Updates on 1 GeV beam π^+-Ar inclusive cross section measurement

Yinrui Liu[1], Ed Blucher[1], Tingjun Yang[2]

Apr.14, 2022 @ Hadron Analysis Meeting

[1]University of Chicago

[2]FNAL

Previous energy slicing result

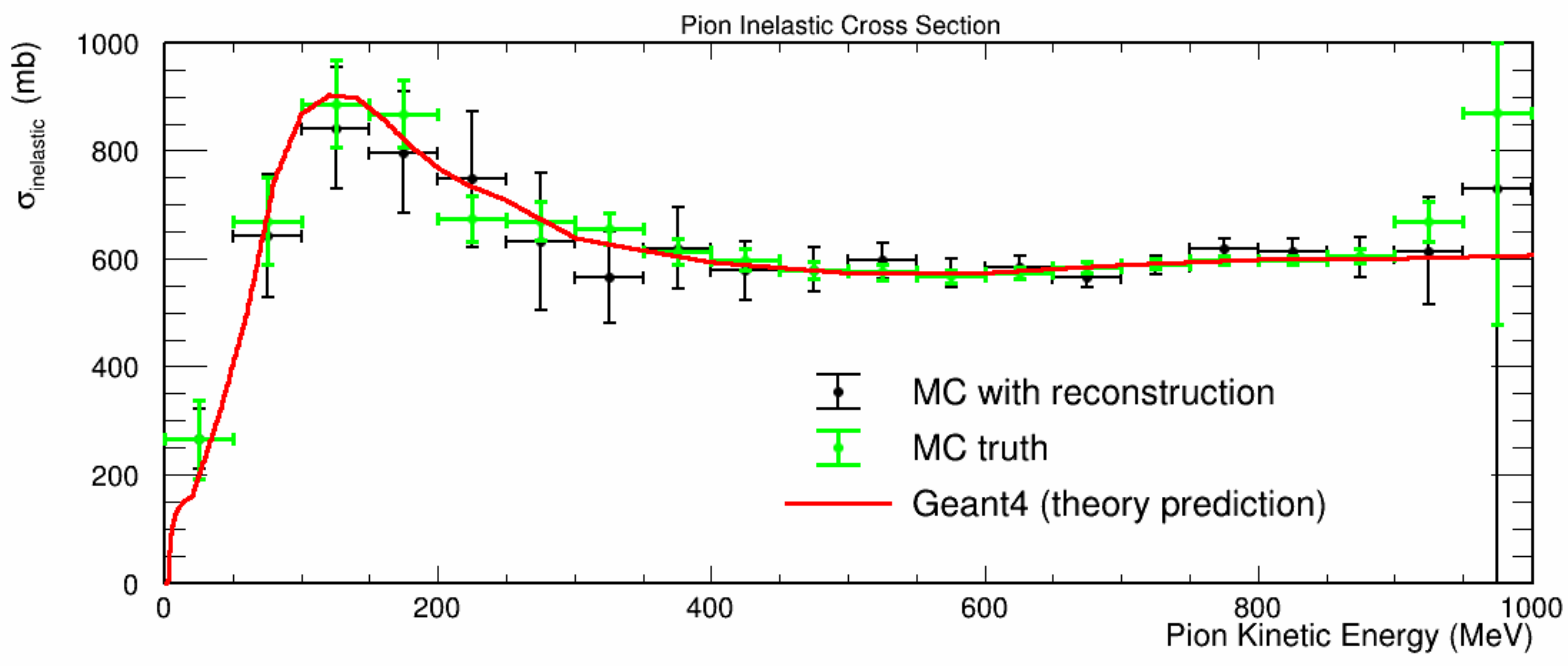
- Last time, we validated the energy slicing method using fake data.

https://indico.fnal.gov/event/53325/contributions/235193/attachments/152412/197326/pionXS_DRA_220223.pdf

$$\sigma = \frac{M_{\mathrm{Ar}}}{\rho N_A \Delta E} \frac{dE}{dx}(E) \ln\left(\frac{N_{\mathrm{inc}}(E)}{N_{\mathrm{inc}}(E) - N_{\mathrm{int}}(E)}\right)$$

$$N_{\mathrm{inc}}(i) = \sum_{j=i}^{N} N_{\mathrm{int}}(j) - \sum_{j=i+1}^{N} N_{\mathrm{ini}}(j)$$

Direct measurements



Validating by fake data

- Reco: measured fake data (after selections, background constraints, unfolding)
- True: truth info of fake data

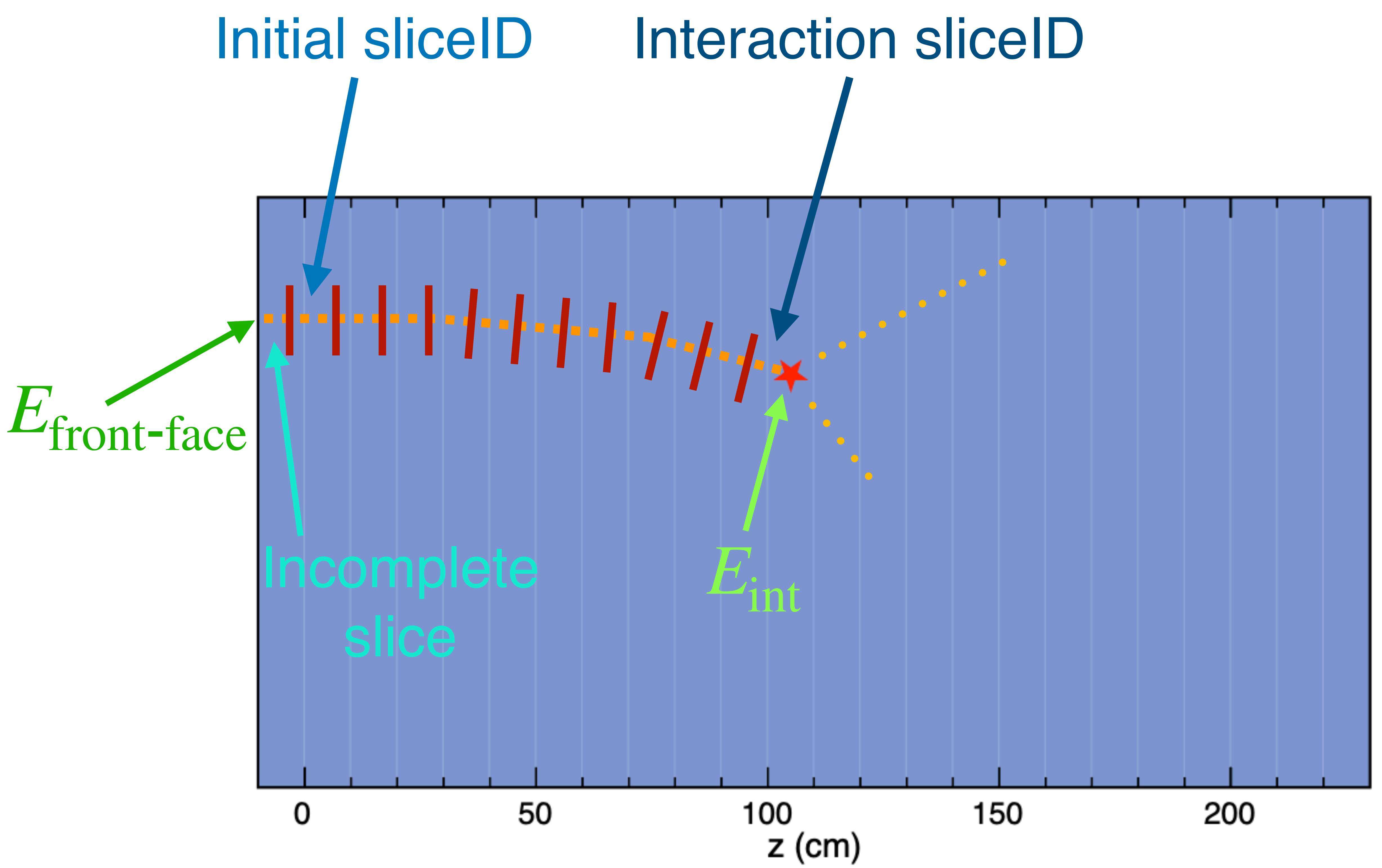
Contents

- We update the implementation of energy slicing method, partly motivated by some problems we encountered in mock-data test.
- In this talk, I will just show the results (some more details will be given in back-ups)
  - The latest implementation of energy slicing method
    - Ignore incomplete slices
    - Removing APA3 cut
  - Mock-data tests using the latest implementation

Ignore incomplete slices

Consider $\Delta E = 50$ (MeV)

- Interaction sliceID: $\text{floor}\left(\frac{1000 - E_{\text{int}}}{\Delta E}\right)$
  - SliceID $= 0 \longleftrightarrow E_{\text{int}} \in [950,1000)$
  - …
  - SliceID $= 19 \longleftrightarrow E_{\text{int}} \in [0,50)$
- Initial sliceID: $\text{ceil}\left(\frac{1000 - E_{\text{front-face}}}{\Delta E}\right)$
  - SliceID $= 0 \longleftrightarrow E_{\text{ff}} \geq 1000$
  - SliceID $= 1 \longleftrightarrow E_{\text{ff}} \in [950,1000)$
  - …



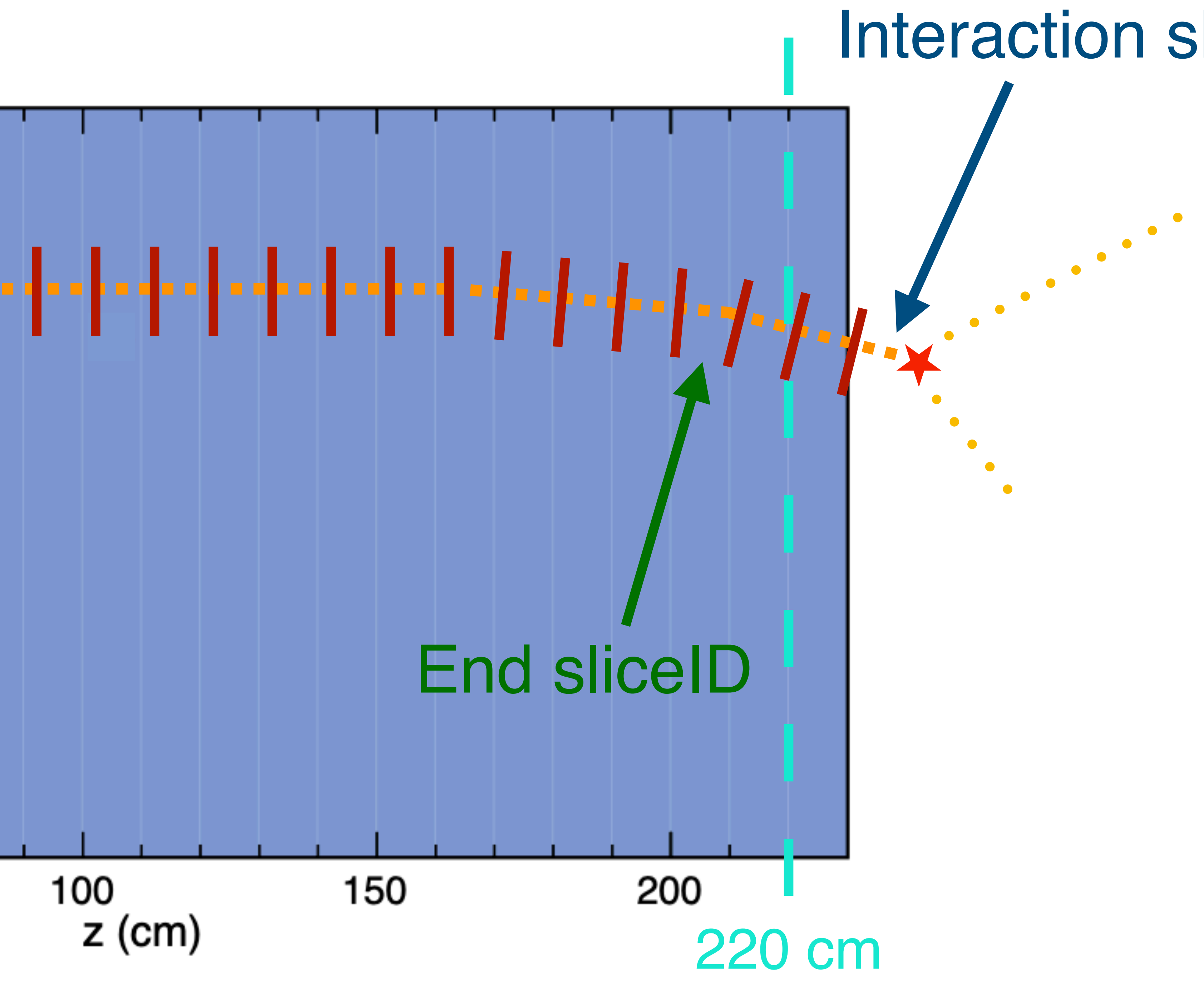
More details are given in back-up 29

Removing APA3 cut

- APA3 cut (reco_beam_calo_endZ < 220) is to cut tracks which extend into the second TPC (since it is likely to be distorted at the boundary of two TPCs), and it can also mitigate muon backgrounds.
  - It removes ~3.5% events, which are majorly long tracks.
- However, this cut can bring bias since the vetoed are all non-interaction events in the high energy slices, which means these vetoed events should have been counted in the incident histogram, and not in the interaction histogram.
  - Thus, with APA3 cut, we are likely to overestimate the cross-section in the high energy slices.

Removing APA3 cut

- Now we don't cut those long tracks, instead, we ignore slices in the second TPC.



- To do this, we define end sliceID
  - If endZ < 220 cm
    - End sliceID = interaction sliceID $\text{floor}\left(\frac{1000 - E_{\text{int}}}{\Delta E}\right)$
  - If endZ >= 220 cm
    - End sliceID = $\text{floor}\left(\frac{1000 - E_{220}}{\Delta E}\right) - 1$

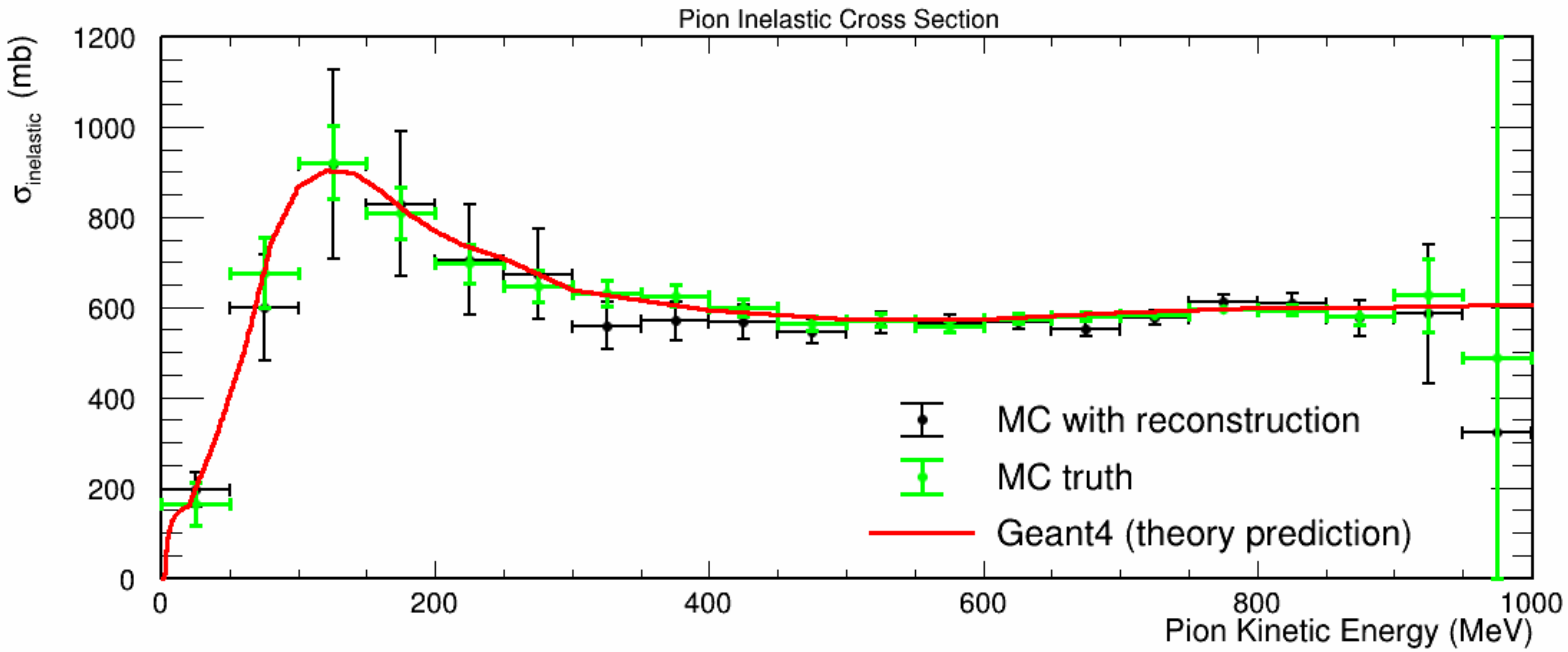
The energy at Z == 220 cm

The slice at 220 cm is incomplete, so we ignore it.

