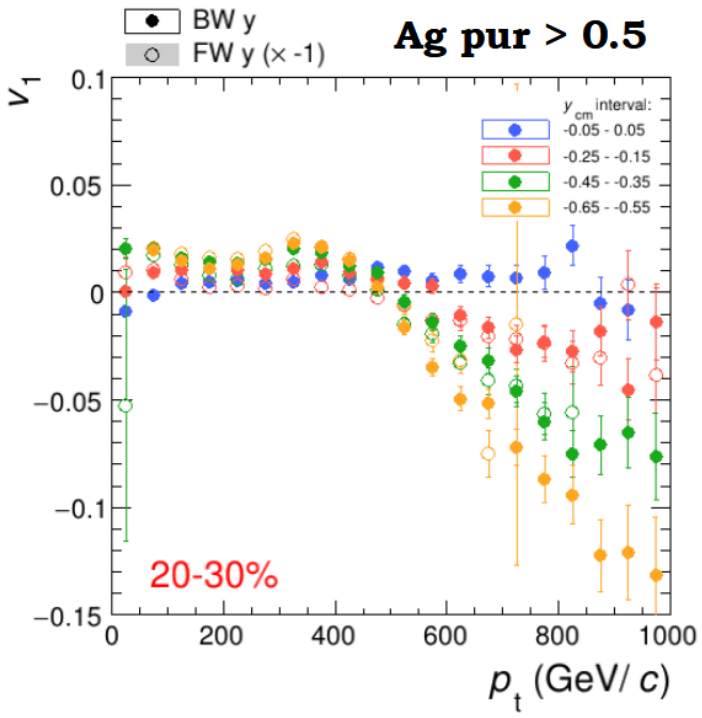


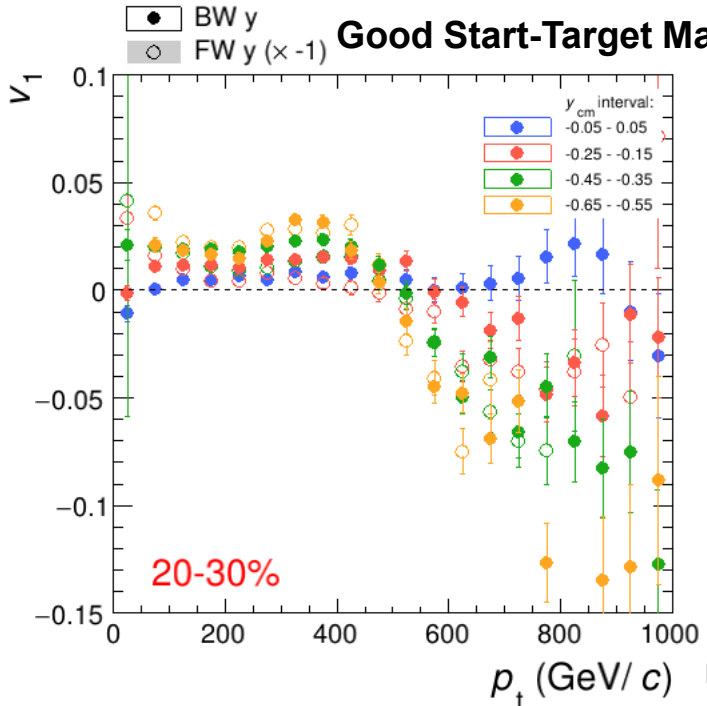
3. MODEL EVALUATION - DATA

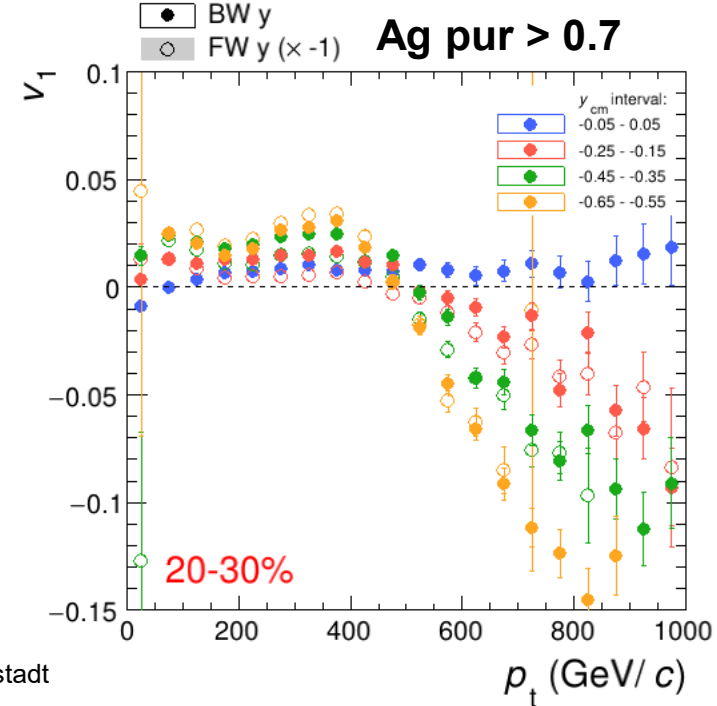
π^-





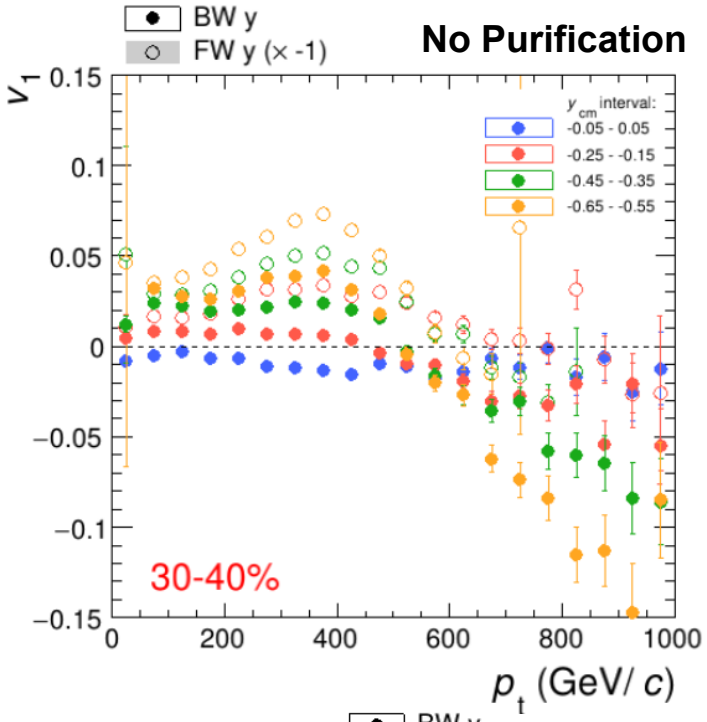


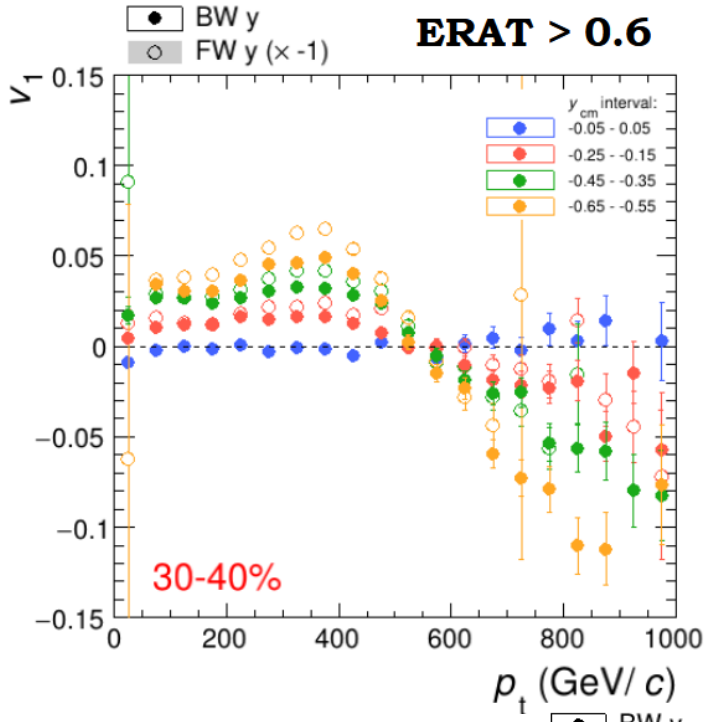


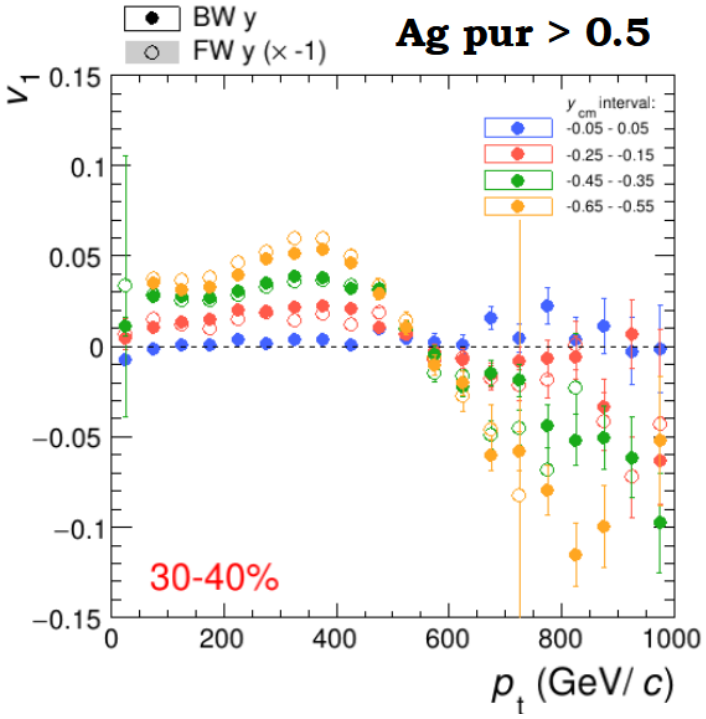


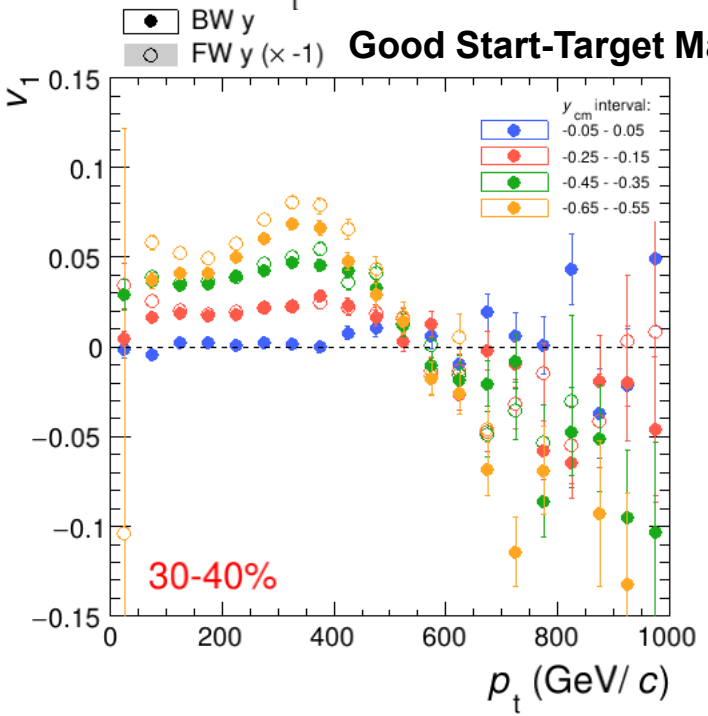
3. MODEL EVALUATION - DATA

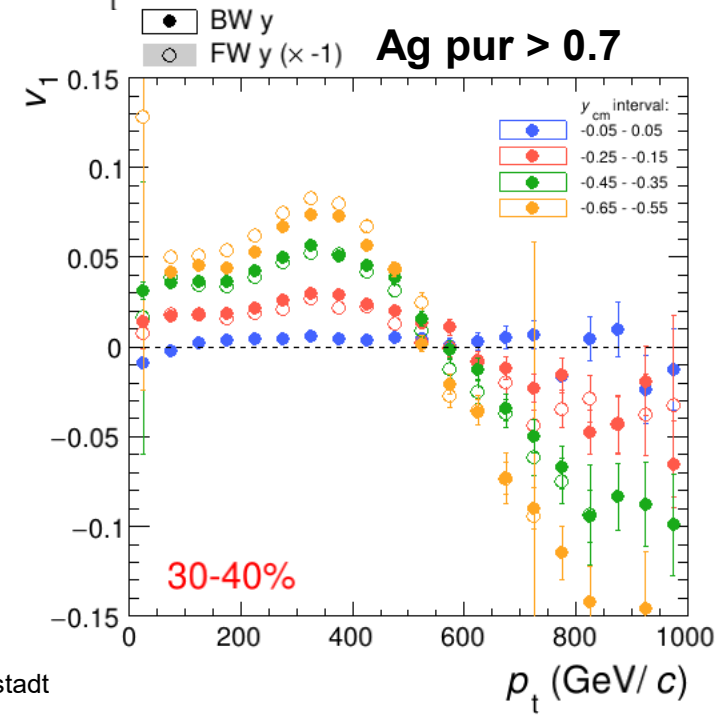
π^-





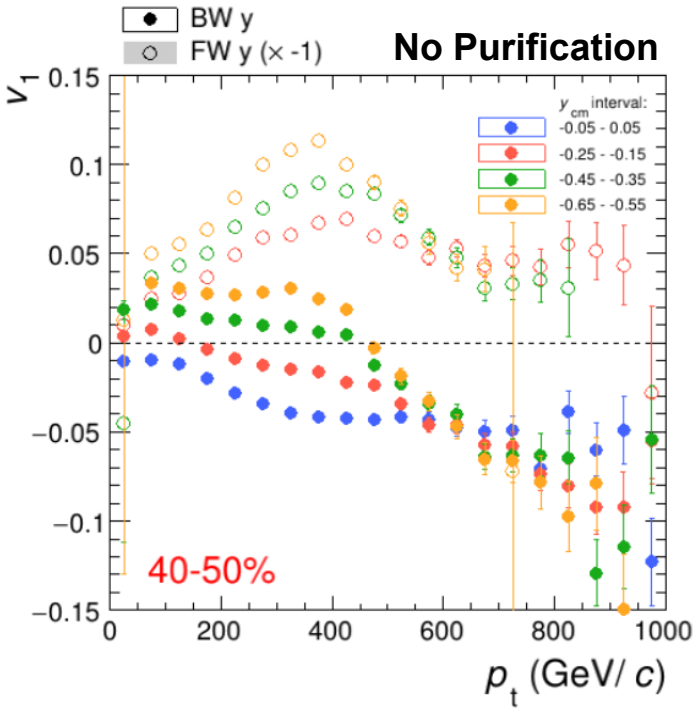


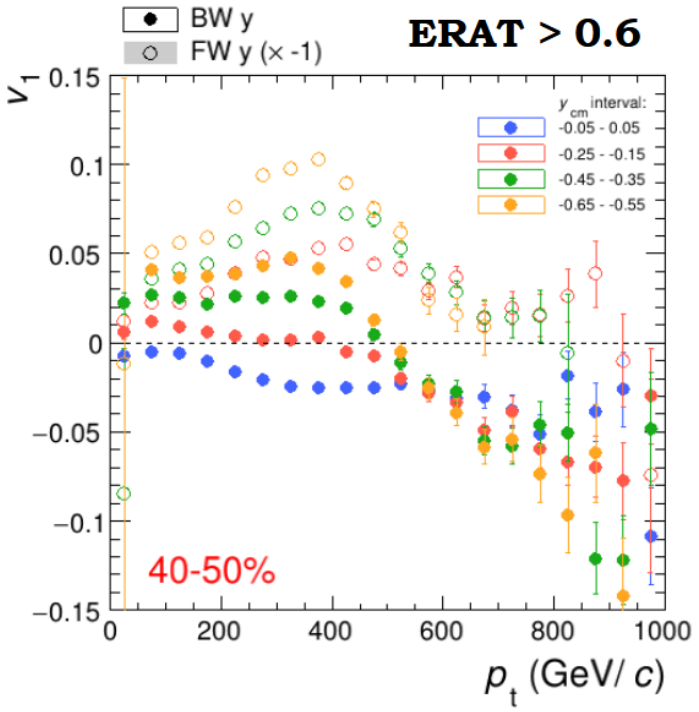


