Studying Geant4 Hadronic Model Parameters

Julia Yarba
Fermilab

HSF Detector Simulation WG Meeting
12 December 2022

DISCLAIMER

This is work in progress

Any results featured in this presentation are preliminary

General Information (I)

- Geant4 is the detector simulation toolkit used in High Energy Physics and in many other domains
- Geant4 employs a set of physics models to simulate interactions of particles with matter across a wide range of interaction energies
- These models, in particular the **hadronic** ones, rely largely on physically motivated parameters; but they generally aim to cover a wide range of possible simulation tasks and may not be optimized for a given process or material
- Starting release 10.4 and onwards, the Geant4 collaboration has been offering and gradually extending a configuration interface to the model parameters
  - https://geant4-userdoc.web.cern.ch/UsersGuides/ForToolkitDeveloper/html/GuideToExtendFunctionality/HadronicPhysics/hadronics.html#changing-internal-parameters-of-an-existing-hadronic-model
- This opens a possibility to fit simulated distributions to experimental datasets and extract optimal values of the model parameters and the associated uncertainties

General Information (II)

- The efforts started as a confluence of
  - Hadronic Physics Validation activity (benchmark of Geant4 simulation results vs a variety of thin target experimental datasets)
    - From time to time, we observed that what we considered as a model improvement gave good MC-data agreement in one area but turned as degradation in some other areas
    - We wanted to be able to give developers a bit more numeric feedback than "works in one corner, jams in two others"
  - Request from the user community (neutrino experiments) for making Geant4 parameters run time configurable so that one could explore the effect of varying such parameters on the simulated observables
    - With the aim to more accurately estimate the systematic errors in physics measurements (e.g., estimates of the neutrino flux) given the role of the detector simulation in performing the physics measurements

The "Exploratory" Phase of the Study

- Model developers offered an **initial** configuration interface to
  - Fritiof (FTF): from 3 GeV to ~1 TeV, hadron+nucleus interactions based on diffractive and non-diffractive quark-gluon string reactions and LUND string fragmentation
  - Bertini: from 0 (better say, 200 MeV) to 12 GeV, hadron+nucleus interactions where hadron-nucleon cross sections and region-dependent nucleon densities are used to sample path lengths of nucleons which follow the Fermi gas momentum distribution
  - PreCompound: below 200 MeV, provides transition from the kinetic stage of reaction to the equilibrium stage described by the de-excitation model
- The purpose was to explore (at first, internally) how close to (or far from) the experimental data the Geant4 predictions would move with varying specific parameters
  - Some parameters were found to be more "impactful" than others
- In cases where varying parameters resulted in substantial impact on the simulated observables, we made an initial attempt to optimize such parameters values via fits vs experimental data

From ”Exploratory” to Focus on Fritiof Model (FTF)

- Results of the “exploratory” phase are largely summarized
  - https://arxiv.org/abs/1910.06417 ; JINST 15 (2020) 02, P02025
  - Conclusions at the time:
    - Varying/optimizing parameters of the Geant4 models in the study generally leads to better agreement with some data
    - However, the number of parameters available at the time, per model, appeared to be too few in order to reach a better agreement across the board
- As the next step, we have decided to focus on one model, namely FTF, because the manpower is limited, and to expand the number of configurable parameters
- Why FTF:
  - Very popular as it is valid over a large range in energy (3 GeV to 1 TeV)
  - In active development
  - Plenty of support and advise from principal developer (Vladimir Uzhinsky, JINR)
- Obviously, we do have plans to extend work to other models as well

High Level View of FTF

- FTF model includes simulation of multiple (sub)processes, and each process can be described with the use of numeric parameters
- Current activity focuses on configurable FTF parameters involved in modeling the following processes
  - Nuclear target destruction
    - Important for modeling production of nucleons in hadron+nucleus interactions at intermediate energies (several GeV)
    - Has impact ("side effect" ?) on modeling production of pions in hadron+nucleus interactions at several GeV beam energy
  - Projectile or target diffraction dissociation
    - Quark exchange (without or with excitation of participants in this case)
      - Important for modeling pion production in hadron+nucleus interactions at intermediate-to-higher energies
- Obviously, FTF involves more elements; we plan to include those in future work

Datasets and Fitting Package

- Baryon (proton) projectile:
  - In the intermediate energy range (3-12 GeV), we are fortunate to have a reasonable collection of experimental data at least for such secondaries as pions, protons, and neutrons, in both forward and backward hemisphere, for a variety of nuclear targets
  - At higher energies (30-158 GeV), we have a number of datasets but they are mainly for light target (e.g., C); high energy data for heavier targets (Cu, Pb) are quite limited
- Pion projectile :
  - At intermediate energies we have a reasonable collection of data on pion production, for a variety of nuclear targets; however, the available data on nucleons production cover theta>~60deg; thus, it is hard to judge how we do in the forward hemisphere
  - At higher energies (60 GeV) we have recently obtained data on hadron production on light targets (e.g., Be or C; although only a subset of data on C have been used so far)
  - There are some limited data at 100 GeV on Cu, Pb
- Fitting Package: **Professor**
  - http://professor.hepforge.org
- Details in backup slides