Updated result

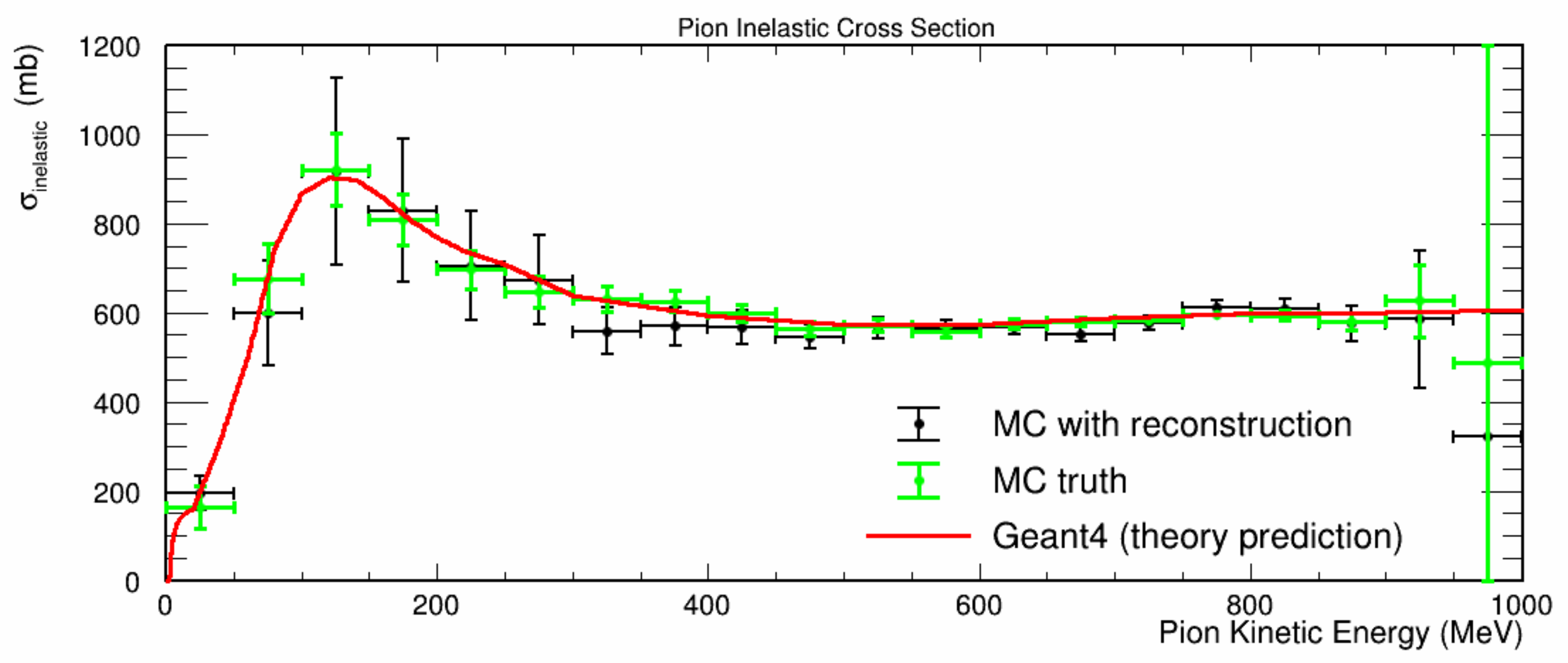
Since end sliceID differs from interaction sliceID, now we have

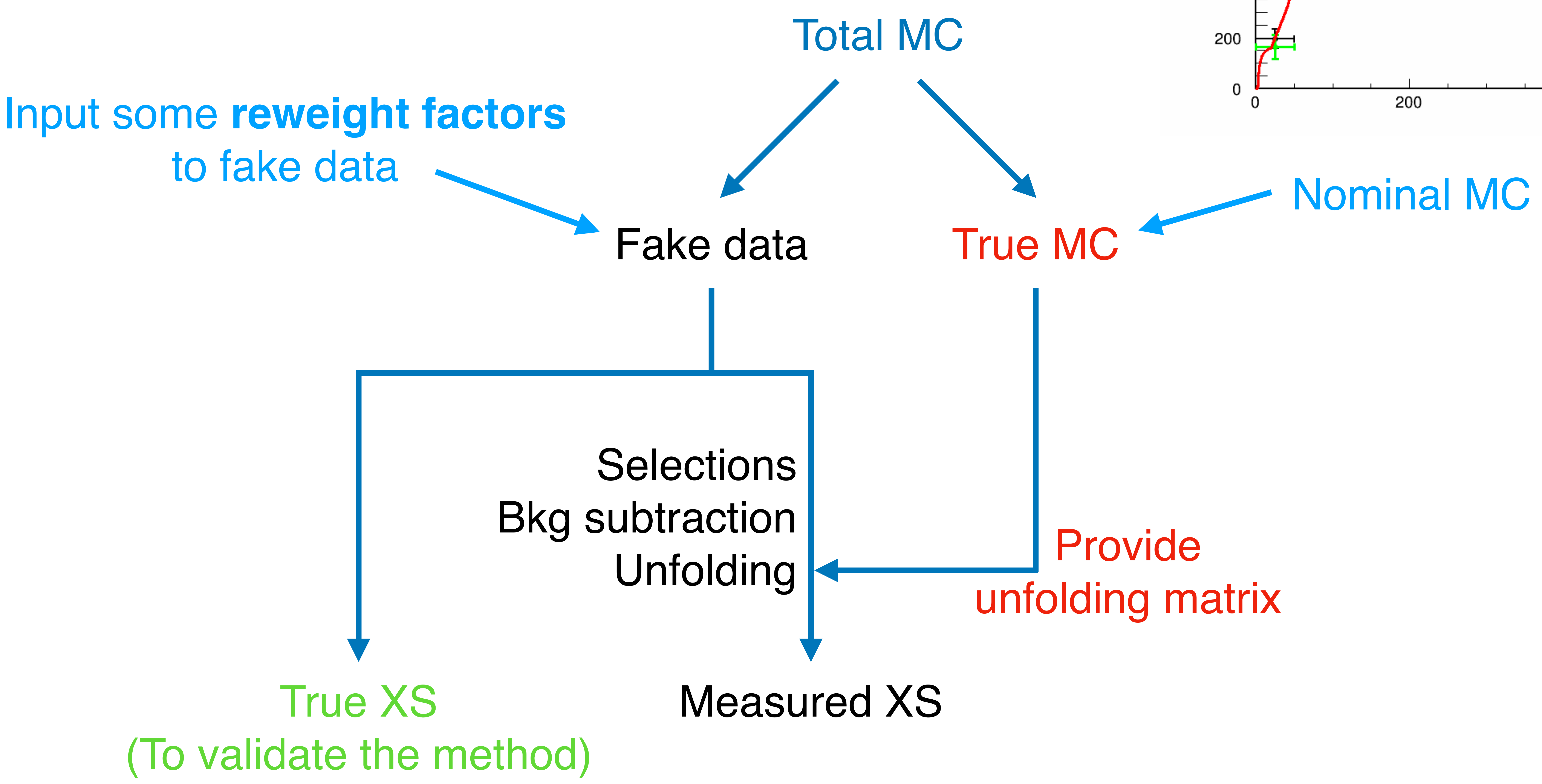
$$N_{\text{inc}}(i) = \sum_{j=i}^{N} N_{\text{end}}(j) - \sum_{j=i+1}^{N} N_{\text{ini}}(j)$$



- Reco: measured fake data (after selections, background constraints, unfolding)
- True: truth info of fake data

Mock-data test

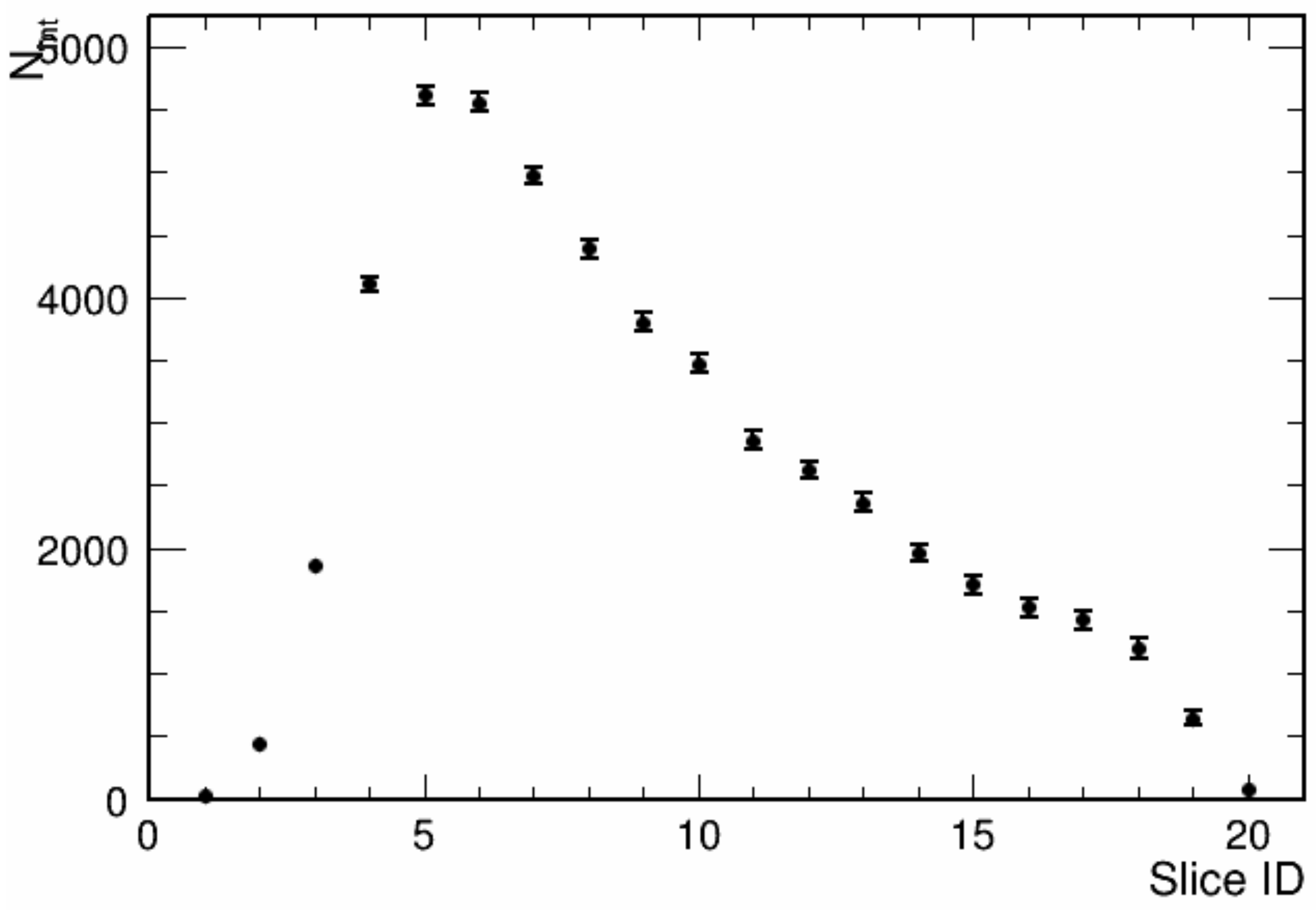




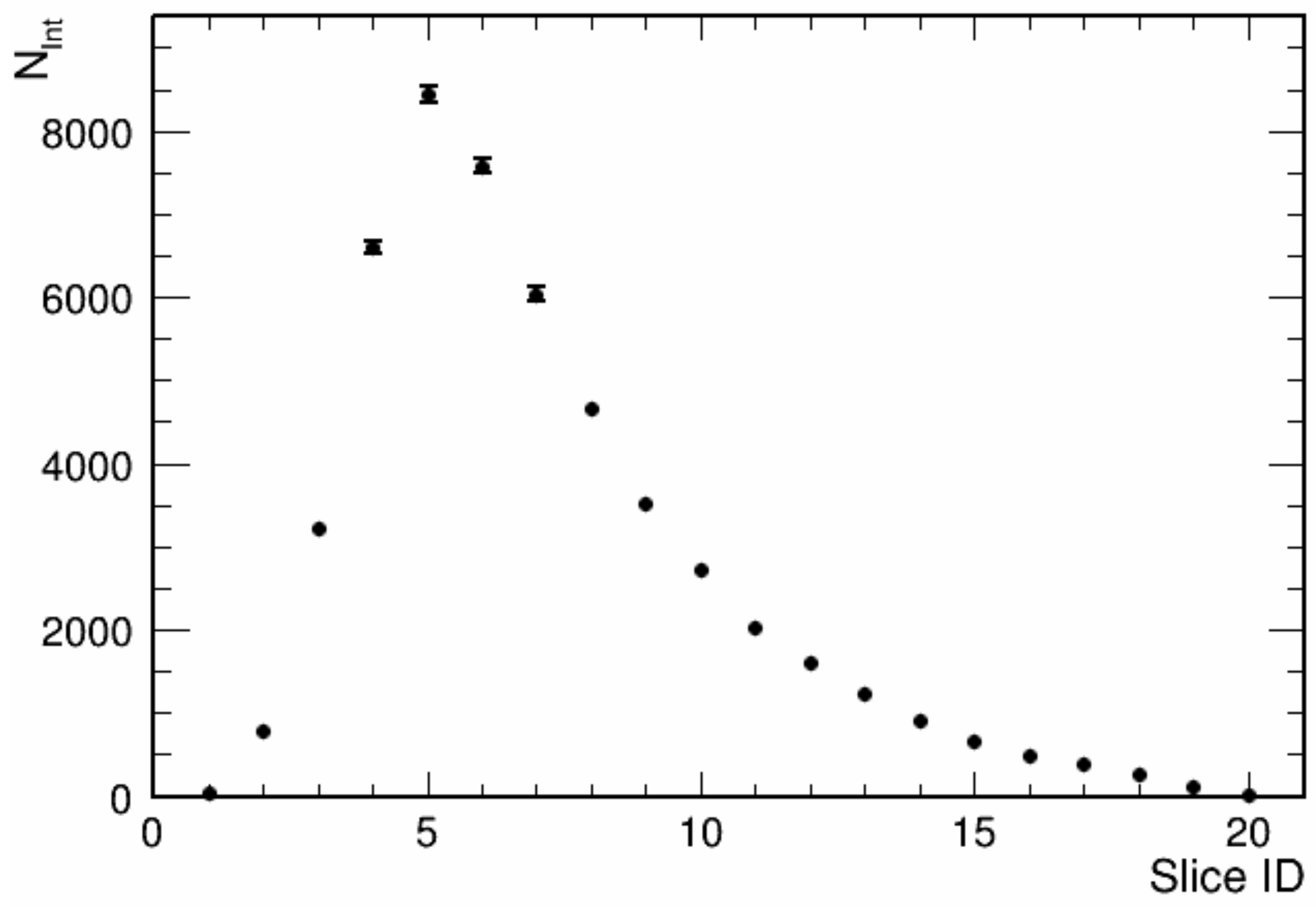
Mock-data test

- We use Geant4reweight to derive samples of reweighted cross-sections (more details in back-up 31-32) Jacob Calcutt, et. al. https://arxiv.org/abs/2105.01744