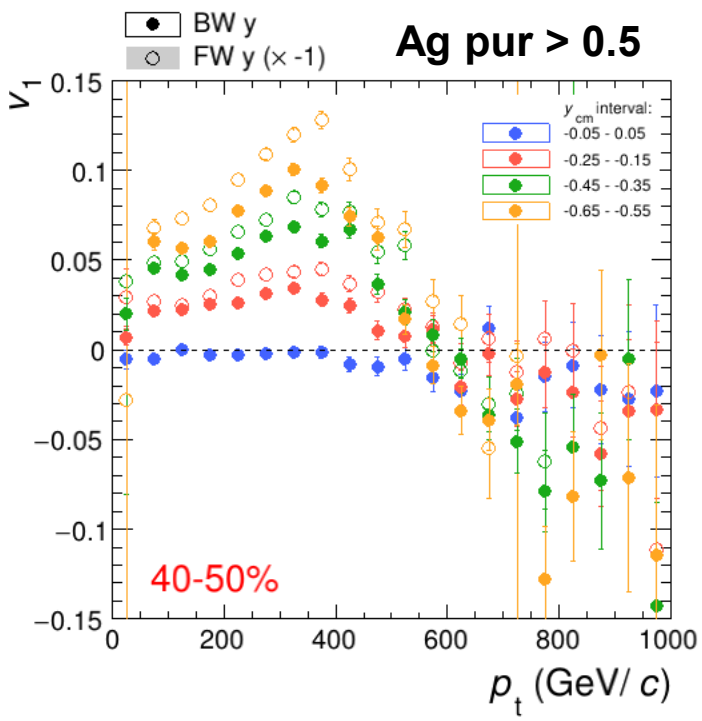


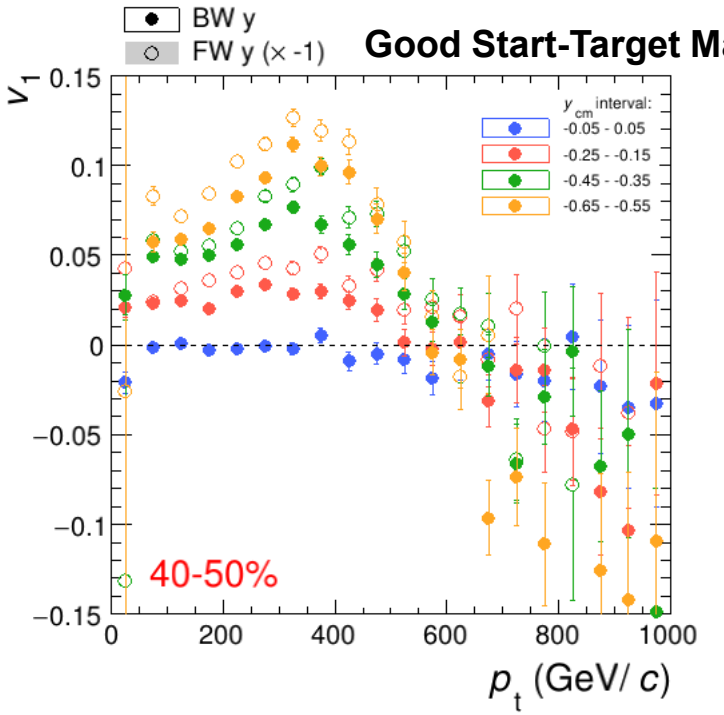
3. MODEL EVALUATION - DATA

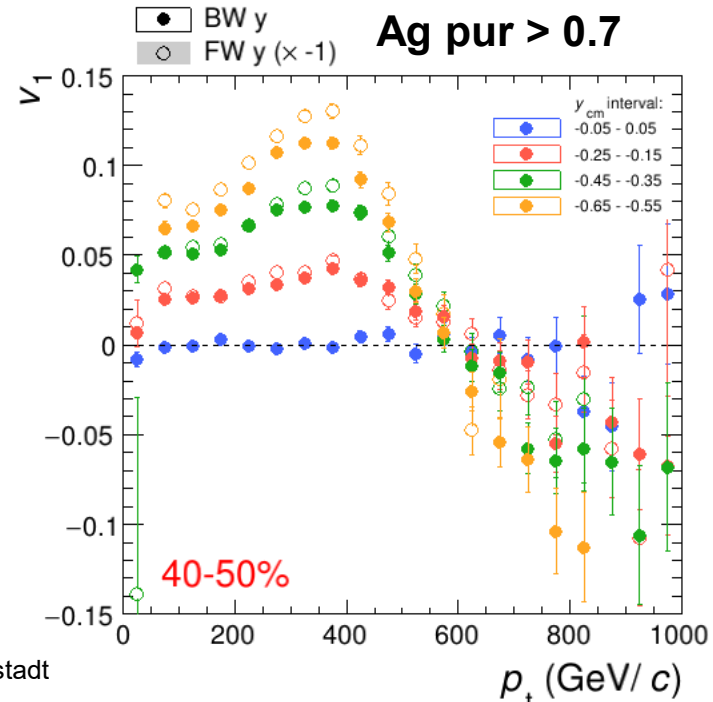
π^-











CONCLUSION

Model brings significant improvement in semi-central centralities

- While not perfect, it can allow extension of analyses to at least 40% centrality

  ➡ Employ in dilepton analysis to maximize statistics

  ➡ Estimate on systematics can be made by varying decision threshhold p

CONCLUSION

Model brings significant improvement in semi-central centralities