FTF: Nuclear Destruction (from the Geant4 documentation)

The GEANT4 FTF model uses reggeon cascade in the impact parameter space to simulate production of fast nucleons in the hadron-nucleus interactions. After the projectile particle interacts with one of the nucleons in the target nucleus, this "wounded" nucleon may involve another nucleon in the cascade with the probability that is given as follows:

$$P(|\vec{s}_i - \vec{s}_j|) = C_{nd}\ exp[-(\vec{s}_i - \vec{s}_j)^2/R_c^2]$$

In this formula $\vec{s}_i$ and $\vec{s}_j$ are projections of the radii of *i*-th and *j*-th nucleons on the impact parameter plane, $R_c^2 = 1.5(fm)^2$, and the coefficient C_{nd} is defined as follows:

This is fixed (D) for baryons
but not for pions/mesons

$$C_{nd} = P_1\ e^{P_2\ (y-P_3)}/[1 + e^{P_2\ (y-P_3)}]$$

where y is the projectile rapidity. The parameter P_1 in the above formula can be a fixed value (DEFAULT), or it can be expressed as a function of

- baryon number of the projectile in the case of the projectile destruction
- number of nucleons in the target nucleus in case of the target destruction

Modeling of momentum distributions of the nucleons involved in the cascade is described in greater details later in this document; however, one of the characteristics we would like to mention here is the average transverse momentum squared which can be expressed in a parametric way:

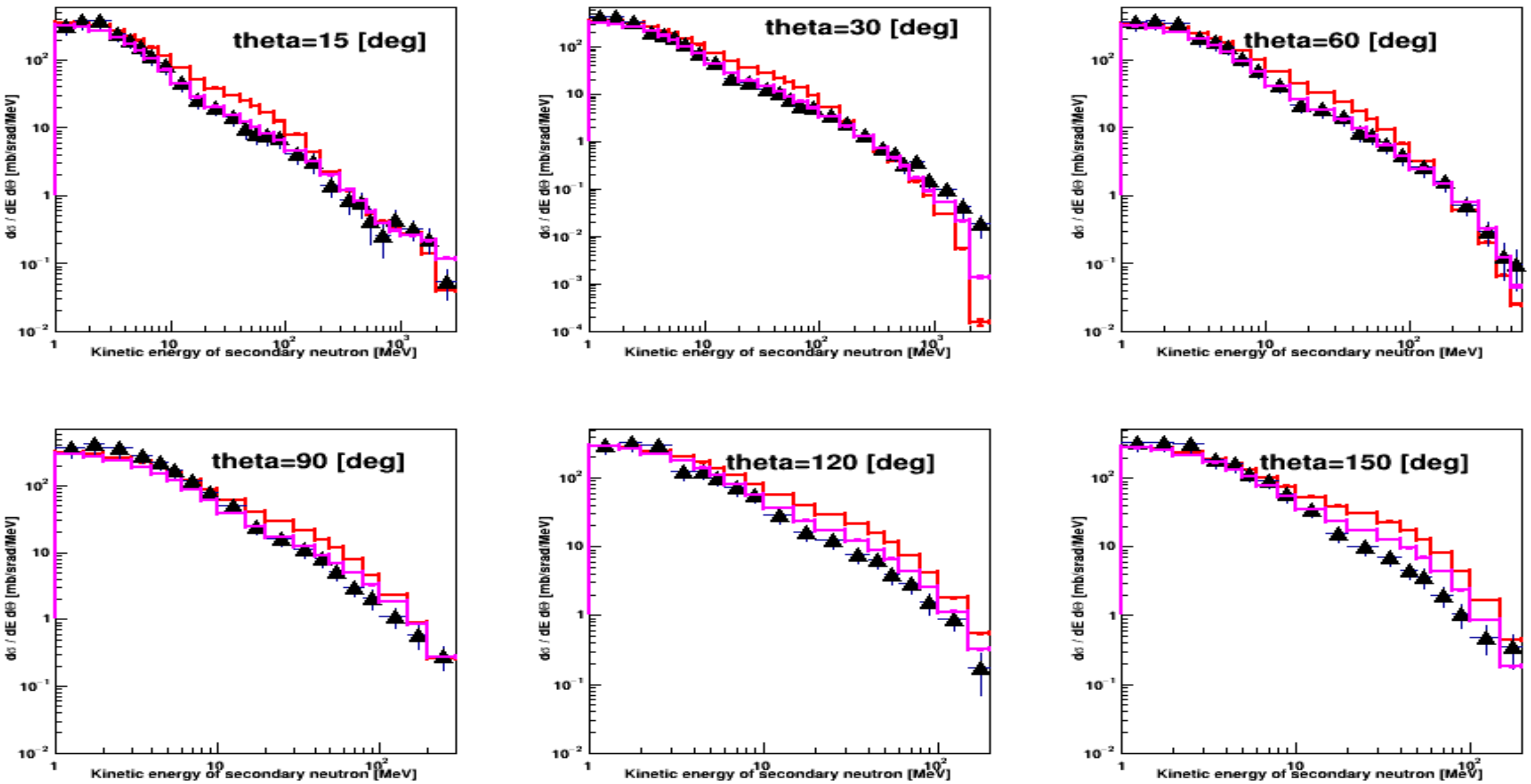
$$\langle P_T^2 \rangle = C_1 + C_2 \frac{e^{C_3\ (y_l ab - C_4)}}{1. + e^{C_3\ (y_l ab - C_4)}} \quad [(GeV/c)^2]$$