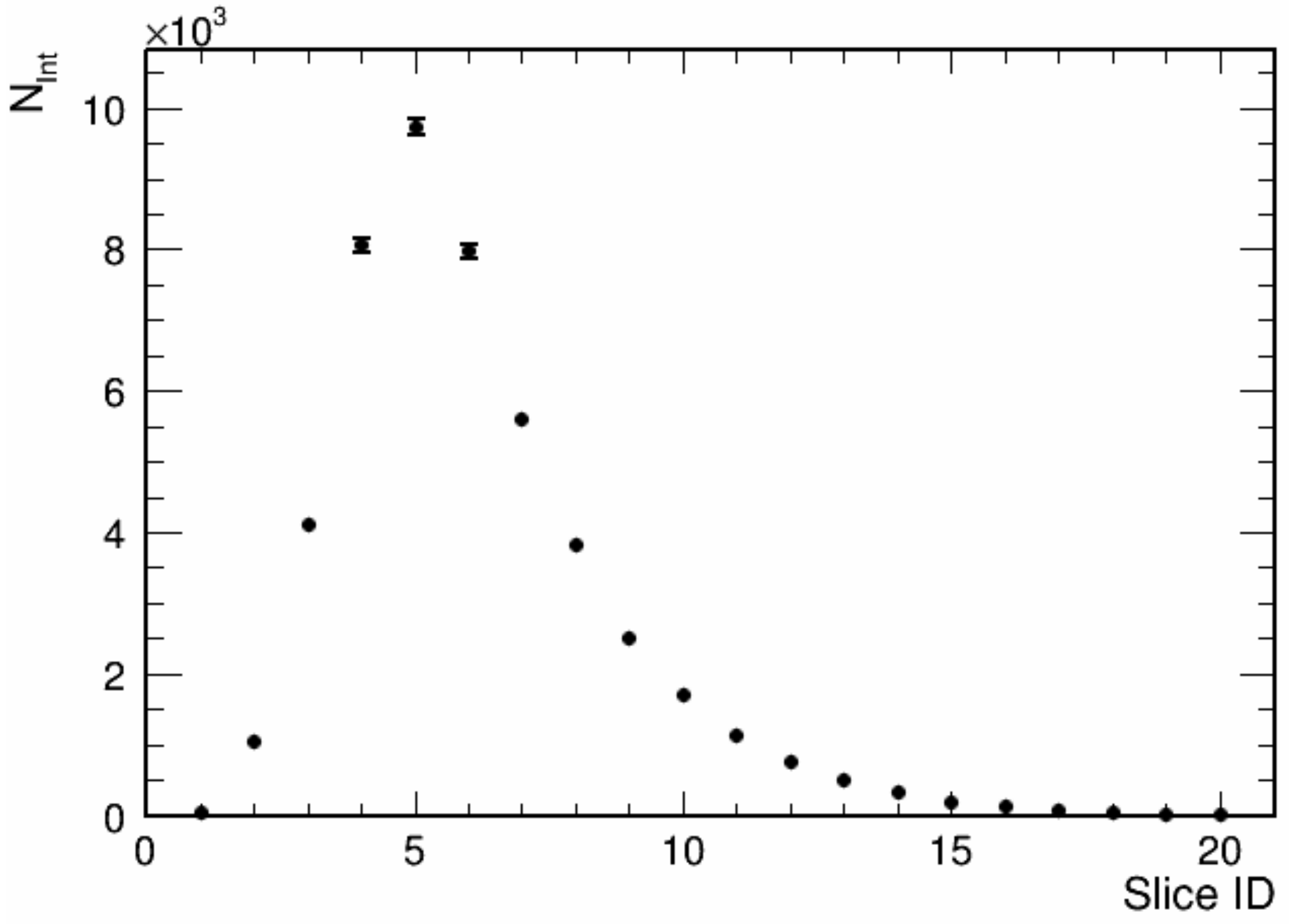
- N_{int} of fake data with different XS



True XS reweighted to 0.5



Nominal true XS

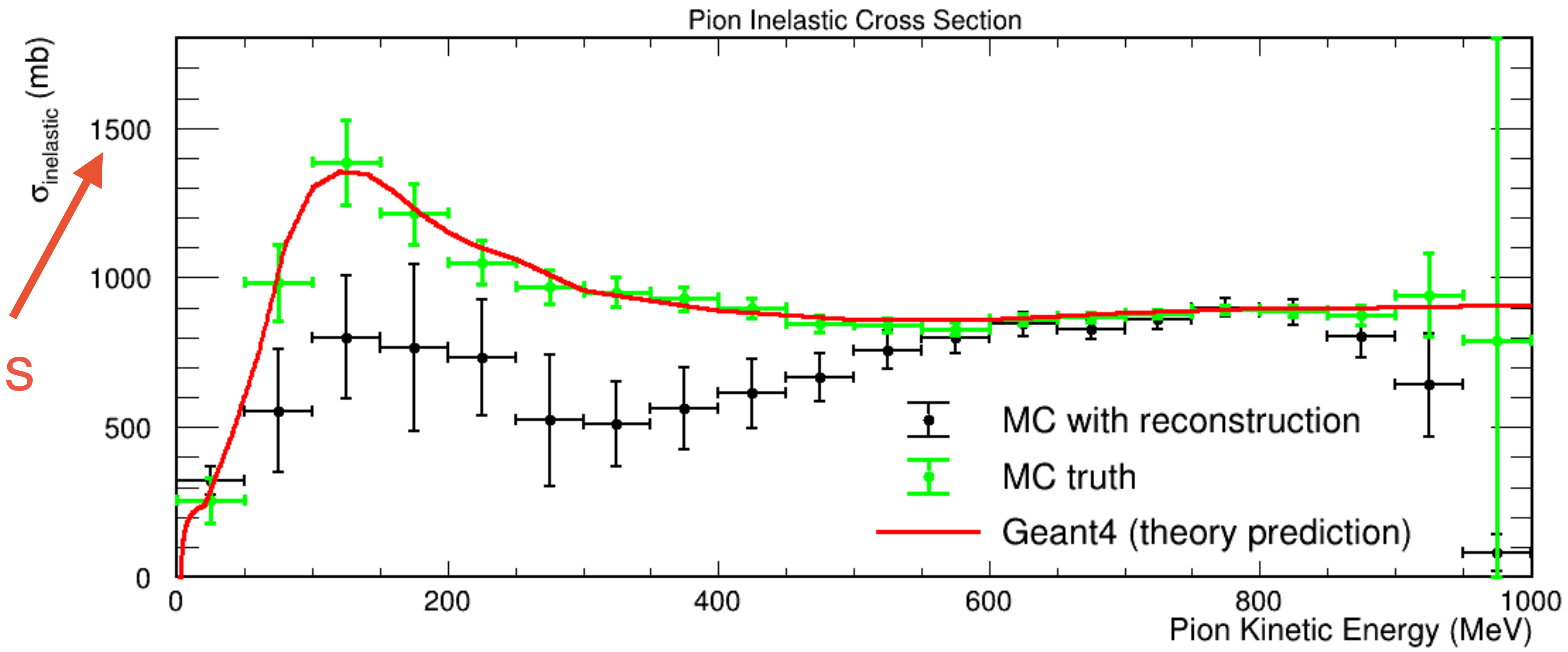


True XS reweighted to 1.5

Example: reweight XS of fake data to 1.5

- The XS calculated using truth info (green points) is as expected (with the red curve), but the measured XS (black points) looks weird.

Note the XS is scaled to 1.5

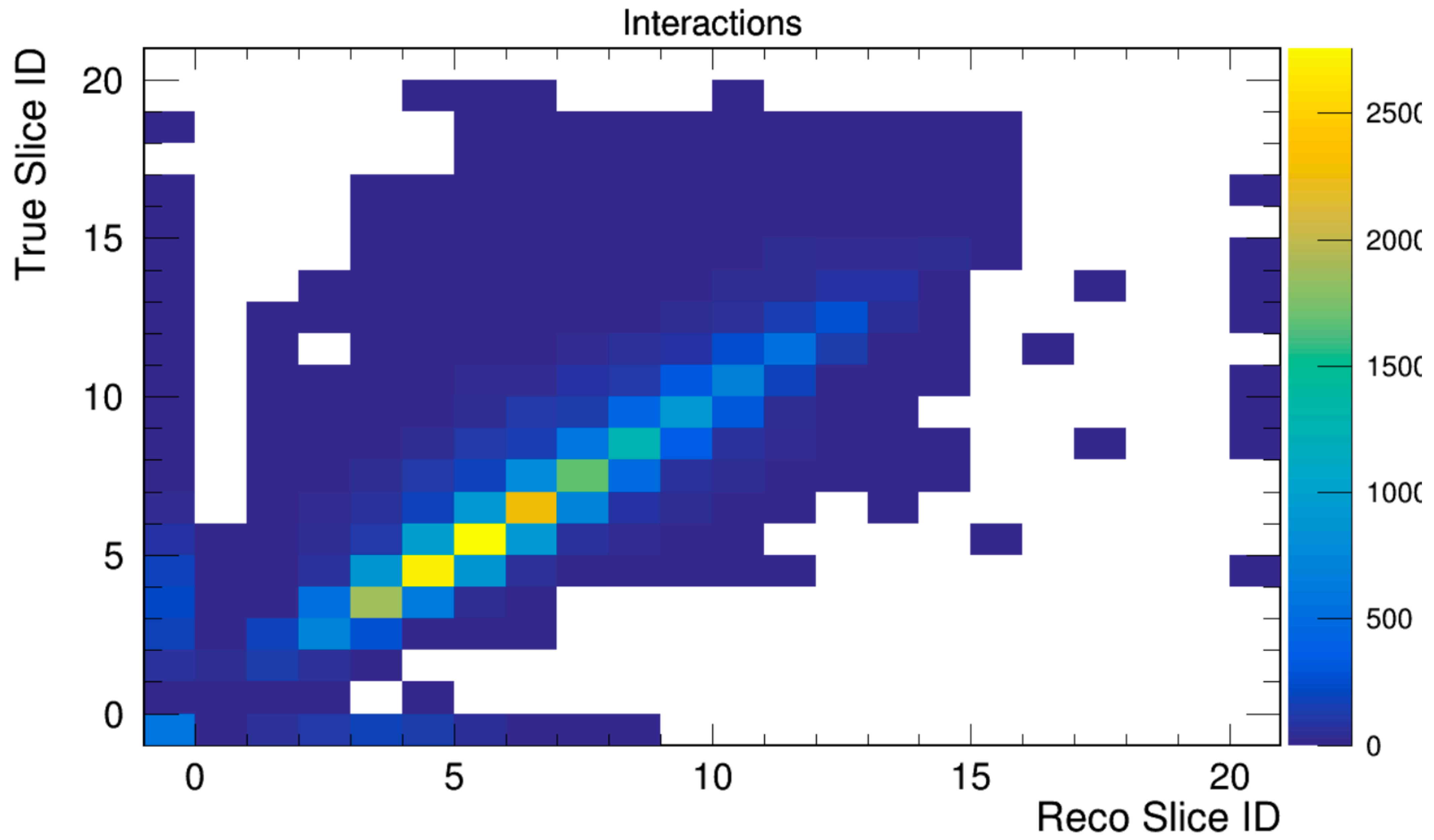


- Reco: measured fake data (after selections, background constraints, unfolding)
- True: truth info of fake data

- We think that's because there is something wrong with unfolding.

Unfolding

- Unfolding accounts for the detection resolution and inefficiency.
- The response matrix R_{ij} is derived by true MC sample, and then applied to fake data.



Response matrix derived by nominal true MC

ν_i expected observed value

μ_i expected true value

$$\nu_i = \sum_{j=1}^{N} R_{ij}\mu_j$$

$$R_{ij} = F(\text{observed value in bin } i \,|\, \text{true value in bin } j)$$

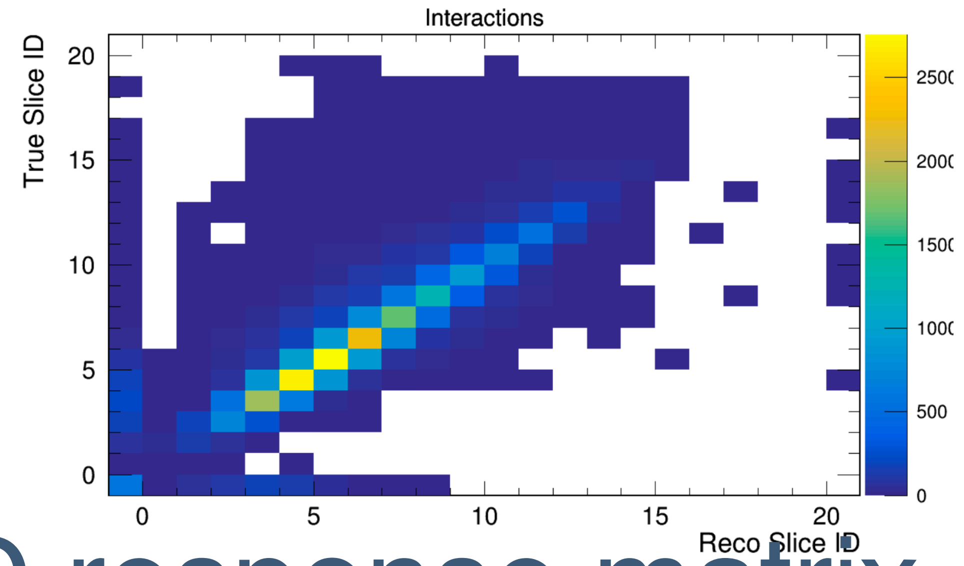
$$= \frac{\int_{\text{bin } i} dx \int_{\text{bin } j} dy\, P(x|y)\varepsilon(y) f_{\text{true}}(y)}{\int_{\text{bin } j} dy\, f_{\text{true}}(y)},$$ calculated by true MC sample