- While not perfect, it can allow extension of analyses to at least 40% centrality
  - ➡ Employ in dilepton analysis to maximize statistics
  - ➡ Estimate on systematics can be made by varying decision threshhold p

HOW TO USE

- Code and analysis details are in dedicated analysis note
- **Test it yourself**:

/u/nschild/Niklas/Publication/CarbonRemoval/MacroPort/

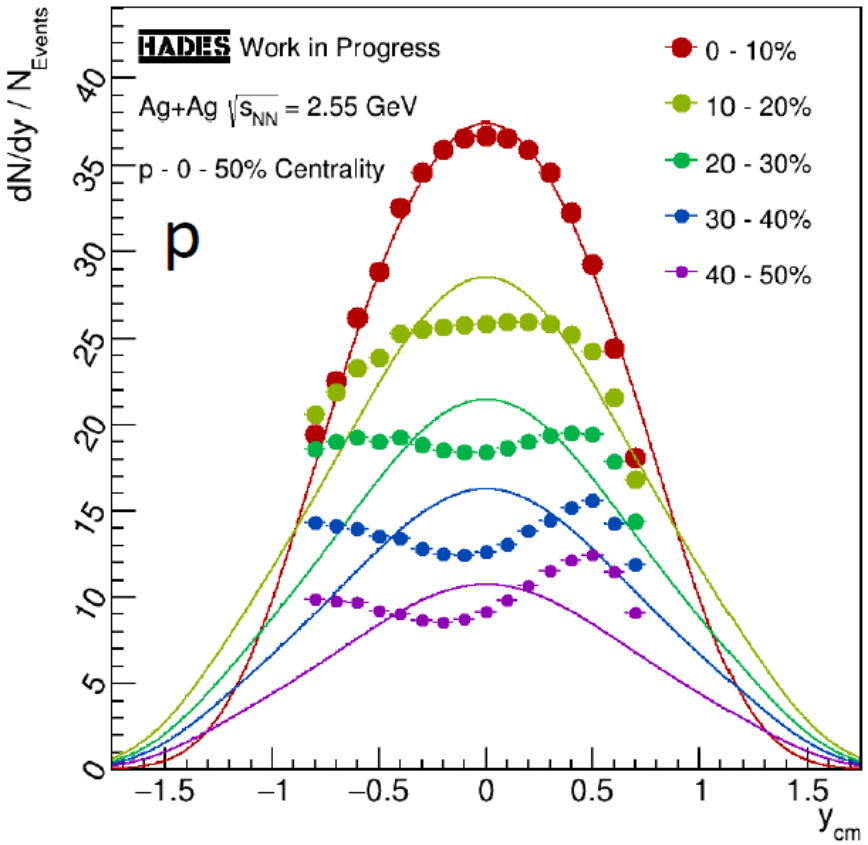
https://git.gsi.de/n.schild/agcremover/-/tree/main/MacroPort