Selected Results

- Subsequent slides show selected results that exemplify how the parameter values extracted via fitting vs experimental data may bring MonteCarlo closer to the data
  - **The red color in the distributions represent results simulated with Geant4/FTF with default settings of parameters**
  - **The magenta color in the distribution show results simulated with Geant4/FTF with the best fit values of selected parameters**
  - **The black triangles represent experimental data**

Nuclear Target Destruction, Proton Projectile (I)

G4/FTF: 3.824GeV proton on Pb → neutron + X; data by IAEA (black triangles)

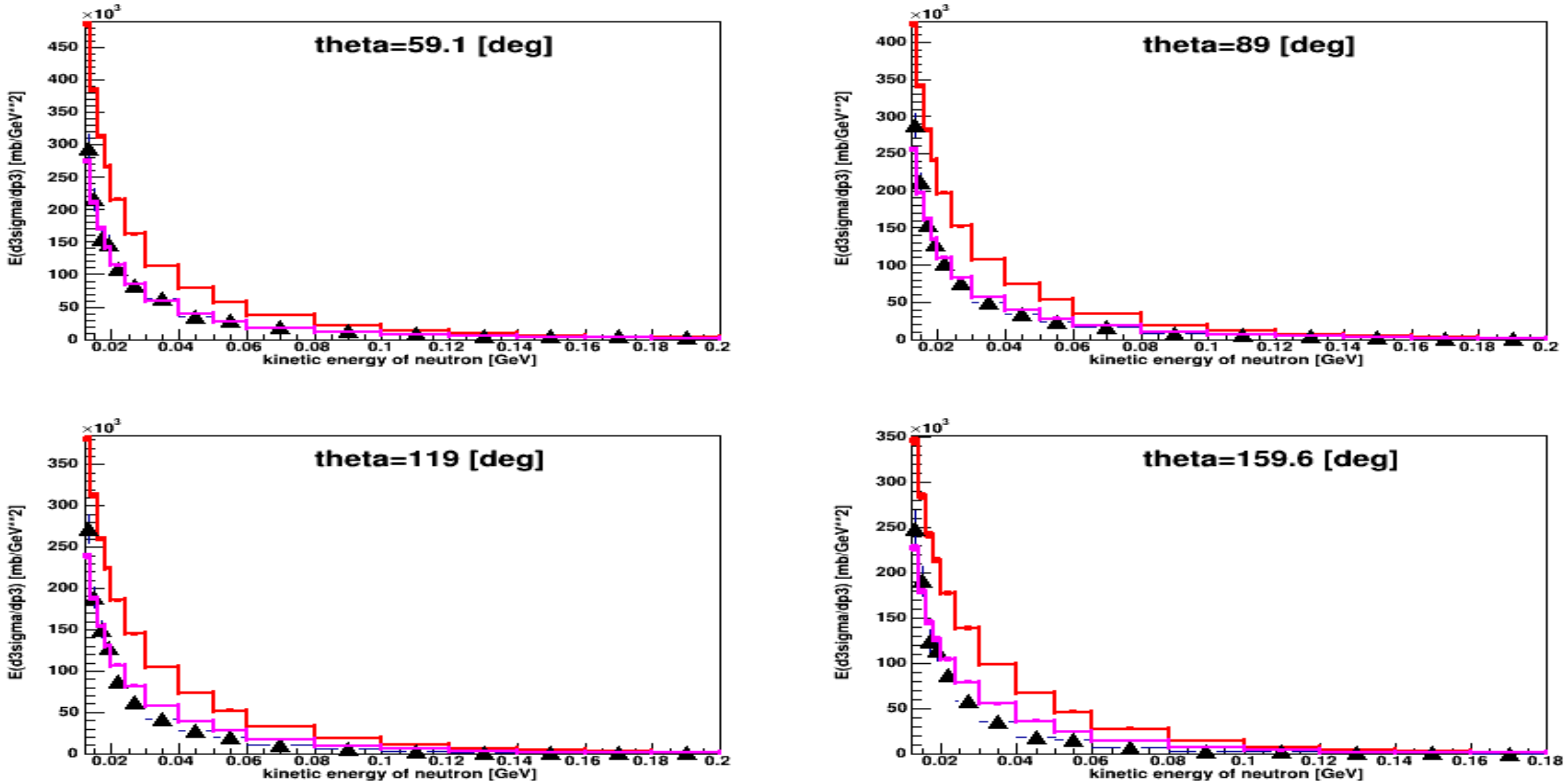