events in bin j of true histogram

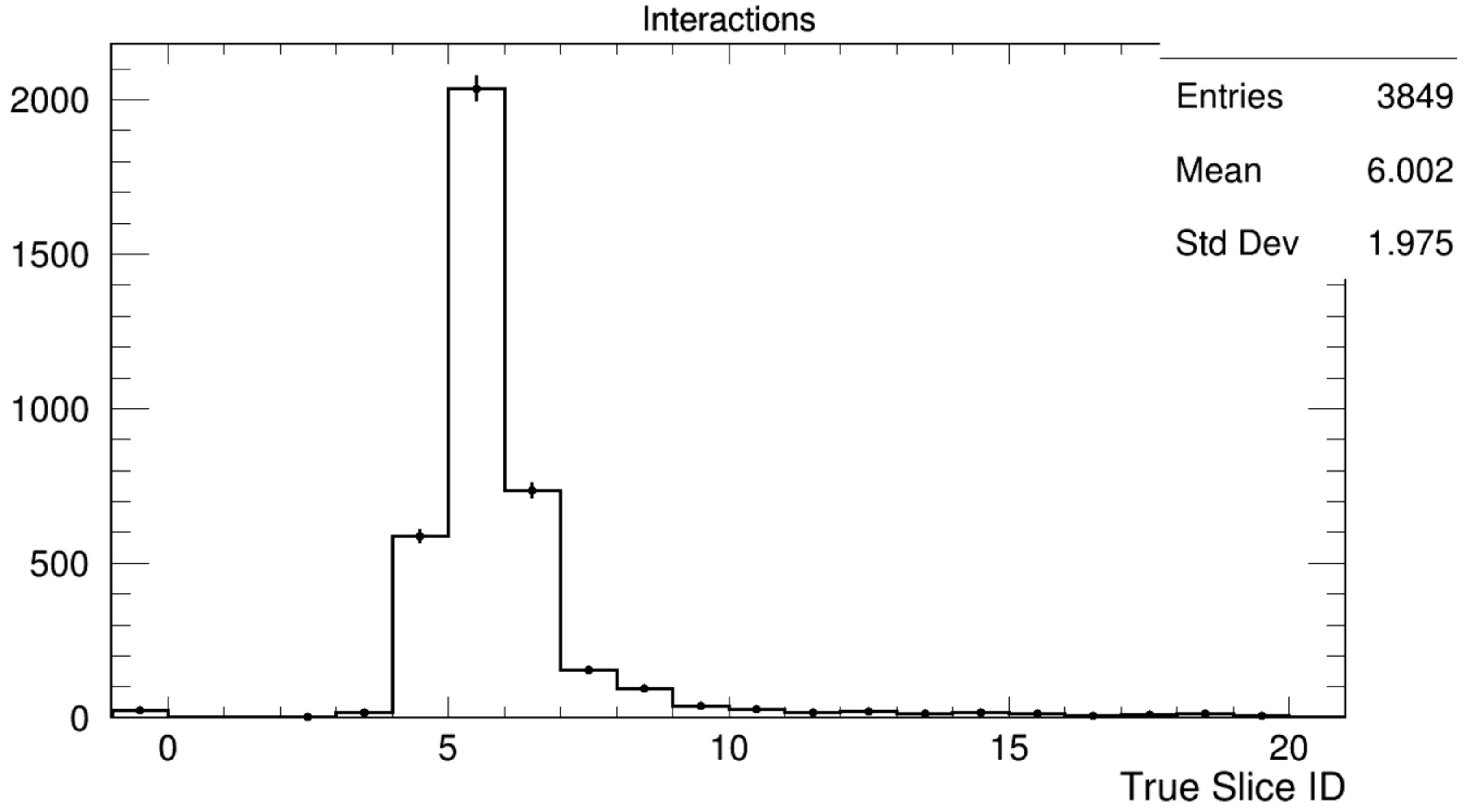
$f_{\text{true}}(y)$ is sample-dependent, which can be different for true MC samples with different cross-sections.

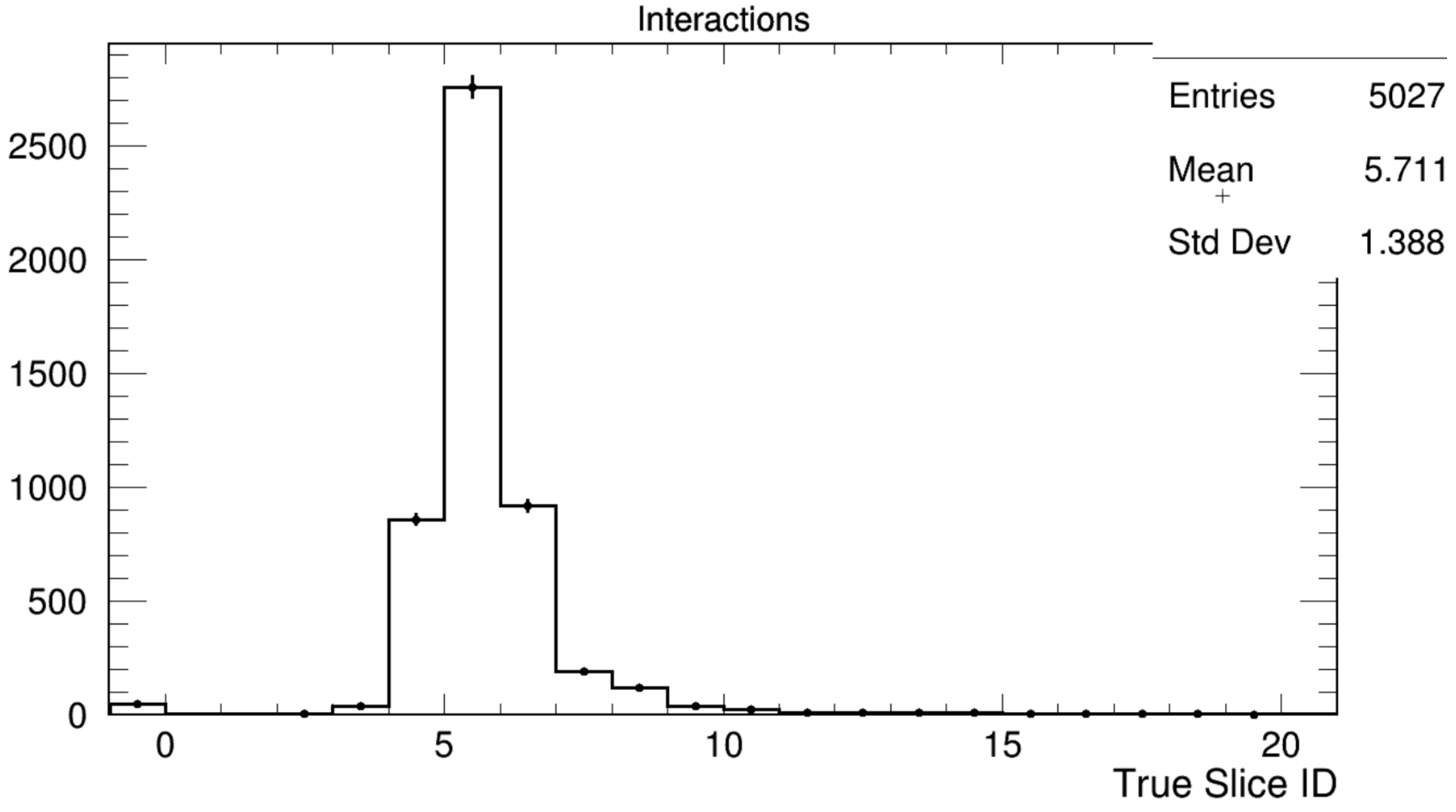
Ref: Glen Cowan, *Statistical Data Analysis*, Chap. 11

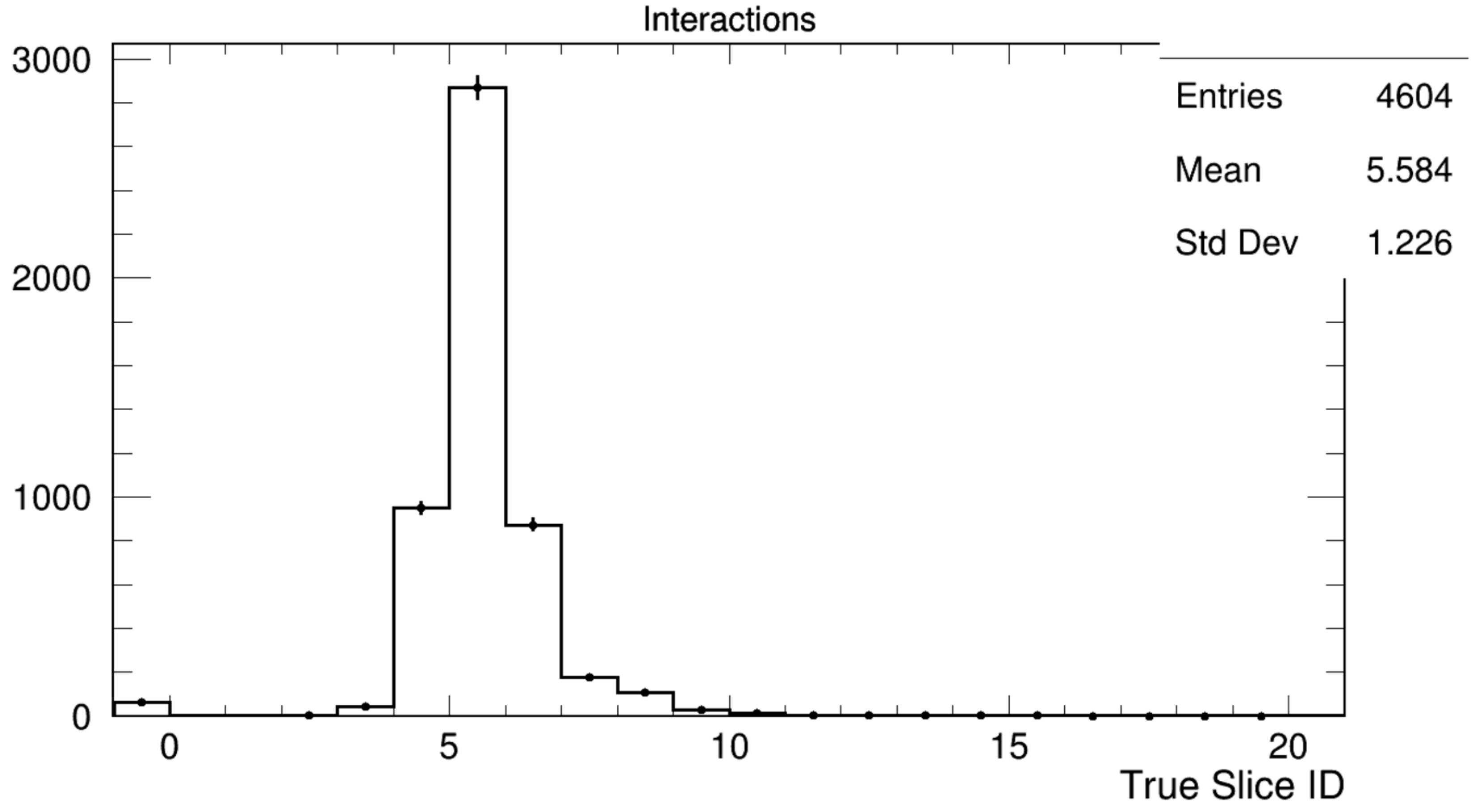
Unfolding



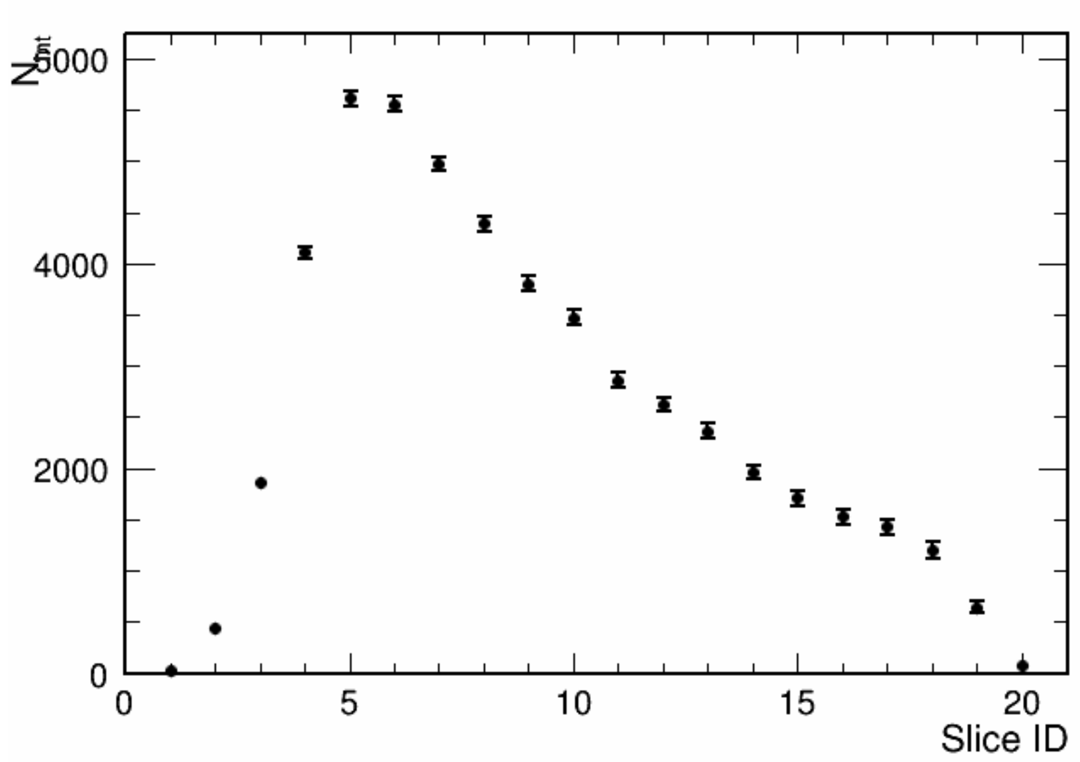
- For data with **reco sliceID == 5**, project the 2D response matrix along true sliceID axis.