Model predictions exist for all mar19 accepted files

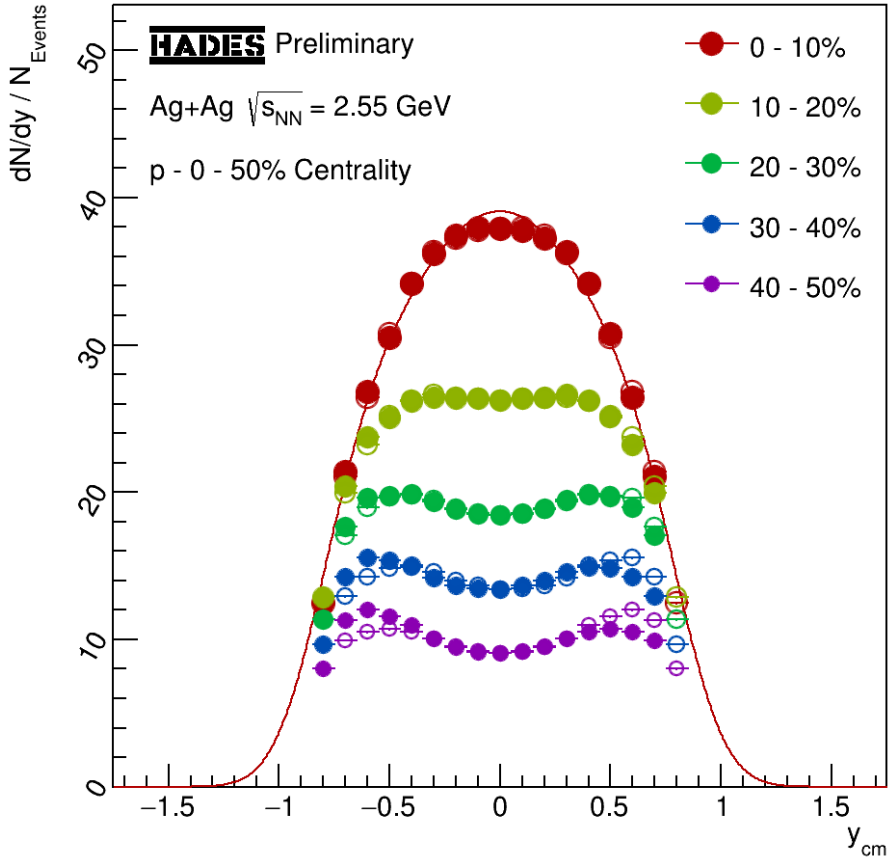
You can also manually select START-Target correlations to get small „pure“ sample

BACKUP

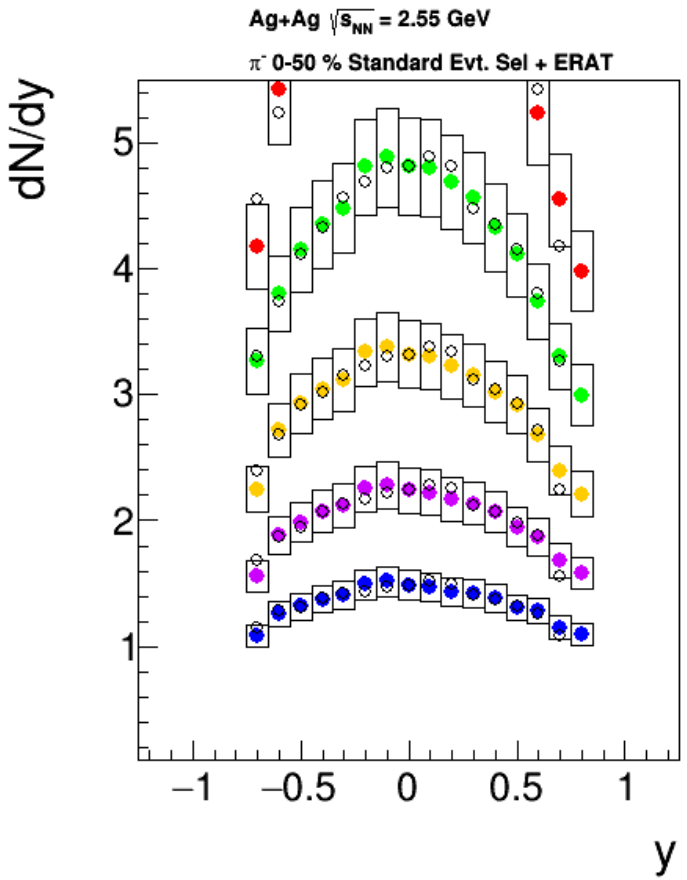
MODEL EVALUATION - DATA

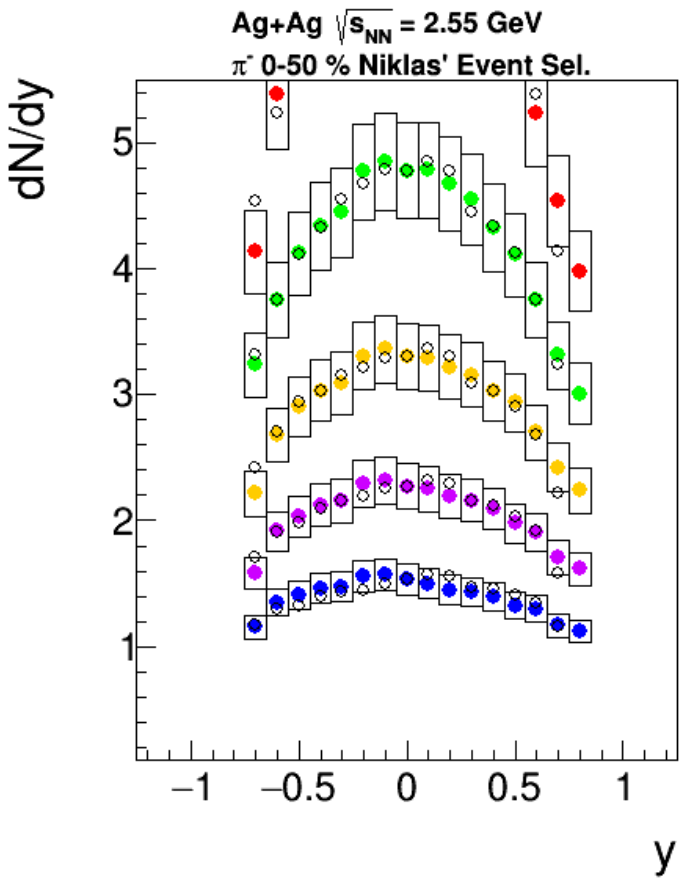


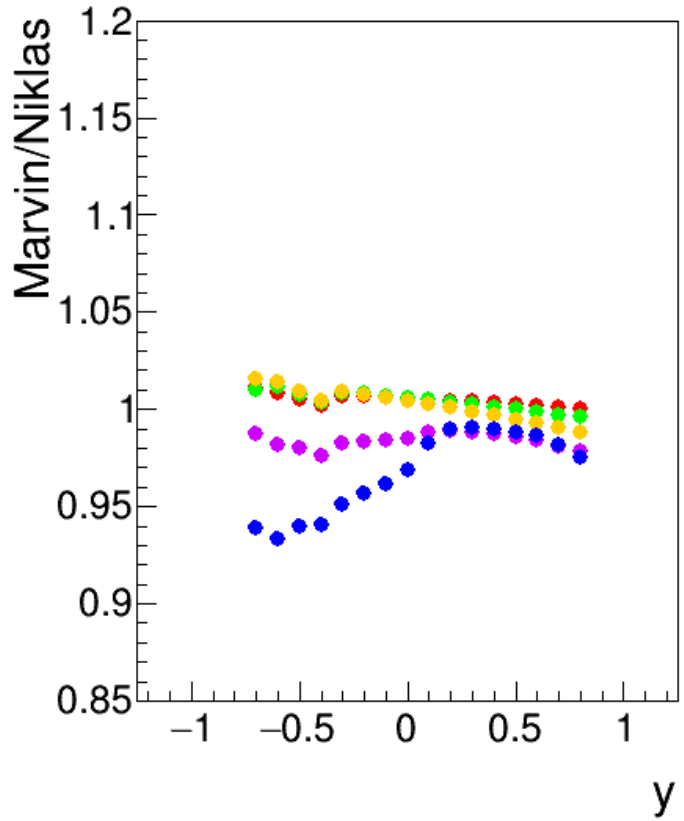
Ag pur > 0.5



MODEL EVALUATION - DATA



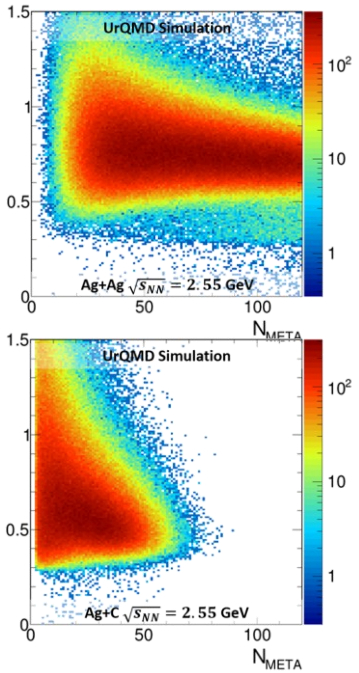




1. SELECTION OF TRAINING DATA

Energy	Ag+C		Ag+Ag	
	Accepted	Rejected	Accepted	Rejected
1.23A GeV	148 424	46 115	450 000	105 000
1.58A GeV	293 302	206 530	700 000	378 299

- Balanced selection between **Ag+C vs Ag+Ag** as well as **accepted vs rejected** events