Sim with G4/FTF Default : chi2/NDF=10.3 Sim with Best Fit : chi2/NDF=1.3

Using best fit values of parameters in simulation brings MC closer to the data in both forward and backward hemisphere

Nuclear Target Destruction, Proton Projectile (II)

G4/FTF: 7.5GeV proton on Pb → neutron + X; data by ITEP (black triangles)

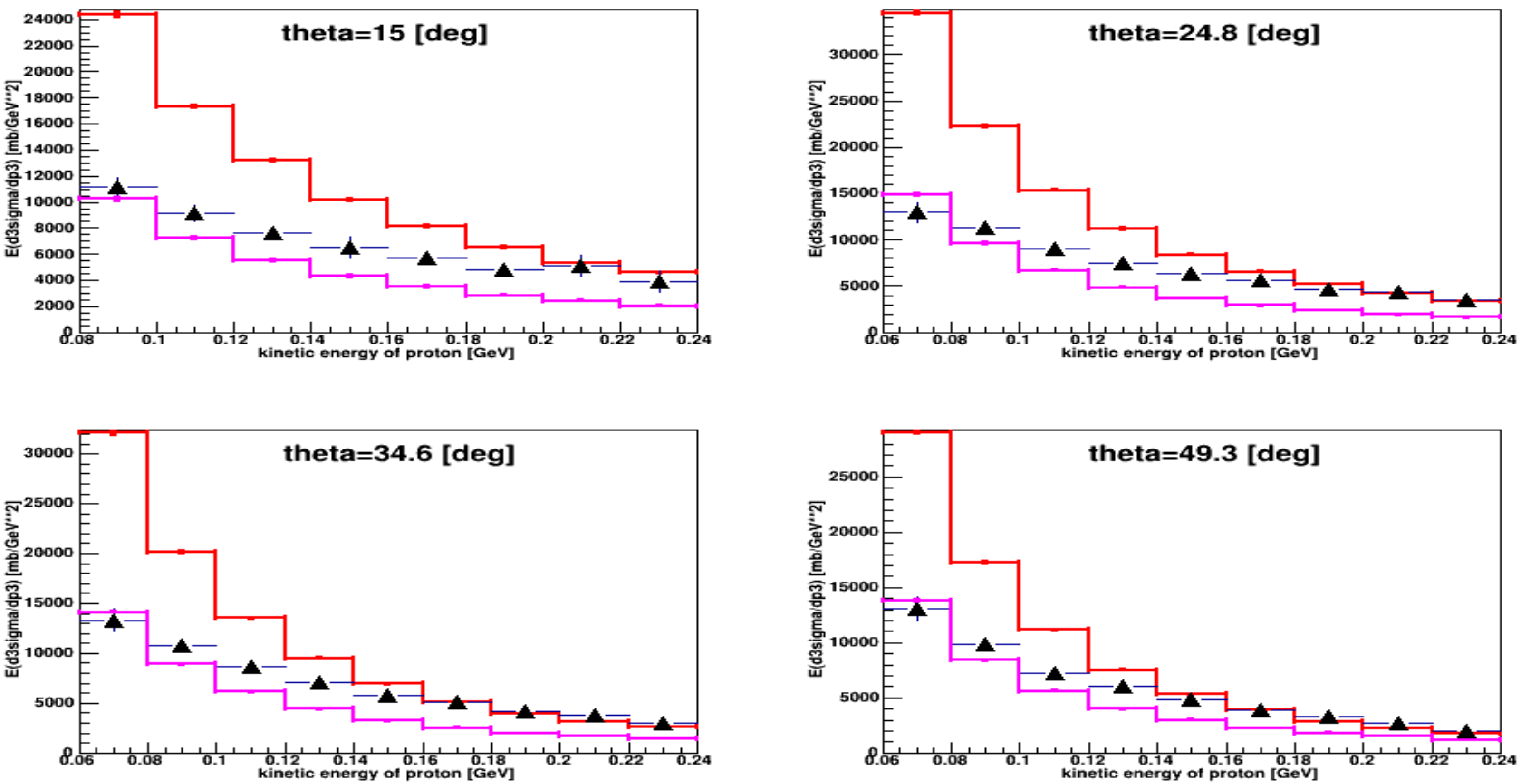


Sim with G4/FTF Default : chi2/NDF=224.6 **Sim using Best Fit : chi2/NDF=12.5**

Using best fit values of parameters in simulation brings MC closer to the data but with this dataset we can only judge at theta>60deg; fortunately, it is complemented by other data on neutron production in proton+neucleus interactions (slides 10)