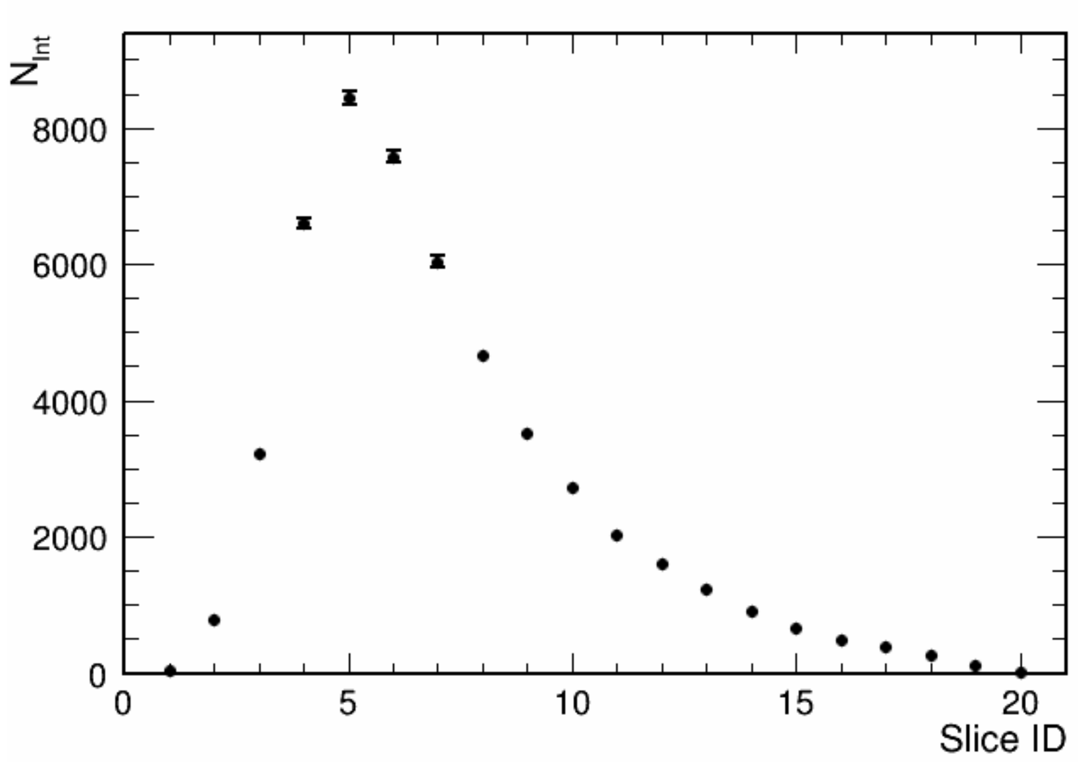


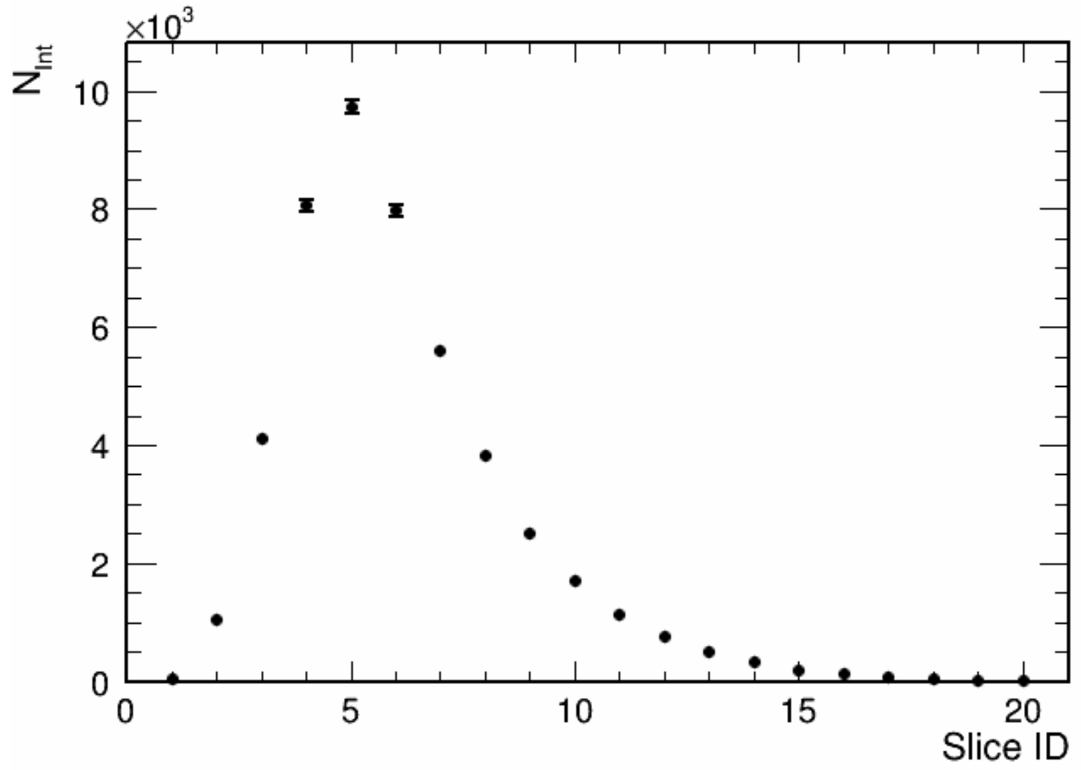




XS of True MC reweighted by 0.5 **Nominal true MC** **XS of True MC reweighted by 1.5**



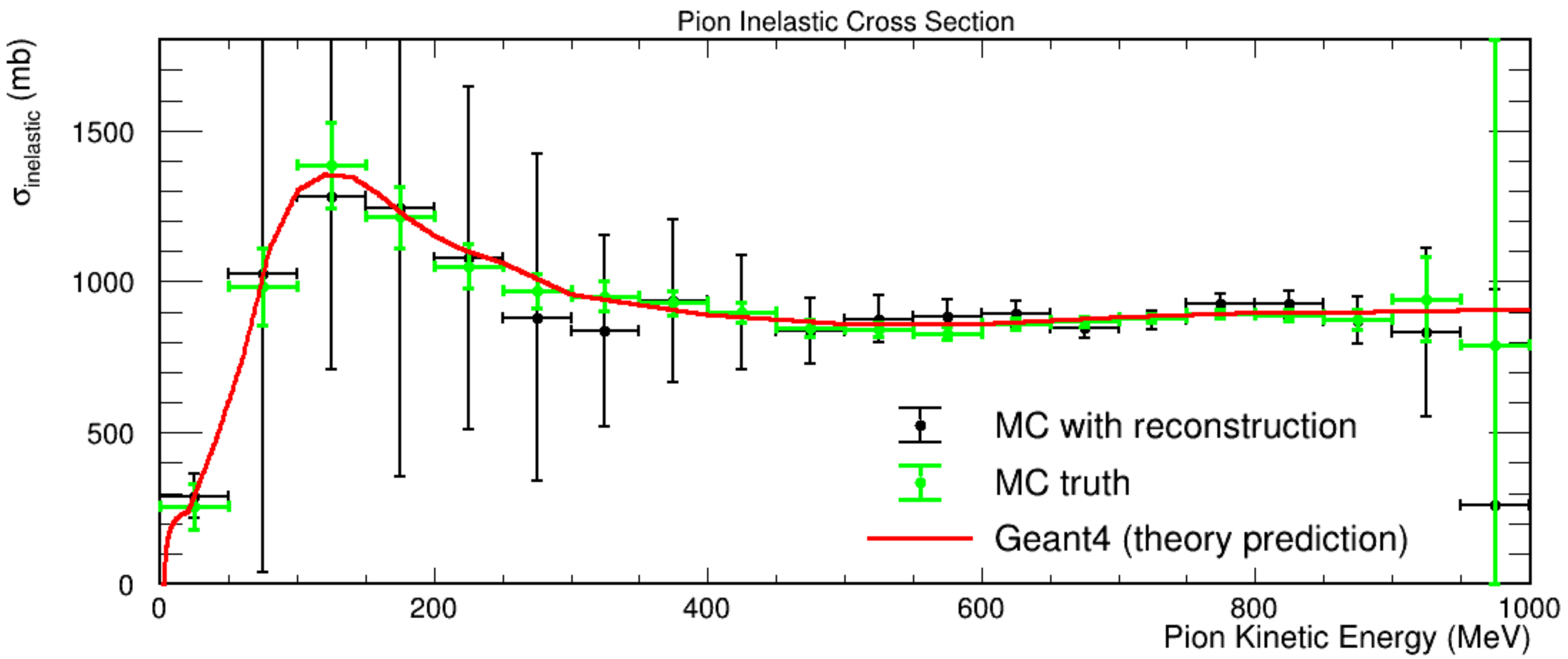




Unfolding matrices derived by different true MC samples are different!

Unfolding

- Unfolding matrices derived by different true MC samples are different!
- Which one is the best? It must be the sample which is most similar to the fake data sample.
- In the example on slide 10, if we also reweight XS of true MC to 1.5 (same as fake data), we get better agreed measured result (black points).



Iterative unfolding

- Therefore, we consider iterative unfolding when measuring XS of real data.
  - In the unfolding, we start from using nominal true MC, then reweight the true MC iteratively until the true MC XS and the measured data XS are consistent with each other.
- In the following slides, we perform a preliminary test of the iterative method on fake data, whose XS is reweighted
  - to 0.5 nominal for momentum < 600 MeV
  - to 1.5 nominal for momentum > 600 MeV
  - Let's see if the iterative method can recover these reweight factors.

These are the only two reweightable ranges for inelastic XS provided by Geant4reweight

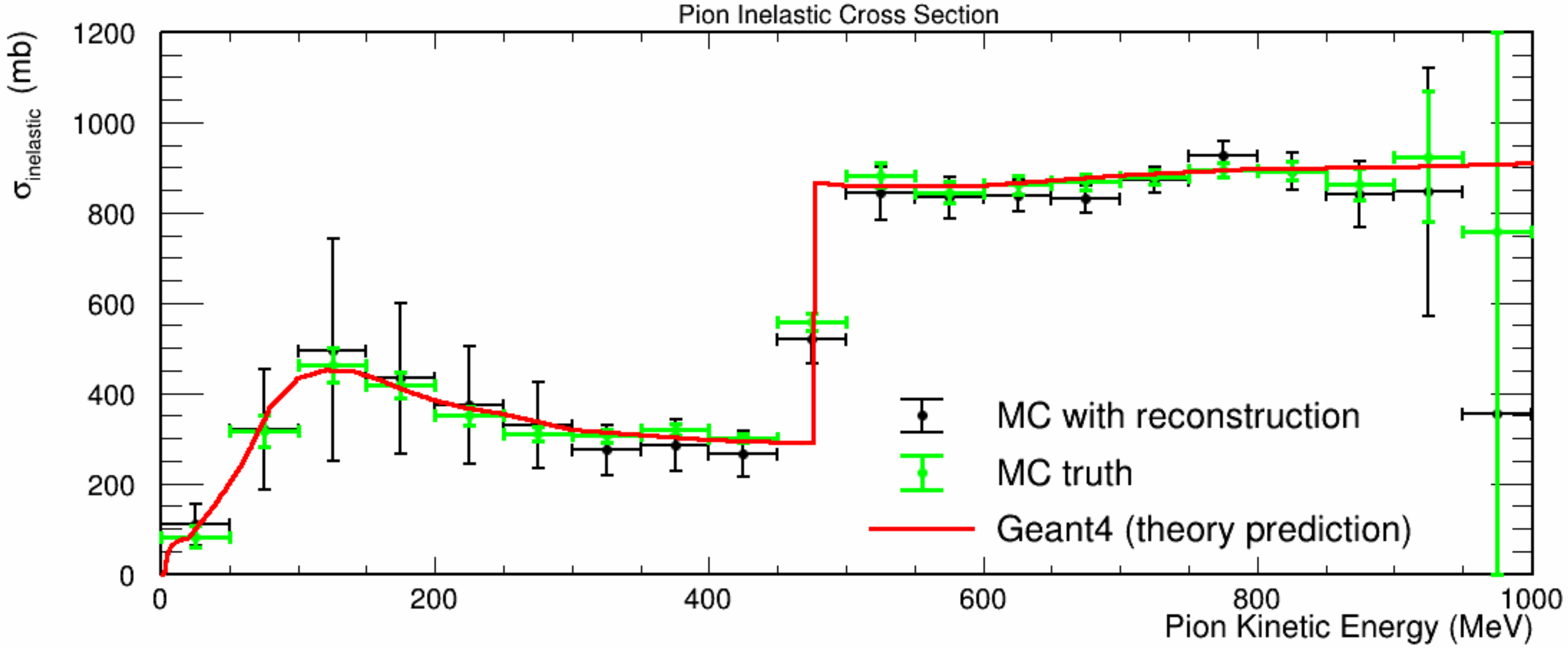
Test iterative unfolding

- True MC scale=[0.5, 1.5], Chi2/Ndf = 0.398362

[P < 600 MeV, P > 600 MeV]

$$\sum\left(\frac{\text{MCreco - Geant4}}{\text{MCreco\_error}}\right)^2$$

Ignore sliceID 0, 10, 19, so Ndf=17



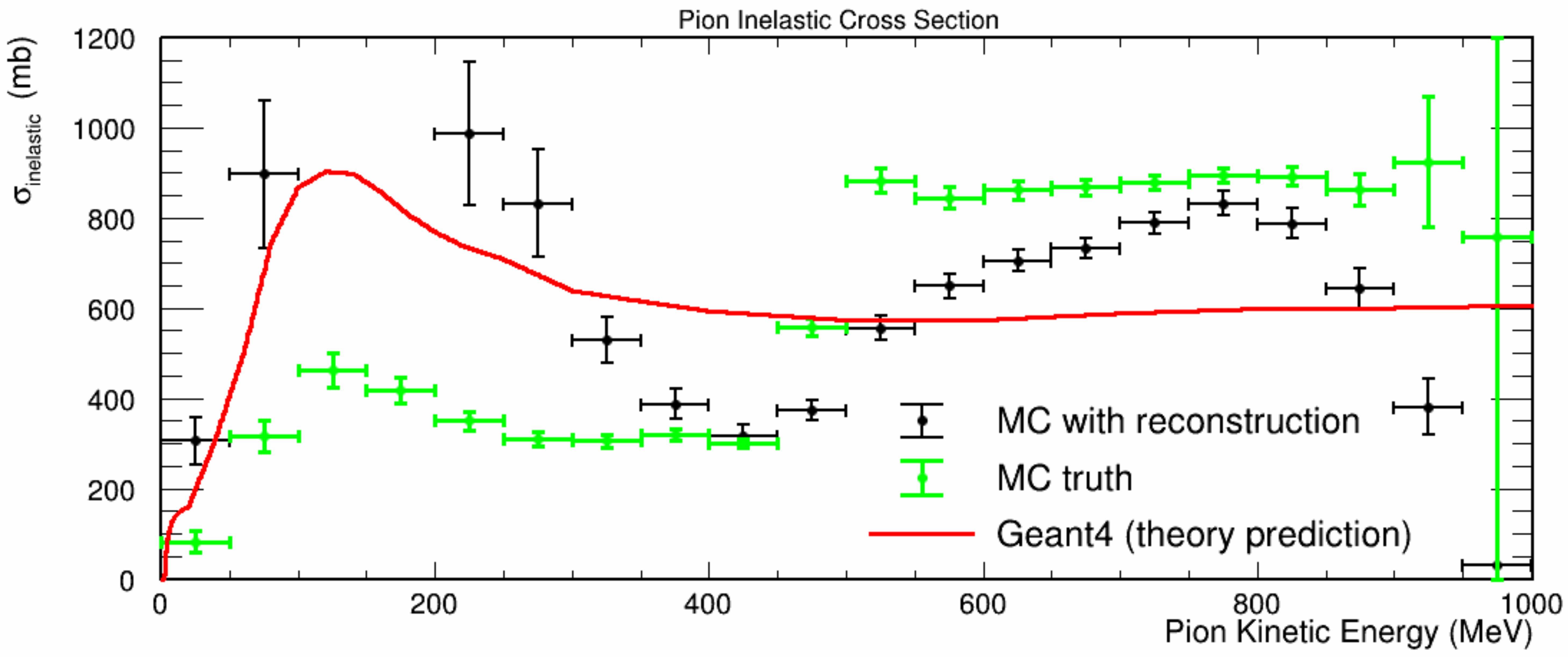
- XS of true MC
- Reconstructed XS of fake data (unfolded by response matrix calculated by true MC)
- True XS of fake data