Need some more Ag+Ag events due to different multiplicity distribution

Model main focus on accepted events

But include some rejected events for more general applicability

2. MODEL ARCHITECTURE

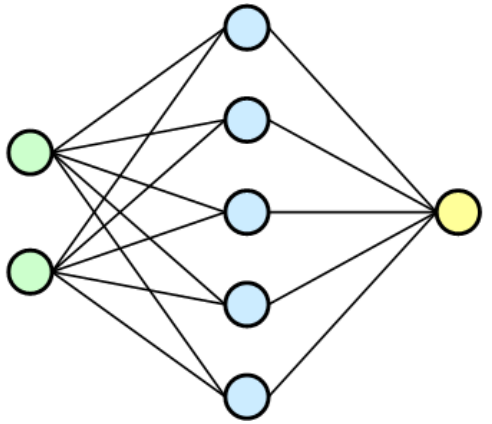
Input Feautres

- N_{TOF}: Number of hits in the TOF detector.
- N_{RPC}: Number of hits in the RPC detector.
- N_{Tracks}: Number of reconstructed tracks after the standard track selection.
- **Average Particle Rapidity**: Mean rapidity of all reconstructed particles within the event.
- **Average Transverse Momentum**: Mean transverse momentum of all reconstructed particles.
- **Average Rapidity**: Event-wise average rapidity of all detected particles.
- **Total Mass**: Sum of the reconstructed particle masses.
- **ERAT**: Ratio of total transverse to longitudinal energy, as defined in equation 1.
- **Sector Correction ID**: Integer identifier encoding the number of active sectors contributing to the event.
- **IsRejected**: Boolean flag indicating whether the event failed any of the standard event selection criteria.

Fully-connected feed-forward neural network

- 3 Hidden layers: 148, 148, 138 neurons
- Activation function: tanh()
- Optimiser: Adam
- Loss function: Categorial cross-entropy

- Model training and predictions made in Python (version 3.6.1) with Keras library
  → output saved in ROOT TTree



2. MODEL ARCHITECTURE

Input Feautres

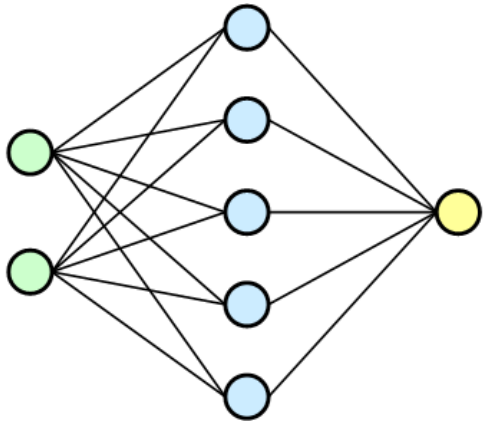
- N_{TOF}: Number of hits in the TOF detector.
- N_{RPC}: Number of hits in the RPC detector.
- N_{Tracks}: Number of reconstructed tracks after the standard track selection.
- **Average Particle Rapidity**: Mean rapidity of all reconstructed particles within the event.
- **Average Transverse Momentum**: Mean transverse momentum of all reconstructed particles.
- **Average Rapidity**: Event-wise average rapidity of all detected particles.
- **Total Mass**: Sum of the reconstructed particle masses.
- **ERAT**: Ratio of total transverse to longitudinal energy, as defined in equation 1.
- **Sector Correction ID**: Integer identifier encoding the number of active sectors contributing to the event.
- **IsRejected**: Boolean flag indicating whether the event failed any of the standard event selection criteria.

Essential as they may impact the other event characteristics

Fully-connected feed-forward neural network

- 3 Hidden layers: 148, 148, 138 neurons
- Activation function: tanh()
- Optimiser: Adam
- Loss function: Categorial cross-entropy

- Model training and predictions made in Python (version 3.6.1) with Keras library
  → output saved in ROOT TTree