Nuclear Target Destruction, Proton Projectile (III)

G4/FTF: 7.5GeV proton on Pb → proton + X; data by ITEP (black triangles)

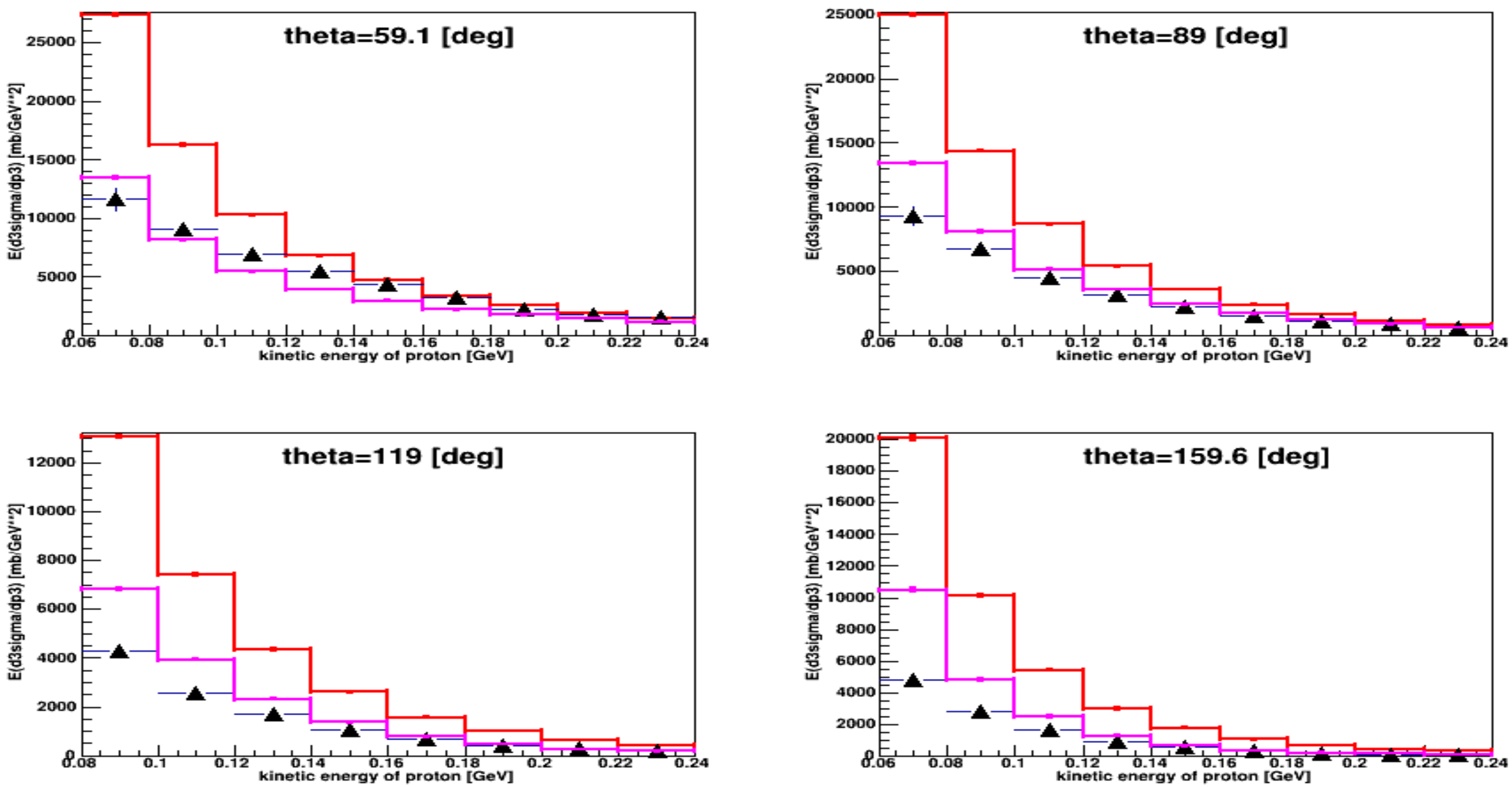


Sim with G4/FTF Default : chi2/NDF=68.4 **Sim using Best Fit : chi2/NDF=22.7**

Using best fit values of parameters in simulation brings MC closer to the data (forward hemisphere)