Iteration 0

- True MC scale=[1., 1.], Chi2/Ndf = 25.3341

$$\sum\left(\frac{\text{MCreco - Geant4}}{\text{MCreco\_error}}\right)^2$$

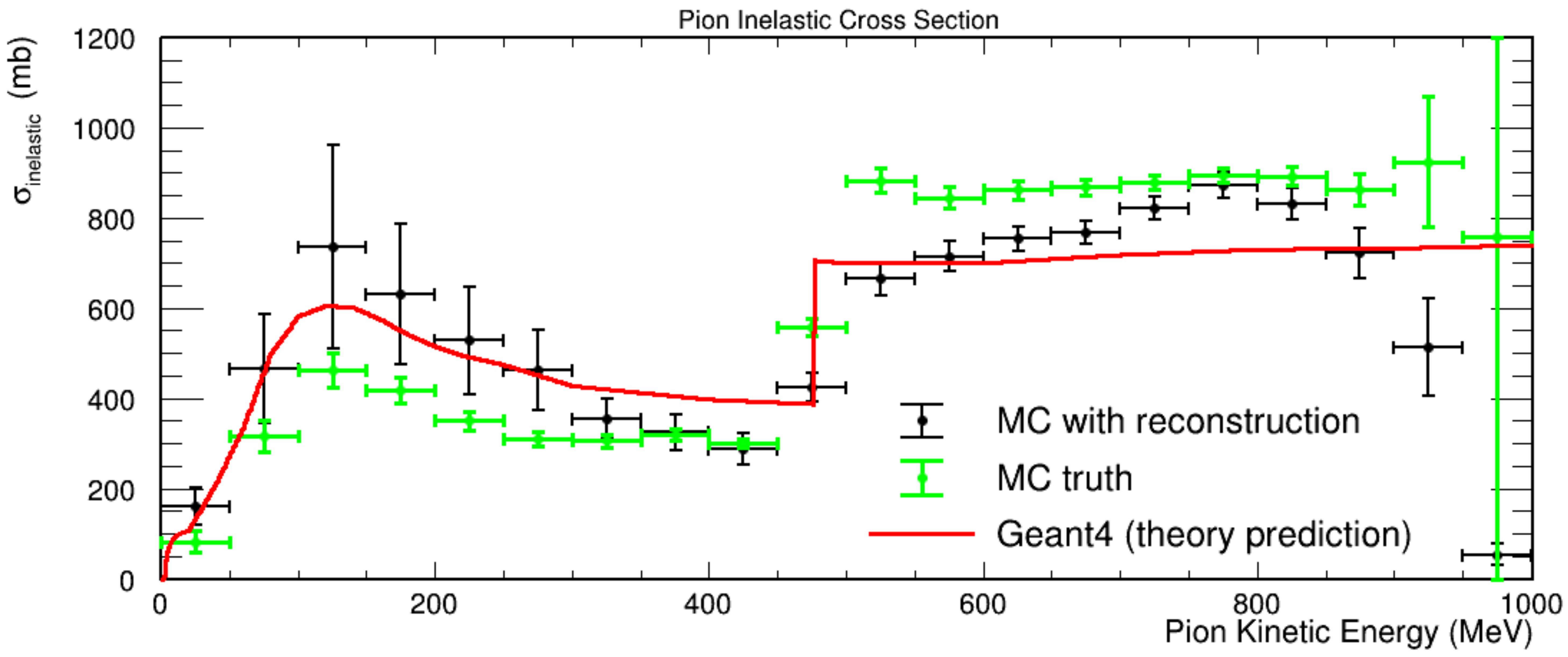
Ignore sliceID 0, 10, 19, so Ndf=17



- XS of true MC
- Reconstructed XS of fake data (unfolded by response matrix calculated by true MC)
- True XS of fake data

Iteration 1

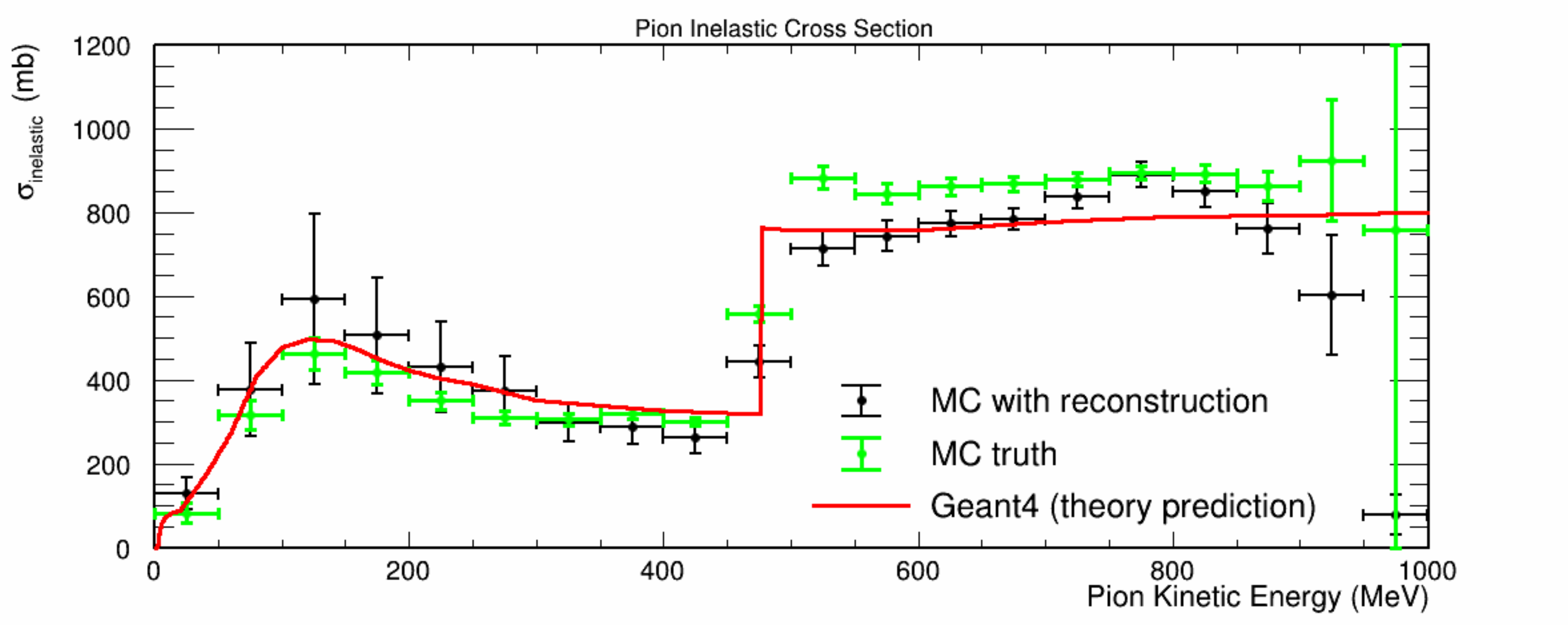
- True MC scale=[0.67, 1.22], Chi2/Ndf = 4.49229



- XS of true MC
- Reconstructed XS of fake data (unfolded by response matrix calculated by true MC)
- True XS of fake data

Iteration 2

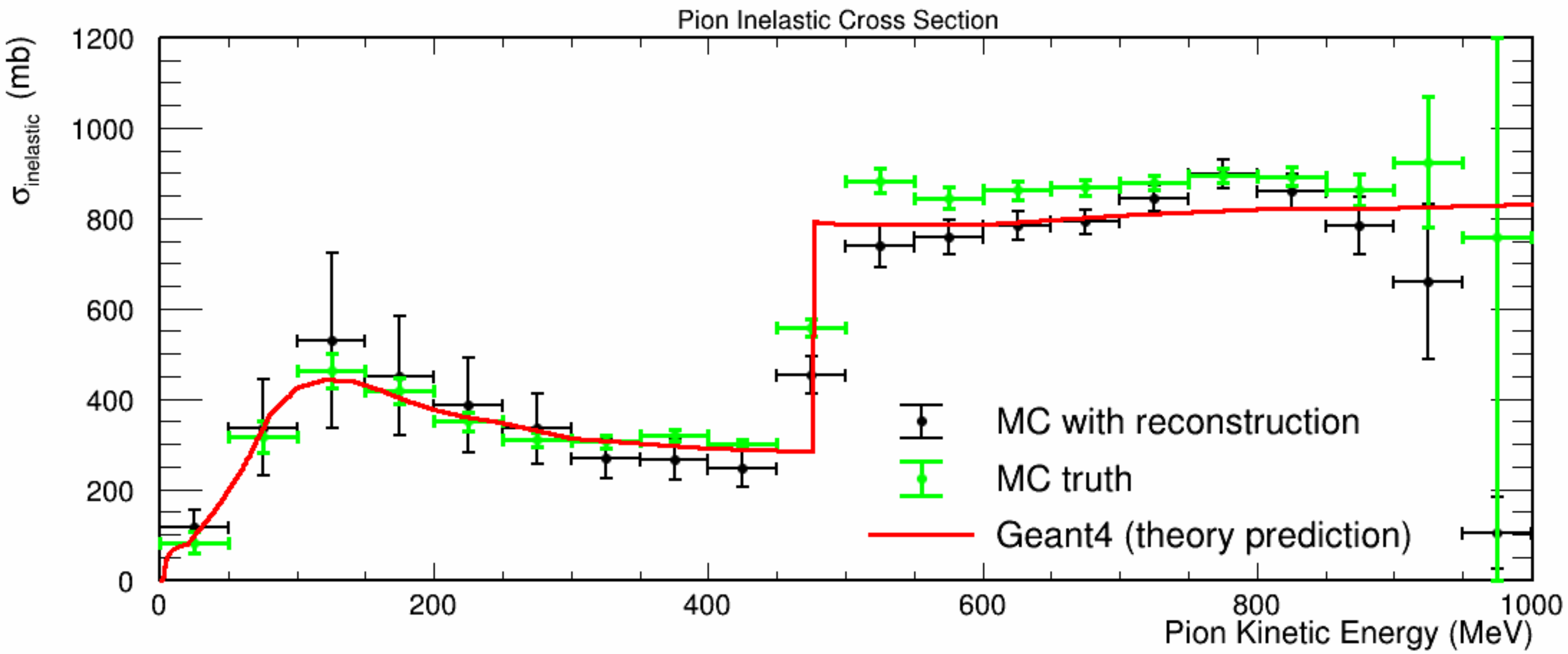
- True MC scale=[0.55, 1.32], Chi2/Ndf = 1.57103



- XS of true MC
- Reconstructed XS of fake data (unfolded by response matrix calculated by true MC)
- True XS of fake data

Iteration 3

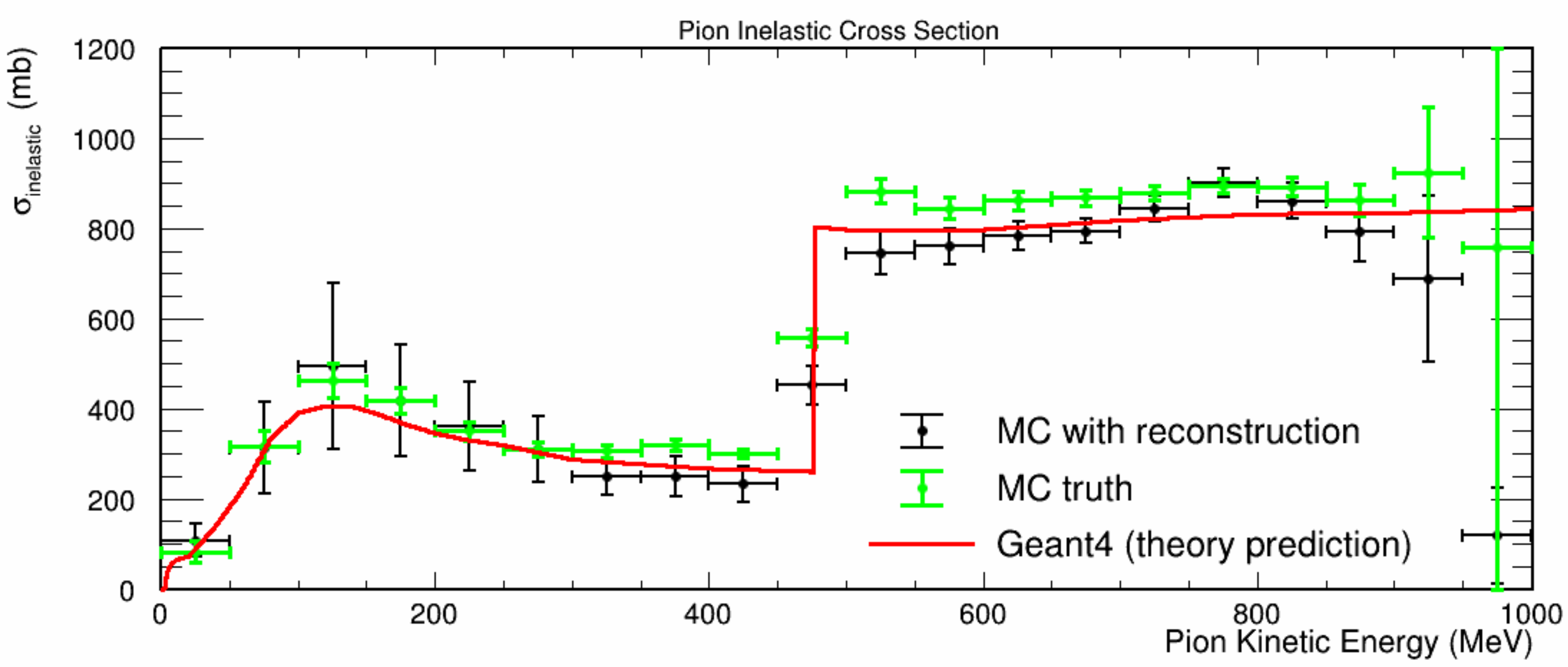
- True MC scale=[0.49, 1.37], Chi2/Ndf = 0.881462



- XS of true MC
- Reconstructed XS of fake data (unfolded by response matrix calculated by true MC)
- True XS of fake data

Iteration 4

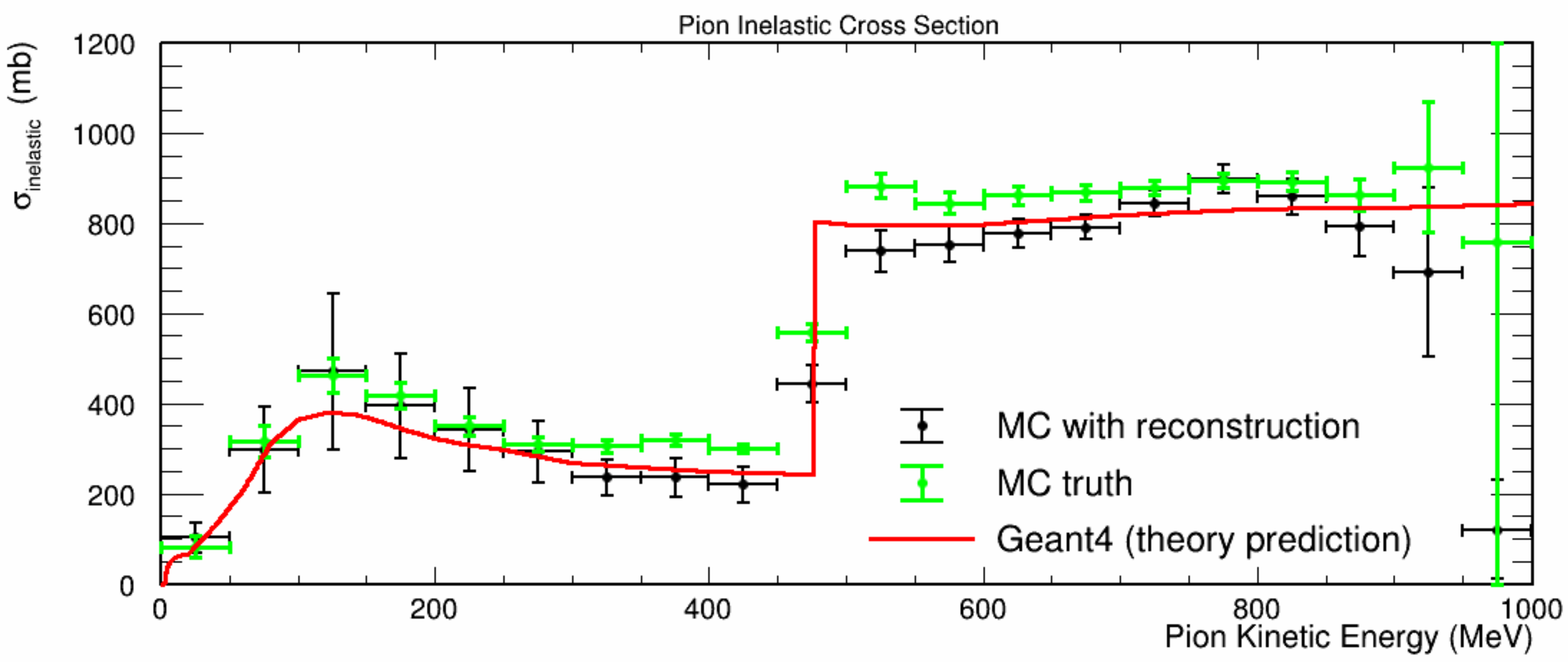
- True MC scale=[0.45, 1.39], Chi2/Ndf = 0.730346



- XS of true MC
- Reconstructed XS of fake data (unfolded by response matrix calculated by true MC)
- True XS of fake data

Iteration 5

- True MC scale=[0.42, 1.39], Chi2/Ndf = 0.755139



- XS of true MC
- Reconstructed XS of fake data (unfolded by response matrix calculated by true MC)
- True XS of fake data

Sum-up

- The iterative method seems to work, but it's possible to converge at some local best result.
  - Random jump could be helpful to find the global best result.