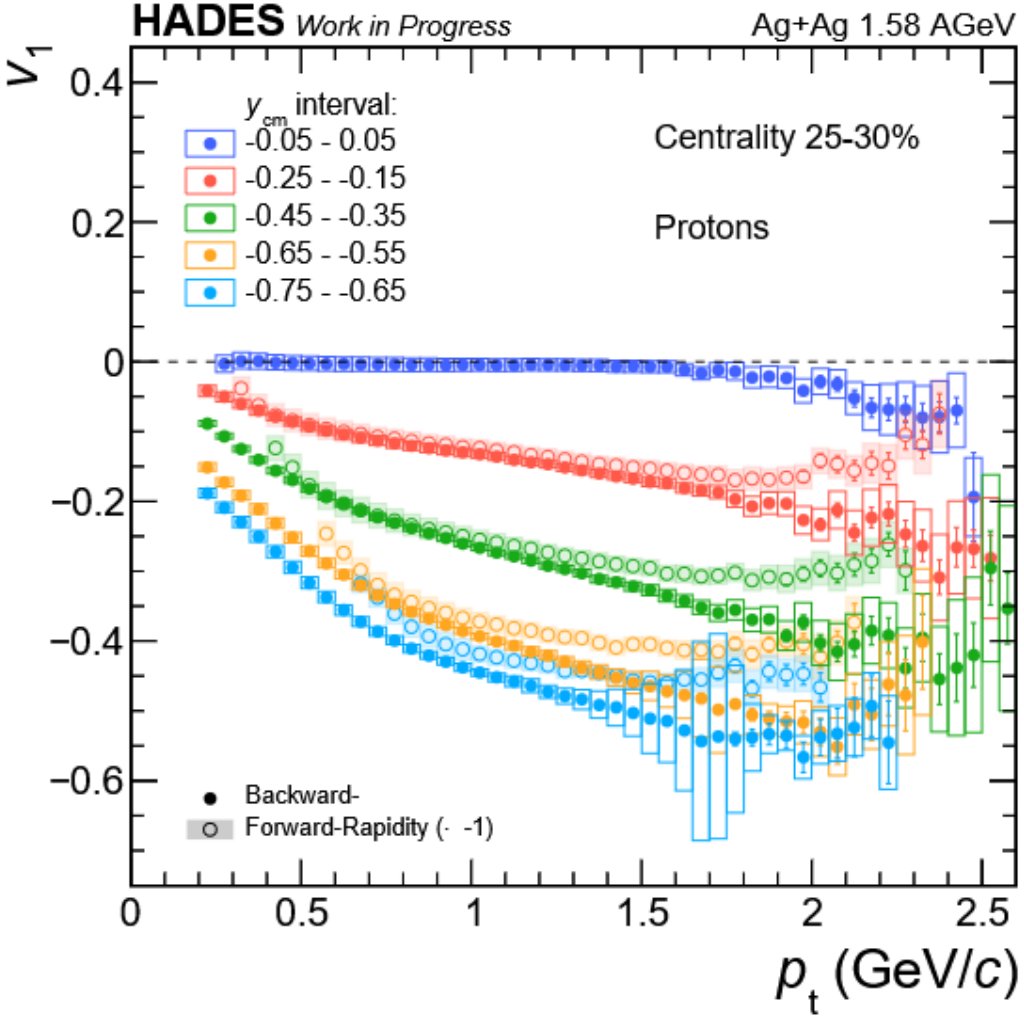


3. MODEL EVALUATION - DATA

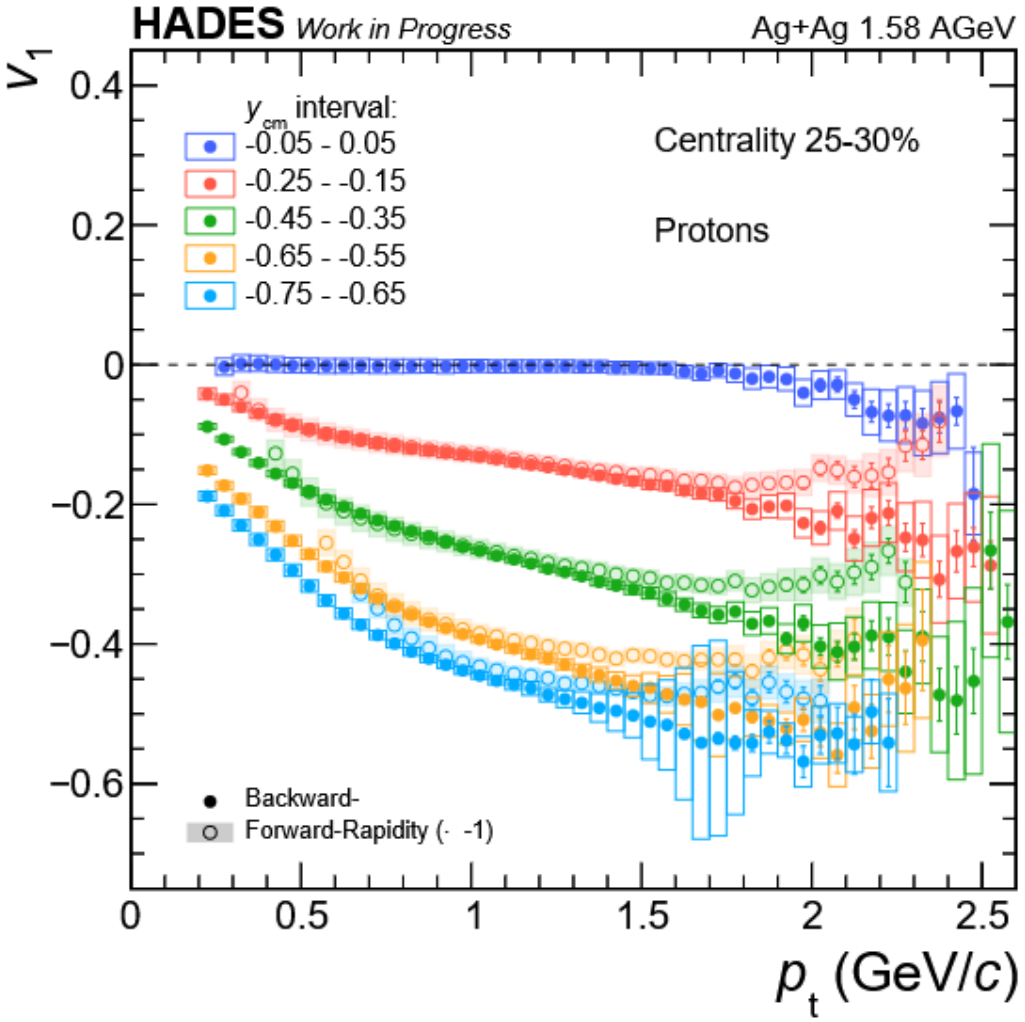
p

HADES

TECHNISCHE UNIVERSITÄT DARMSTADT



Ag pur > 0.5

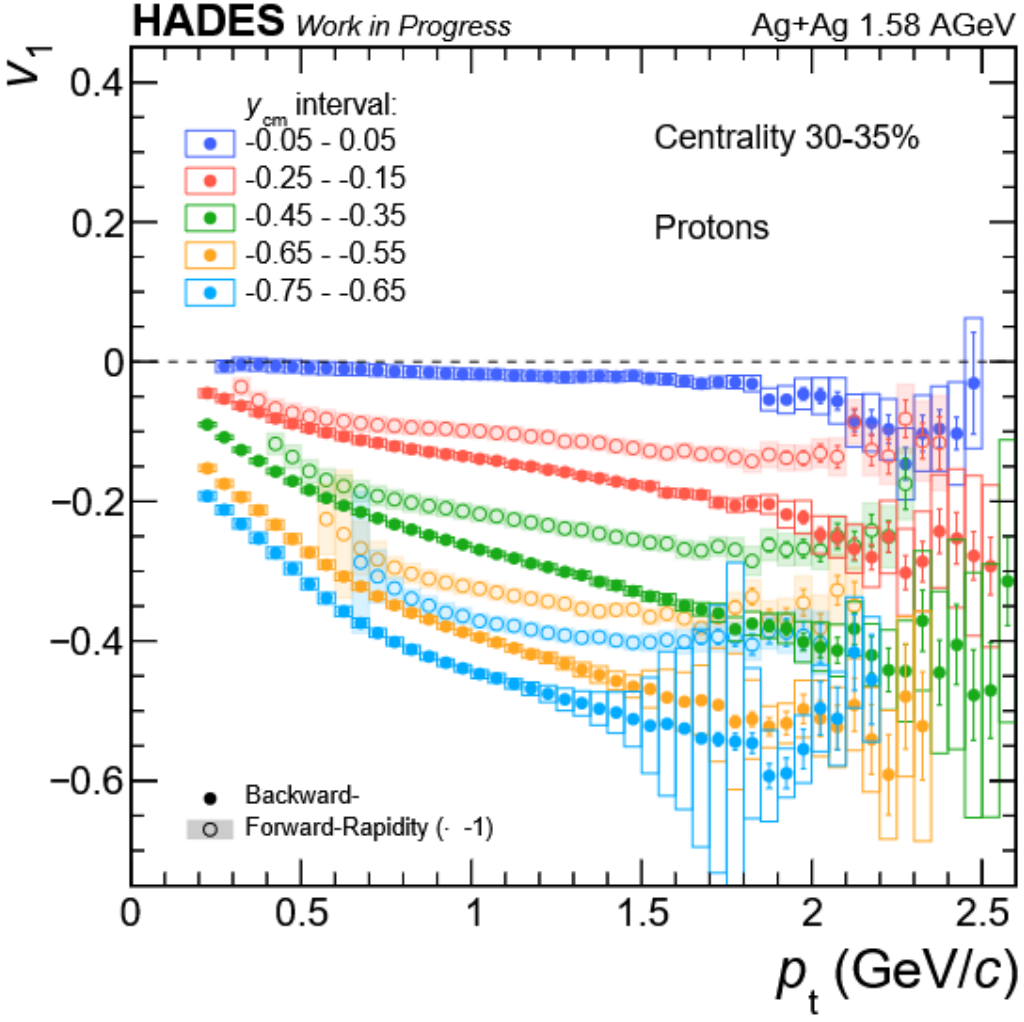


3. MODEL EVALUATION - DATA