Nuclear Target Destruction, Proton Projectile (IV)

G4/FTF: 7.5GeV proton on Pb → proton + X; data by ITEP (black triangles)



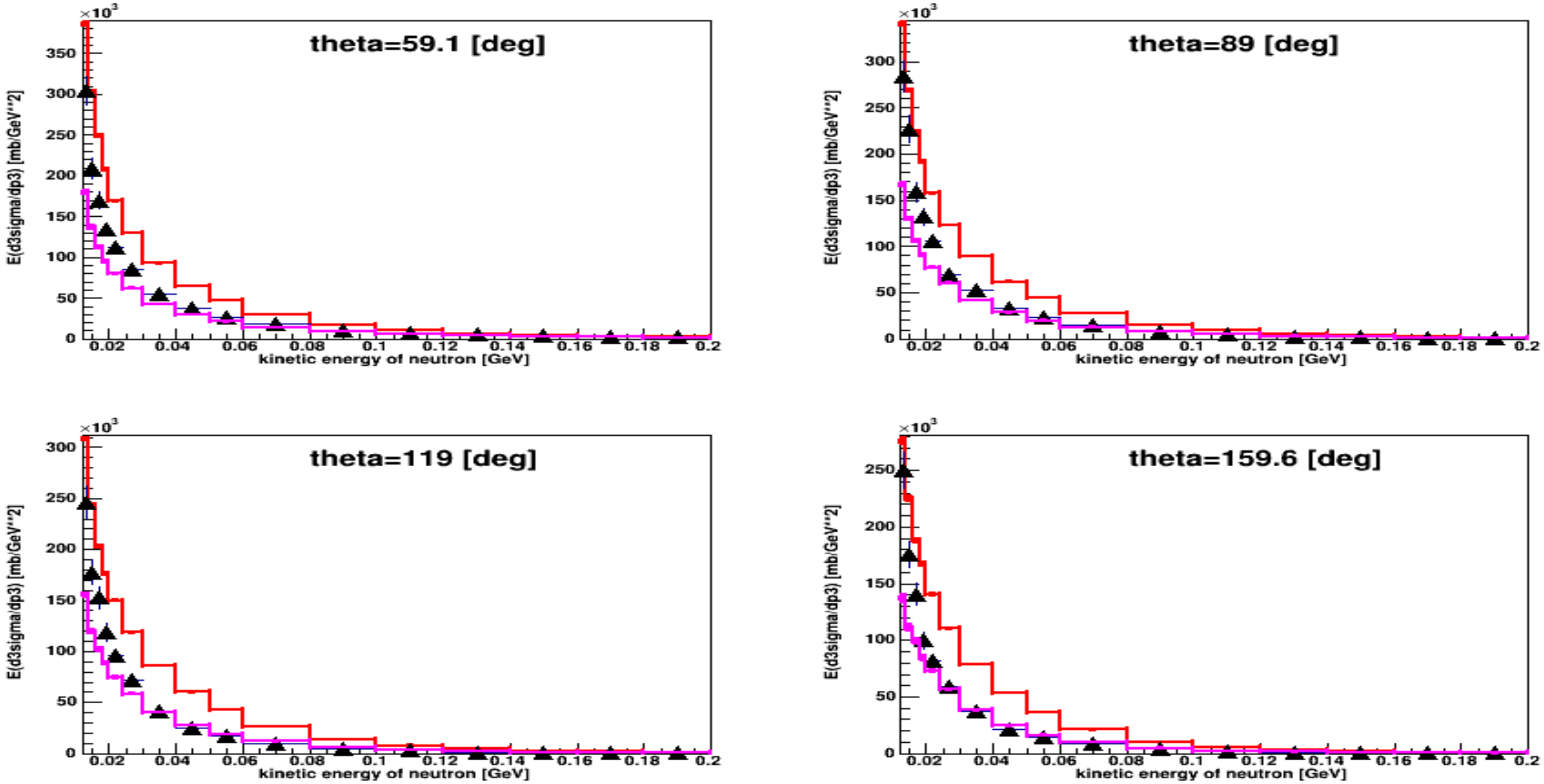
Sim with G4?FTF Default : chi2/NDF=308 **Sim using Best Fit : chi2/NDF=21.6**

Using best fit values of parameters in simulation brings MC closer to the data (backward hemisphere)

Fermilab

Nuclear Target Destruction, Pion Projectile (I)

G4/FTF: 5.0GeV piminus on Pb → neutron + X; data by ITEP (black triangles)