- We would like to request Geant4reweight to provide separate reweight factors every 50 MeV in KE.

Plan

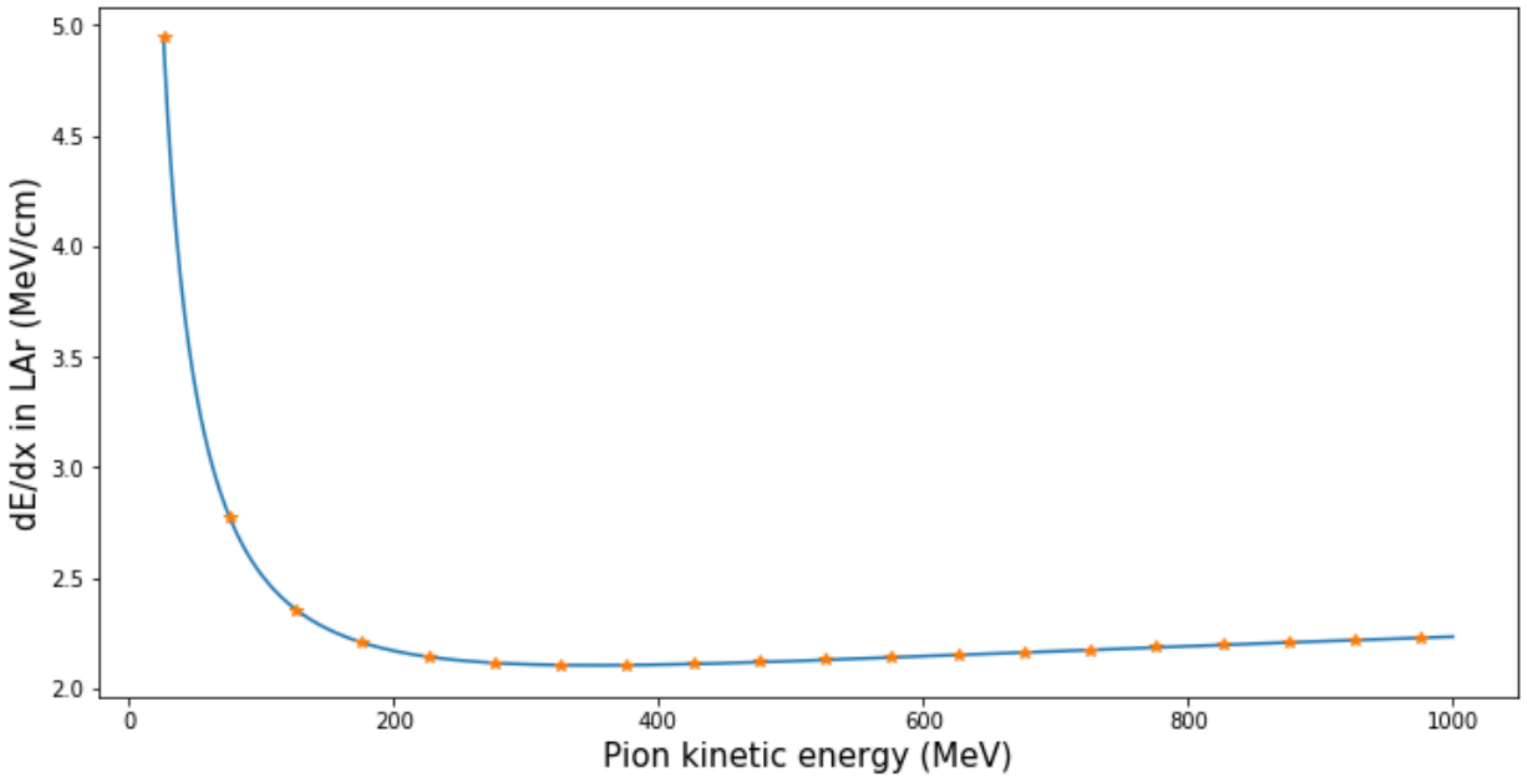
- We still seek better understanding of this “misfunction” of unfolding.
  - Suggestions are very welcome!

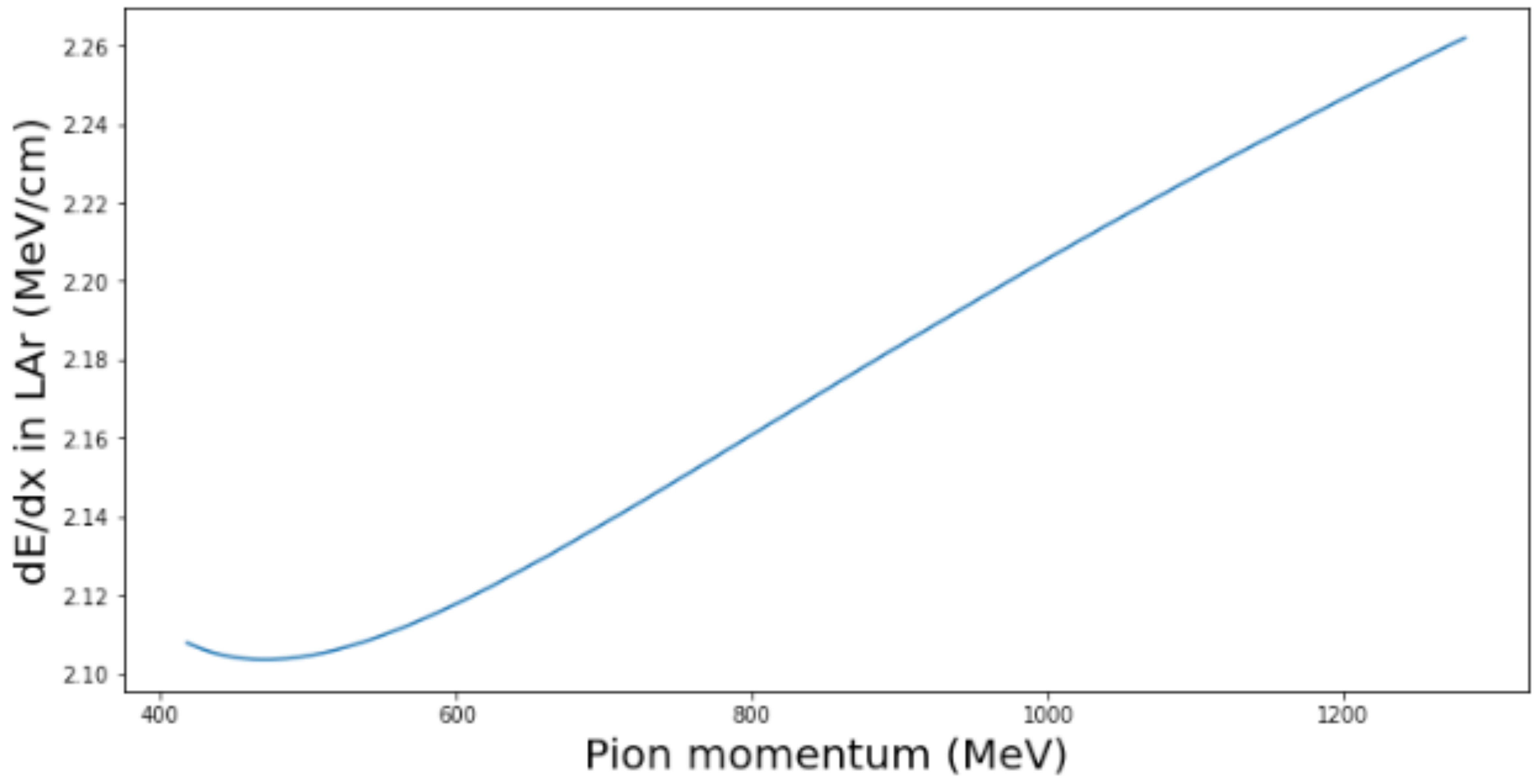
- If we can pass the test of the iterative method, we would like to proceed to apply it to real data soon.

Back-ups

dE/dx curve of pion in LAr

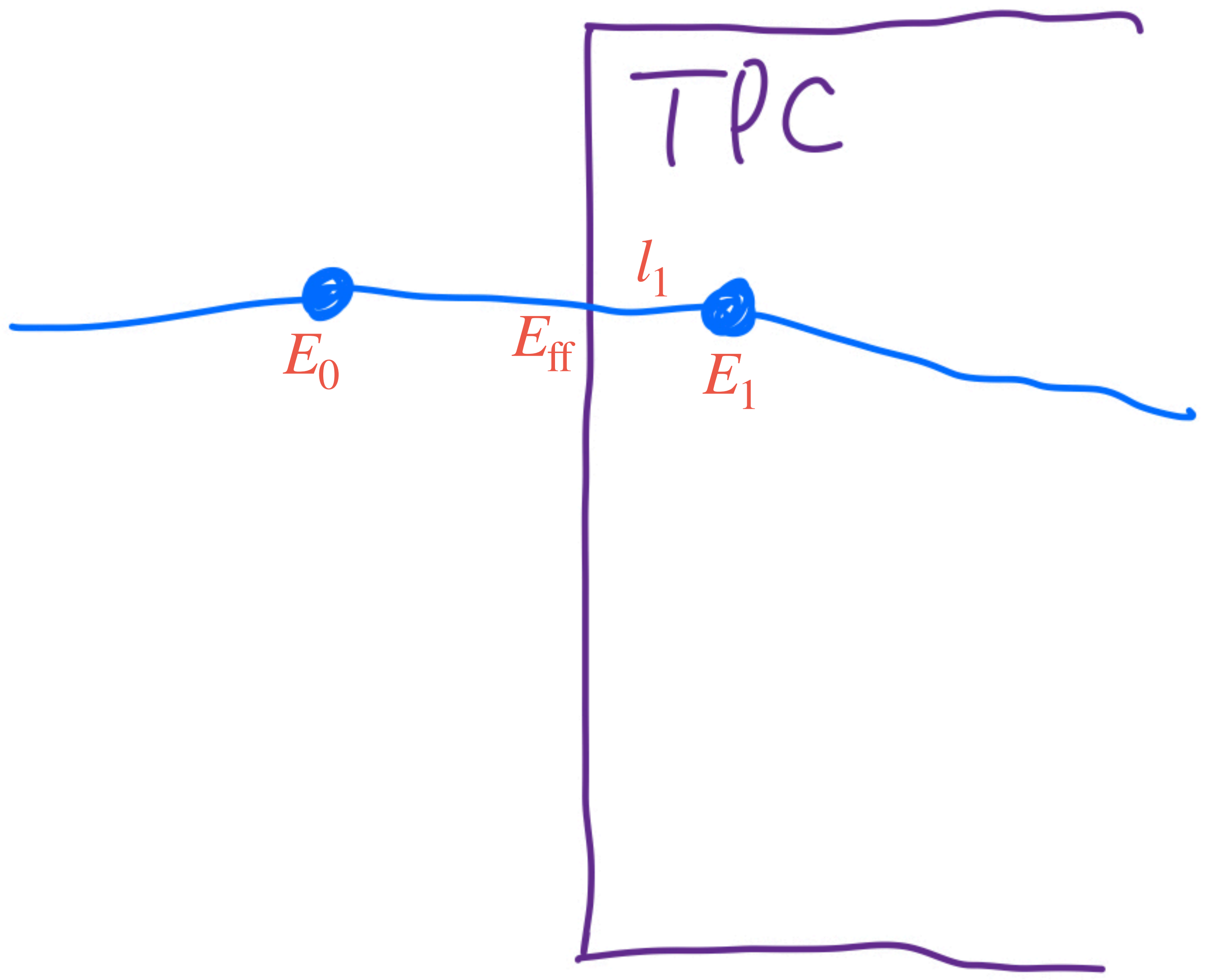
- In $\sigma = \frac{M_{\text{Ar}}}{\rho N_A \Delta E} \frac{dE}{dx}(E) \ln\left(\frac{N_{\text{inc}}(E)}{N_{\text{inc}}(E) - N_{\text{int}}(E)}\right)$, $\frac{dE}{dx}(E)$ is derived according to the Bethe-Bloch formula.



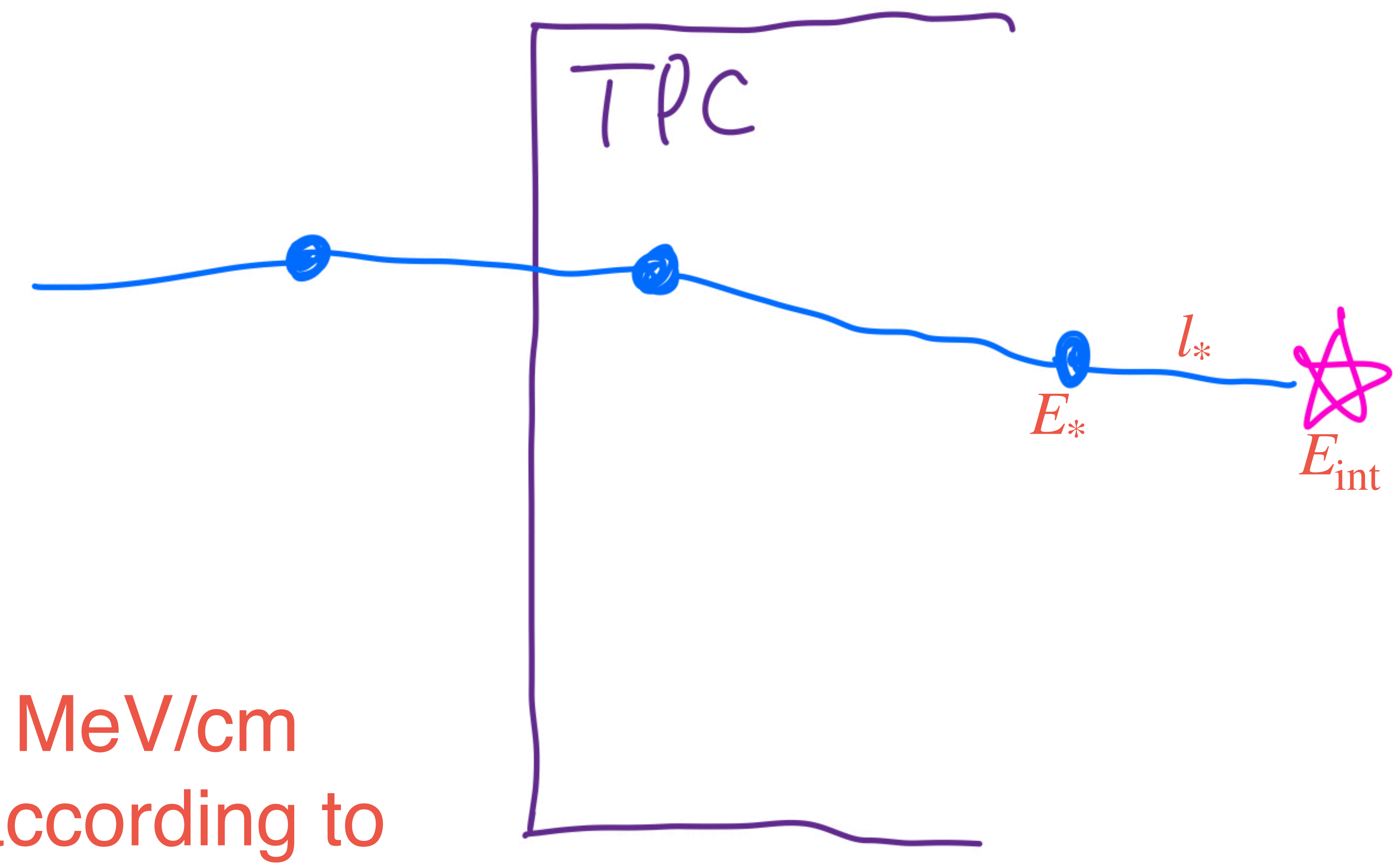


True $E_{\text{front-face}}$ and E_{int}

- For $E_{\text{front-face}}$
  - Previously, I estimate $E_{\text{ff}} = E_0$
  - Now, I use $E_{\text{ff}} = E_1 + 2.18 \cdot l_1$