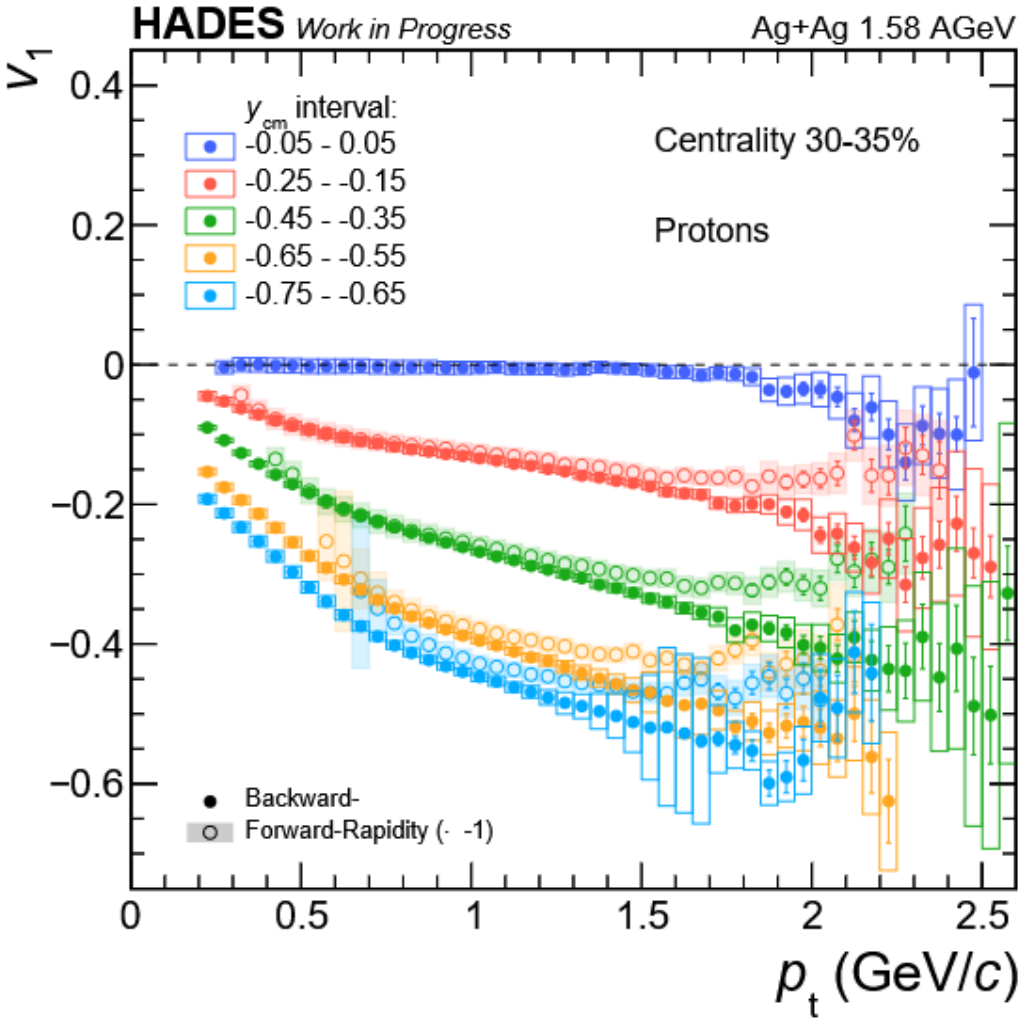
p

HADES

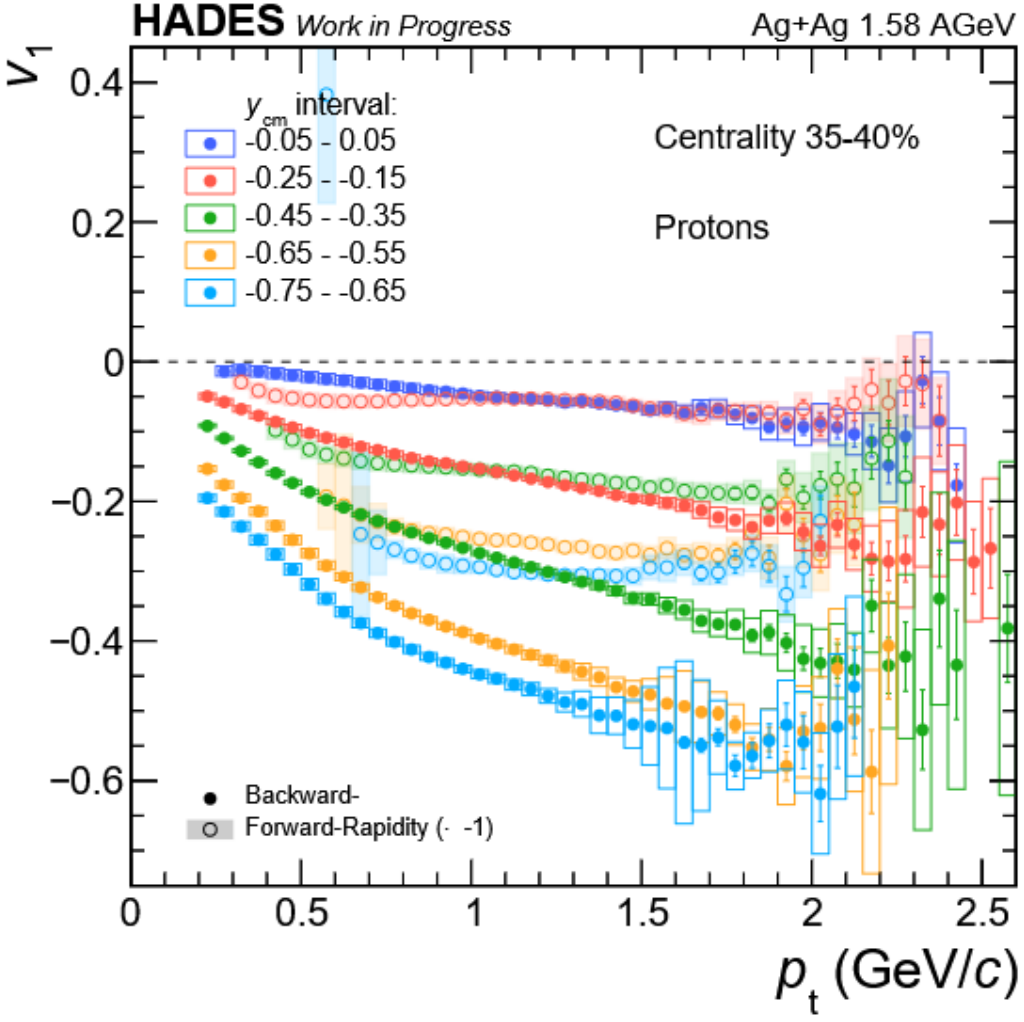
TECHNISCHE UNIVERSITÄT DARMSTADT



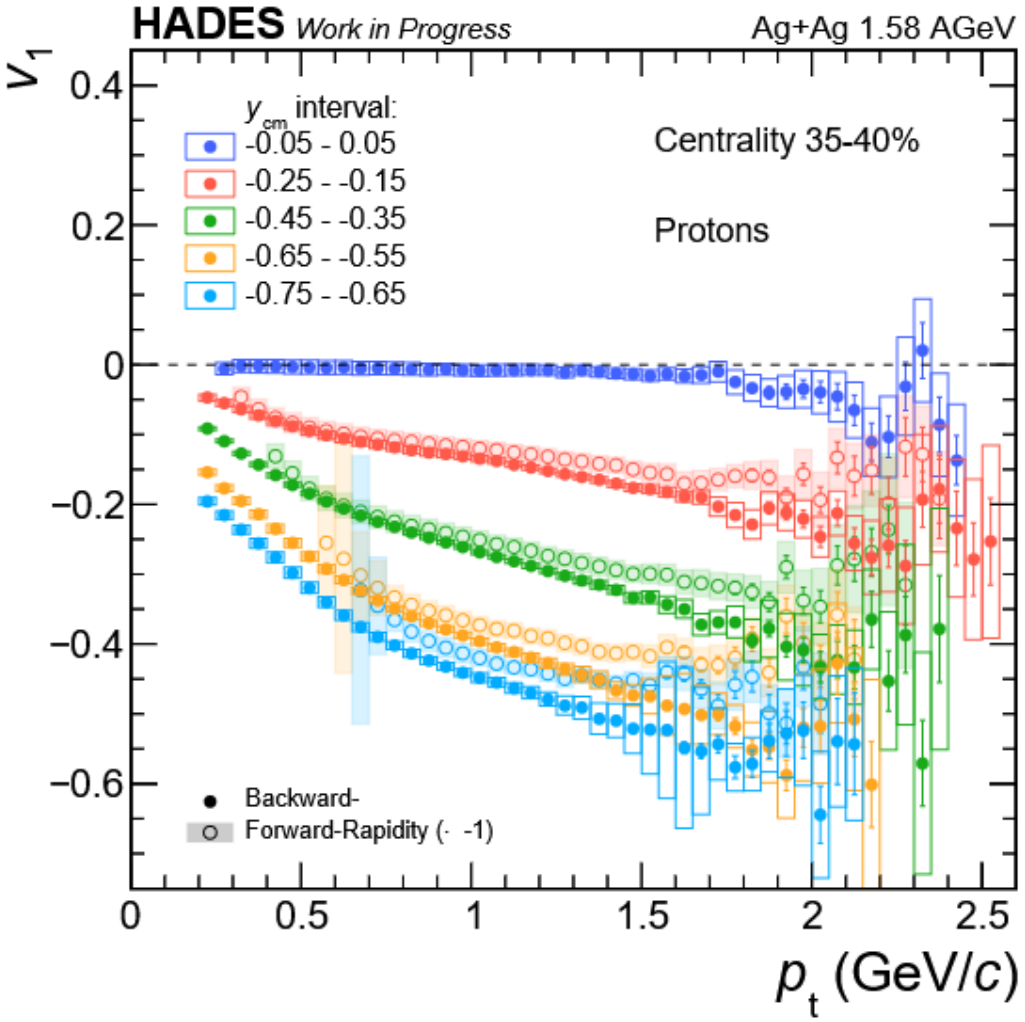
Ag pur > 0.5



3. MODEL EVALUATION - DATA



Ag pur > 0.5

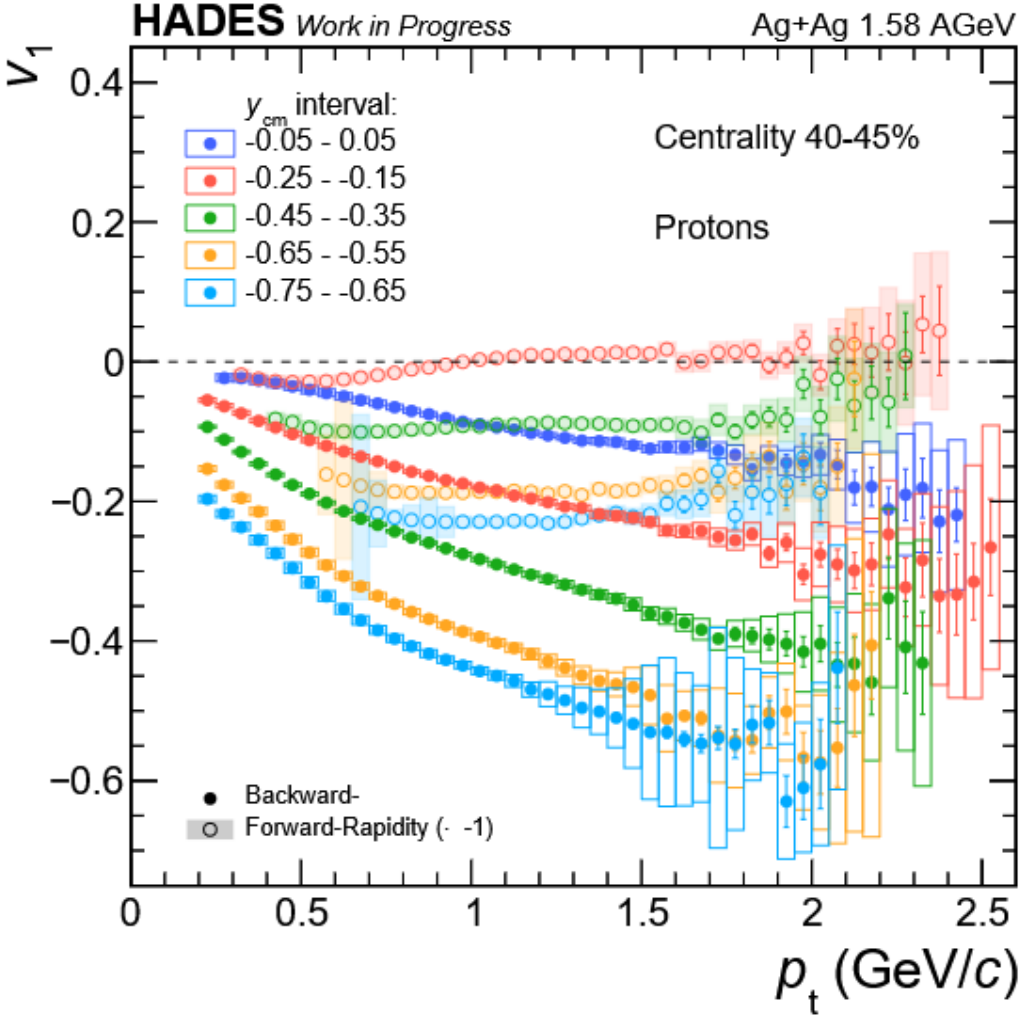