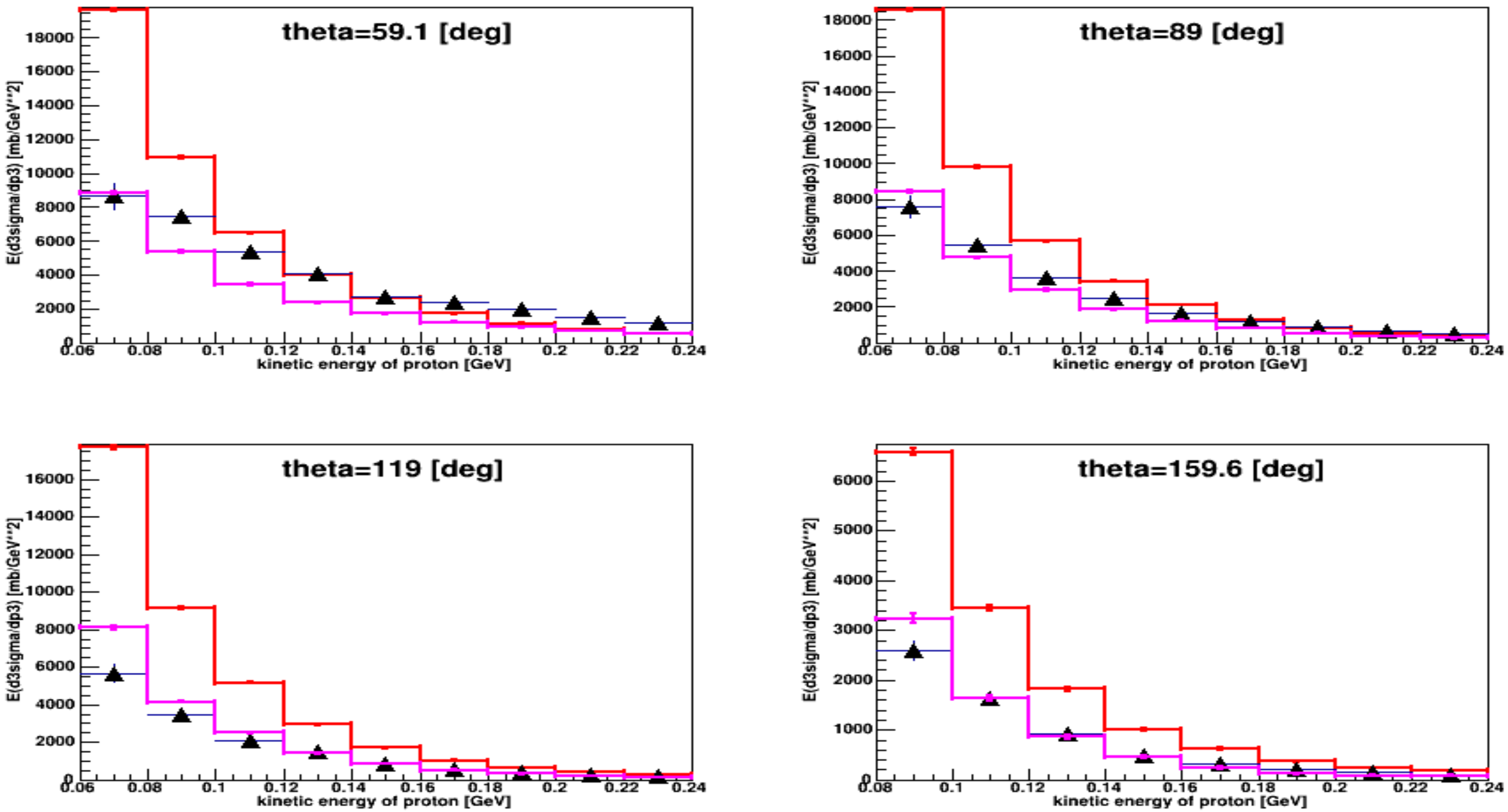


Sim with G4/FTF Default: chi2/NDF = 123.9 **Sim using Best Fit: chi2/NDF = 12.9**

We can judge the improvement in the MC-to-data agreement only at theta>60deg; we can not (yet) fully attest the situation in the forward direction

Nuclear Target Destruction, Pion Projectile (II)

G4/FTF: 5.0GeV piminus on Pb → proton + X; data by ITEP (black triangles)