- For E_{int}
  - I use trajectory point to estimate E_{int} with a small correction.
  - Estimate $E_{\text{int}} = E_* - 2.1 \cdot l_*$



2.18 MeV/cm and 2.1 MeV/cm are rough estimates according to dE/dx curve of pion in LAr

Compare instrumented beam KE and $E_{\text{front-face}}$

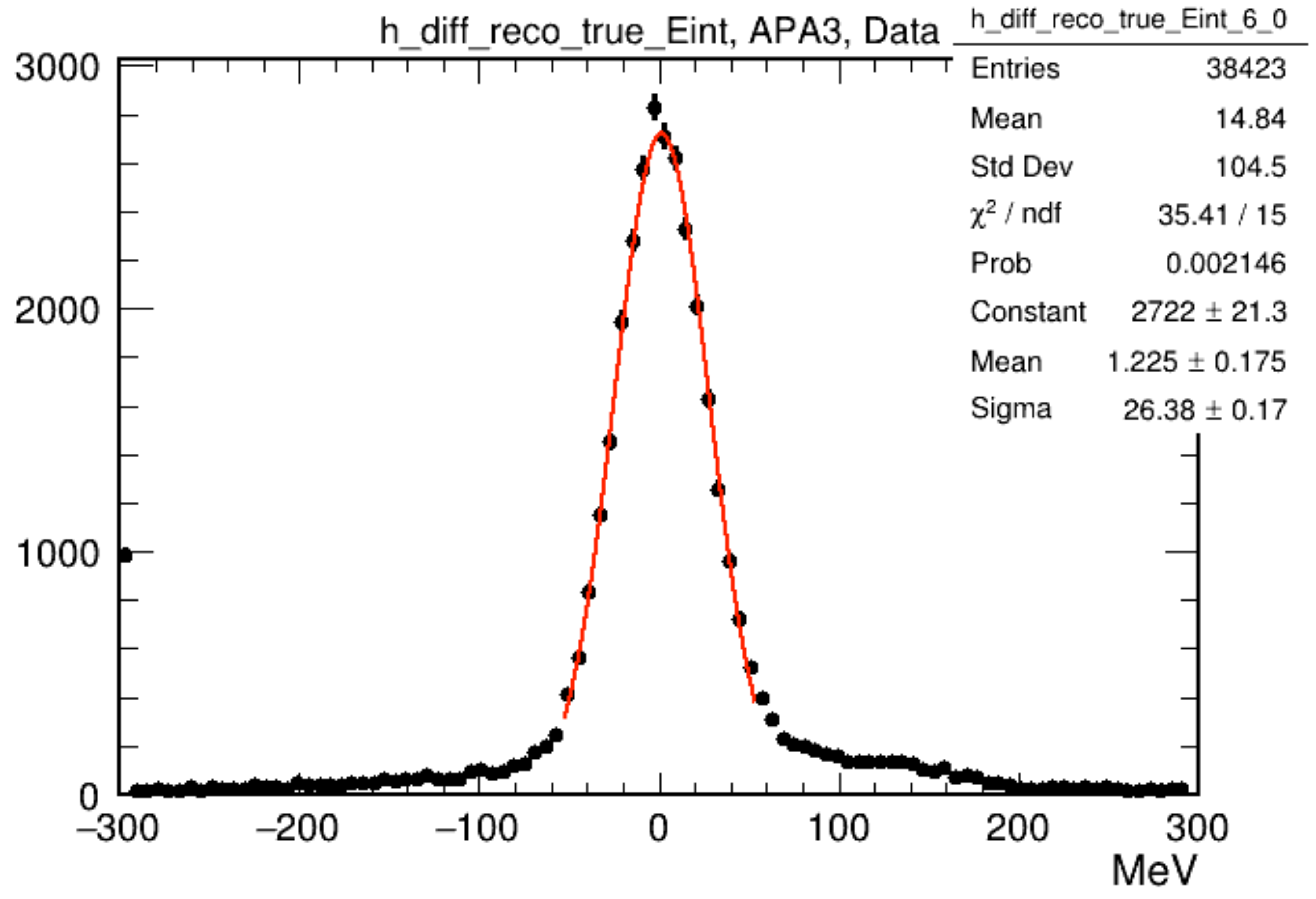
- Using true pion sample

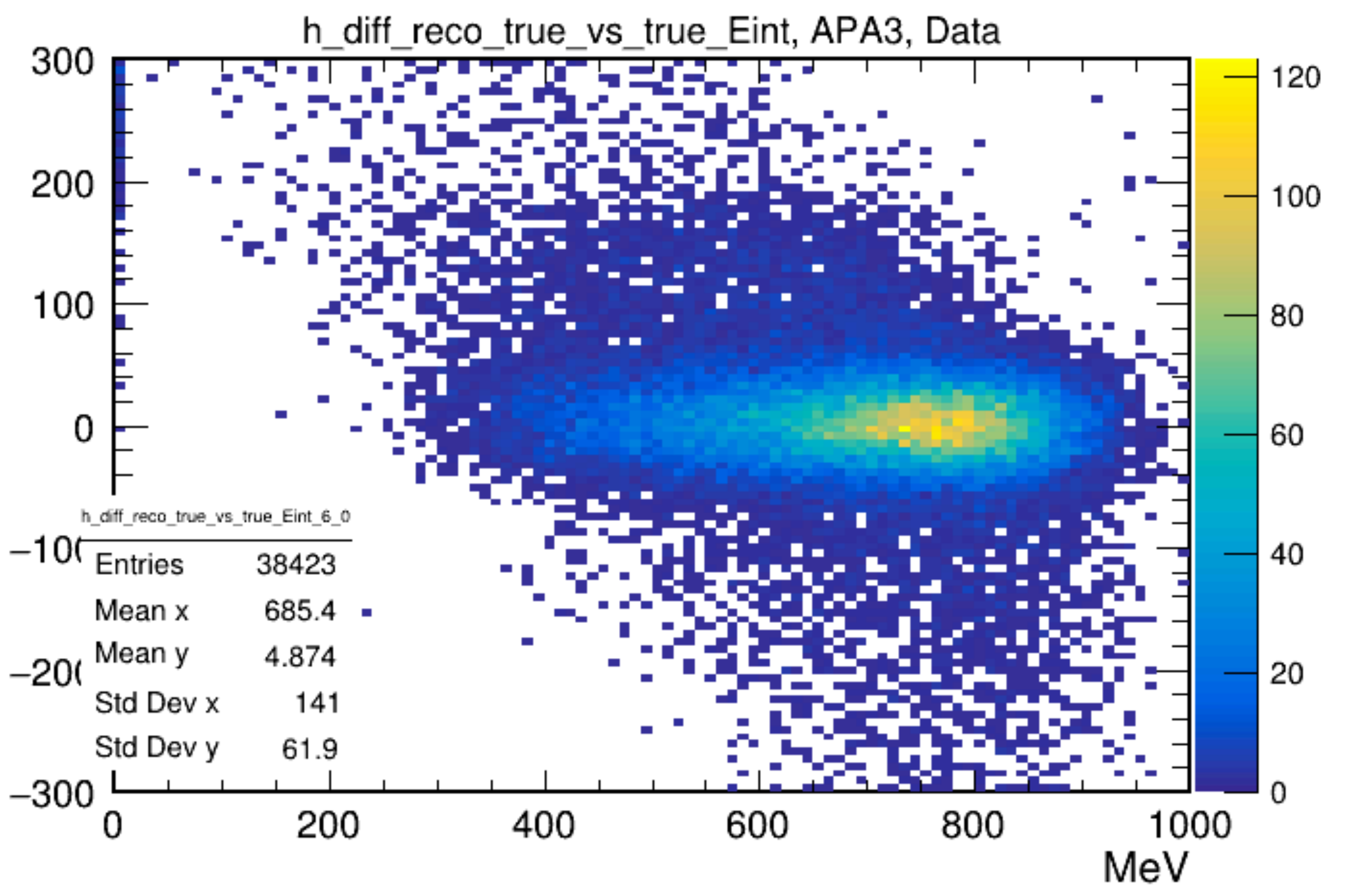
Reconstructed E_{ff} and E_{int}

- Reco $E_{\text{ff}} = E_{\text{inst}} - 12.74$ ← Upstream energy loss calculated using true pion sample

- Reco $E_{\text{int}} = E_{\text{inst}} - 12.74 - \int \frac{dE}{dx} dx_{\text{reco}}$

- Compare to true

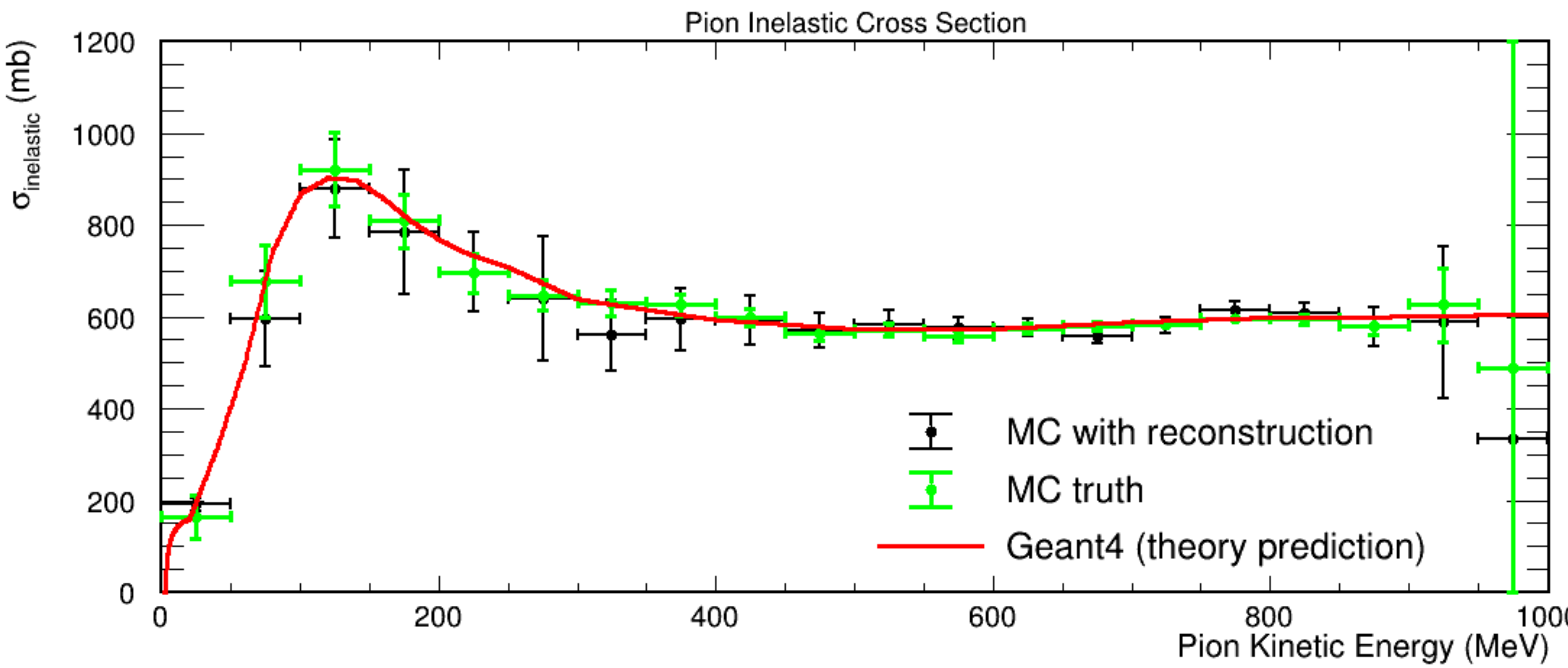




It suggests it is appropriate to choose energy interval as 50 MeV.

Ignore incomplete slice

- Interaction sliceID: $\text{int}\left(\frac{1000 - E_{\text{int}}}{\Delta E}\right)$ ➡ $\text{floor}\left(\frac{1000 - E_{\text{int}}}{\Delta E}\right)$
- Initial sliceID: $\text{int}\left(0.5 + \frac{1000 - E_{\text{front-face}}}{\Delta E}\right)$ ➡ $\text{ceil}\left(\frac{1000 - E_{\text{front-face}}}{\Delta E}\right)$



Ignore incomplete slices can help us get rid of the ambiguity of that confusing 0.5 in initial sliceID, which may cause small bias in the first couple of slices (on the right side)

Some notes about sliceID histograms

So for sliceID $i \in [0, N-1]$ (underflow -1; overflow N),

we have $N_{\text{inc}}(i) = \sum_{j=i}^{N} N_{\text{end}}(j) - \sum_{j=i+1}^{N} N_{\text{ini}}(j)$

Equivalently $N_{\text{inc}}(i) = \sum_{j=0}^{i} N_{\text{ini}}(j) - \sum_{j=0}^{i-1} N_{\text{end}}(j)$

because ideally $\sum_{i=0}^{N} N_{\text{ini}}(i) = \sum_{i=0}^{N} N_{\text{end}}(i)$, which is the total number of true signal events.

(This does not necessarily hold for reco histograms, because separate unfoldings on N_{end} and N_{ini} do not guarantee the total numbers still exactly equal to each other.)

Geant4reweight

Jacob Calcutt, et. al. https://arxiv.org/abs/2105.01744

- Divide a track into several steps.
- In each step, $P_{\text{survive}} = e^{-\sigma \Delta L}$, $P_{\text{interact}} = 1 - e^{-\sigma \Delta L}$.
- The probability that the particle survives after step f is $P_{\text{survive}} = \prod_{i}^{f} e^{-\sigma_i \Delta L_i}$.
- The probability that the particle interacts at step f is

$$P_{\text{interact}} = \prod_{i}^{f-1} e^{-\sigma_i \Delta L_i} \cdot \left(1 - e^{-\sigma_f \Delta L_f}\right).$$

Geant4reweight

Jacob Calcutt, et. al. https://arxiv.org/abs/2105.01744

- The probability that the particle interacts at step f is

$$P_{\text{interact}} = \prod_{i}^{f-1} e^{-\sigma_i \Delta L_i} \cdot \left(1 - e^{-\sigma_f \Delta L_f}\right).$$

- So we can express $P_{\text{interact}} = P_{\text{interact}}(\sigma)$.
- If we scale σ to $\alpha \cdot \sigma$, then $P_{\text{interact}} = P_{\text{interact}}(\alpha \cdot \sigma)$.
- Therefore, we assign a weight $w = \dfrac{P_{\text{interact}}(\alpha \cdot \sigma)}{P_{\text{interact}}(\sigma)}$ to each event, and thus we have a sample whose cross-section is equivalently $\alpha \cdot \sigma$.