3. MODEL EVALUATION - DATA

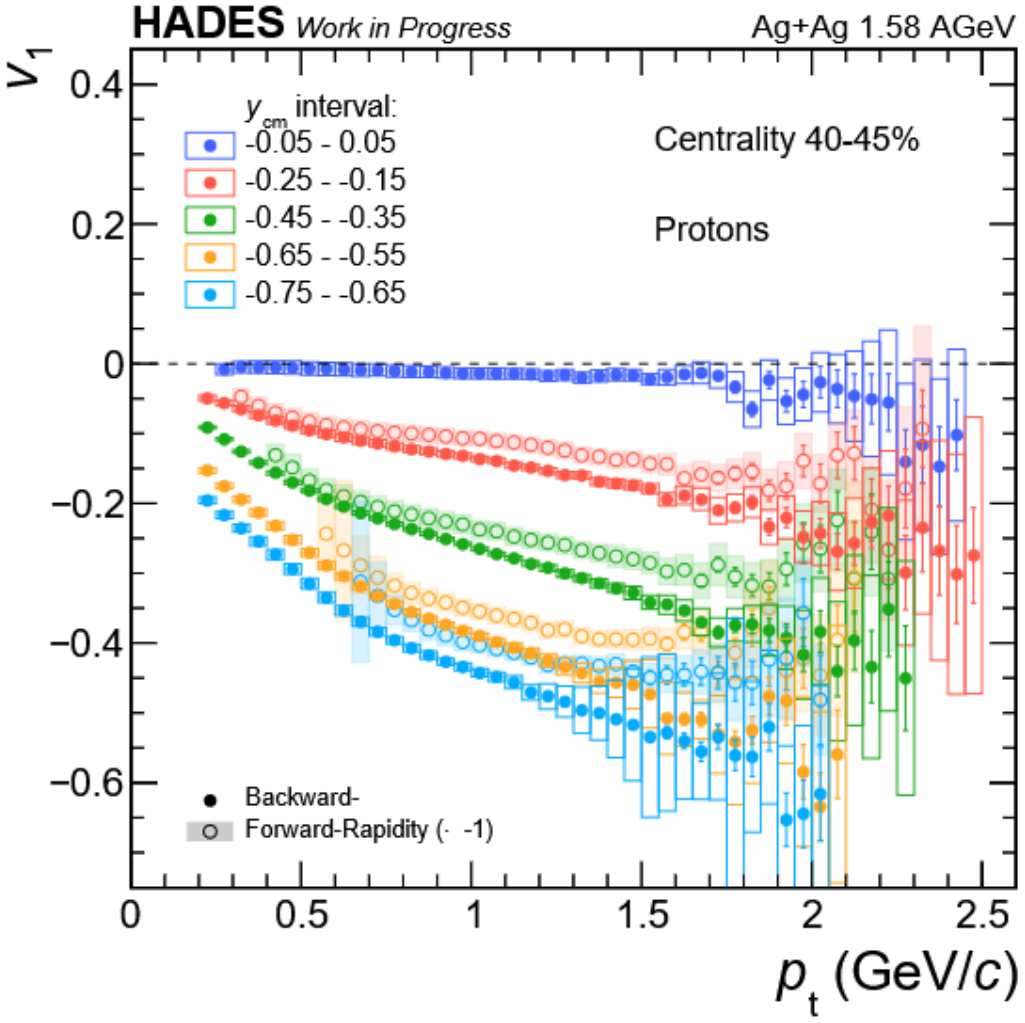
p

HADES

TECHNISCHE UNIVERSITÄT DARMSTADT



Ag pur > 0.5

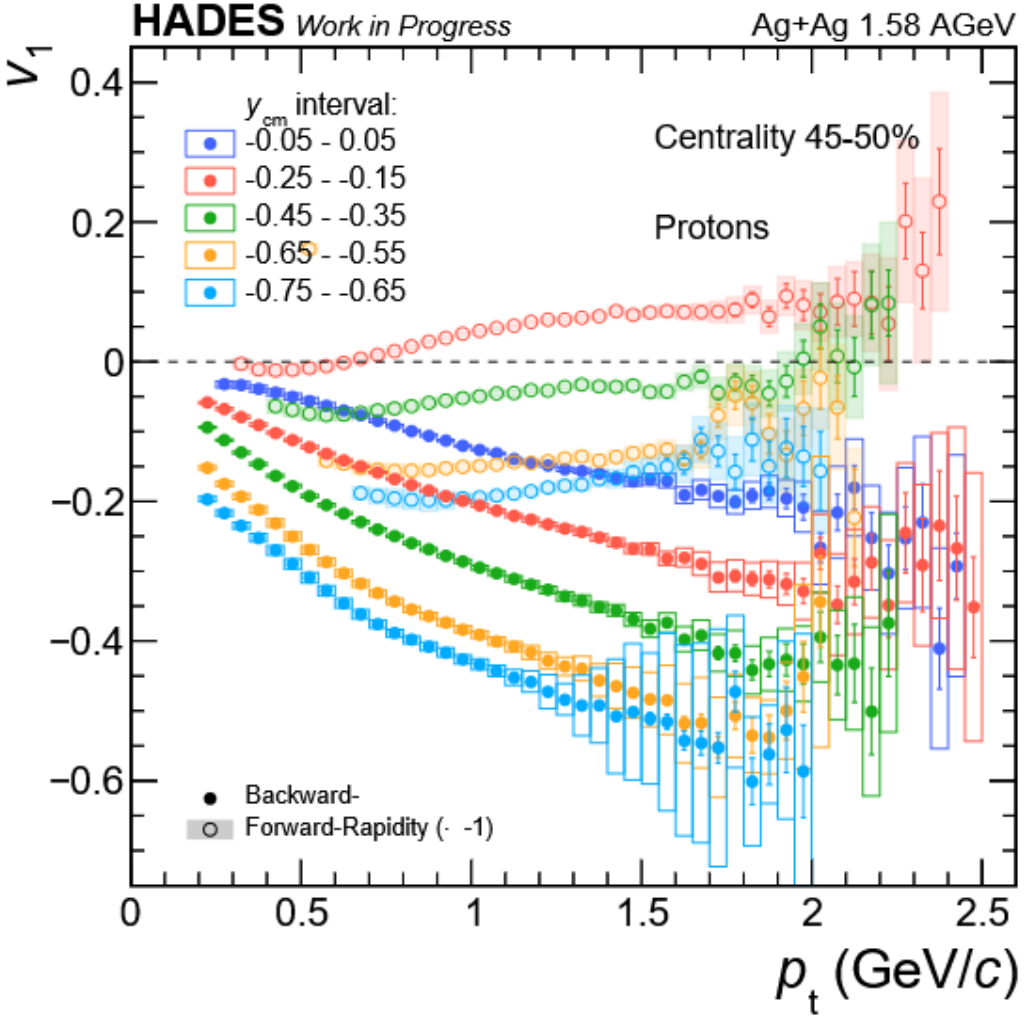


3. MODEL EVALUATION - DATA

p

HADES

TECHNISCHE UNIVERSITÄT DARMSTADT



Ag pur > 0.5