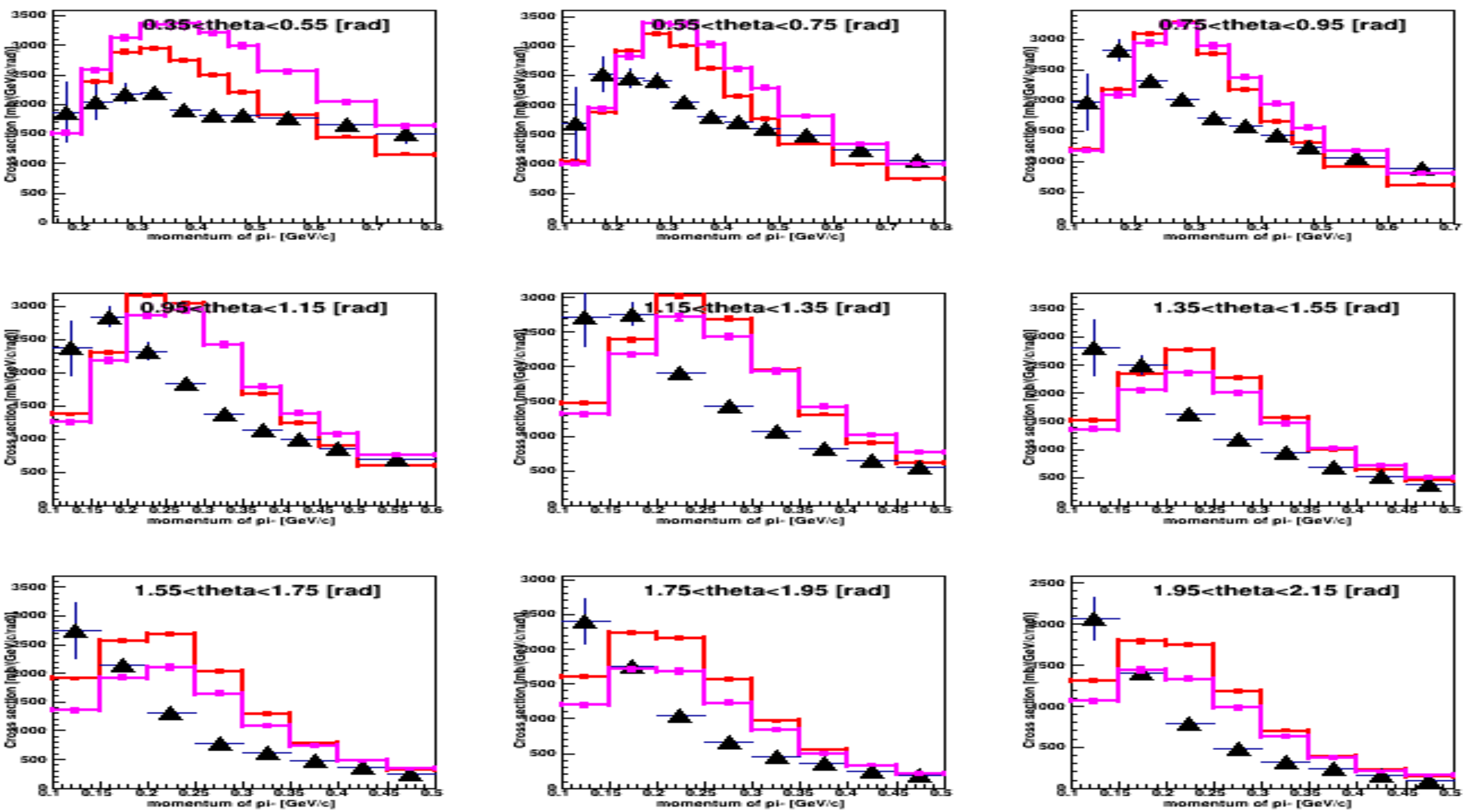


Sim with G4/FTF Default : chi2/NDF = 108.2 **Sim using Best Fit : chi2/NDF = 13.6**

We can judge the improvement in the MC-to-data agreement only at theta>60deg; we can not (yet) fully attest the situation in the forward direction

Nuclear Target Destruction, Pion Projectile (III)

G4/FTF: 5.0GeV piminus on Pb → piminus + X; data by HARP (black triangles)



Sim with G4/FTF Default : chi2/NDF = 66.6 **Sim using Best Fit : chi2/NDF = 55.5**

Using best fit values of parameters in simulation may cause "side effects" in modeling pion production in pion+nucleus interactions; we are exploring how to improve the agreement

FTF: Quark Exchange (from the Geant4 documentation)

The original Fritiof model contains only the pomeron exchange process shown in Fig. 44*(d)*. It would be useful to extend the model by adding the exchange processes shown in Fig. 44*(b)* and Fig. 44*(c)*, and the annihilation process of Fig. 44*(a)*. This could probably be done by introducing a restricted set of mesonic and baryonic resonances and a corresponding set of parameters. This procedure was employed in the binary cascade model of GEANT4 (BIC) [BIC] and in the Ultra-Relativistic-Quantum-Molecular-Dynamic model (UrQMD) [UrQMD1], [UrQMD2]. However, it is complicated to use this solution for the simulation of hadron-nucleus and nucleus-nucleus interactions. The problem is that one has to consider resonance propagation in the nuclear medium and take into account their possible decays which enormously increases computing time. Thus, in the current version of the FTF model only quark exchange processes have been added to account for meson and baryon interactions with nucleons, without considering resonance propagation and decay. This is a reasonable hypothesis at sufficiently high energies.

For each projectile hadrons the following probabilities are set up:

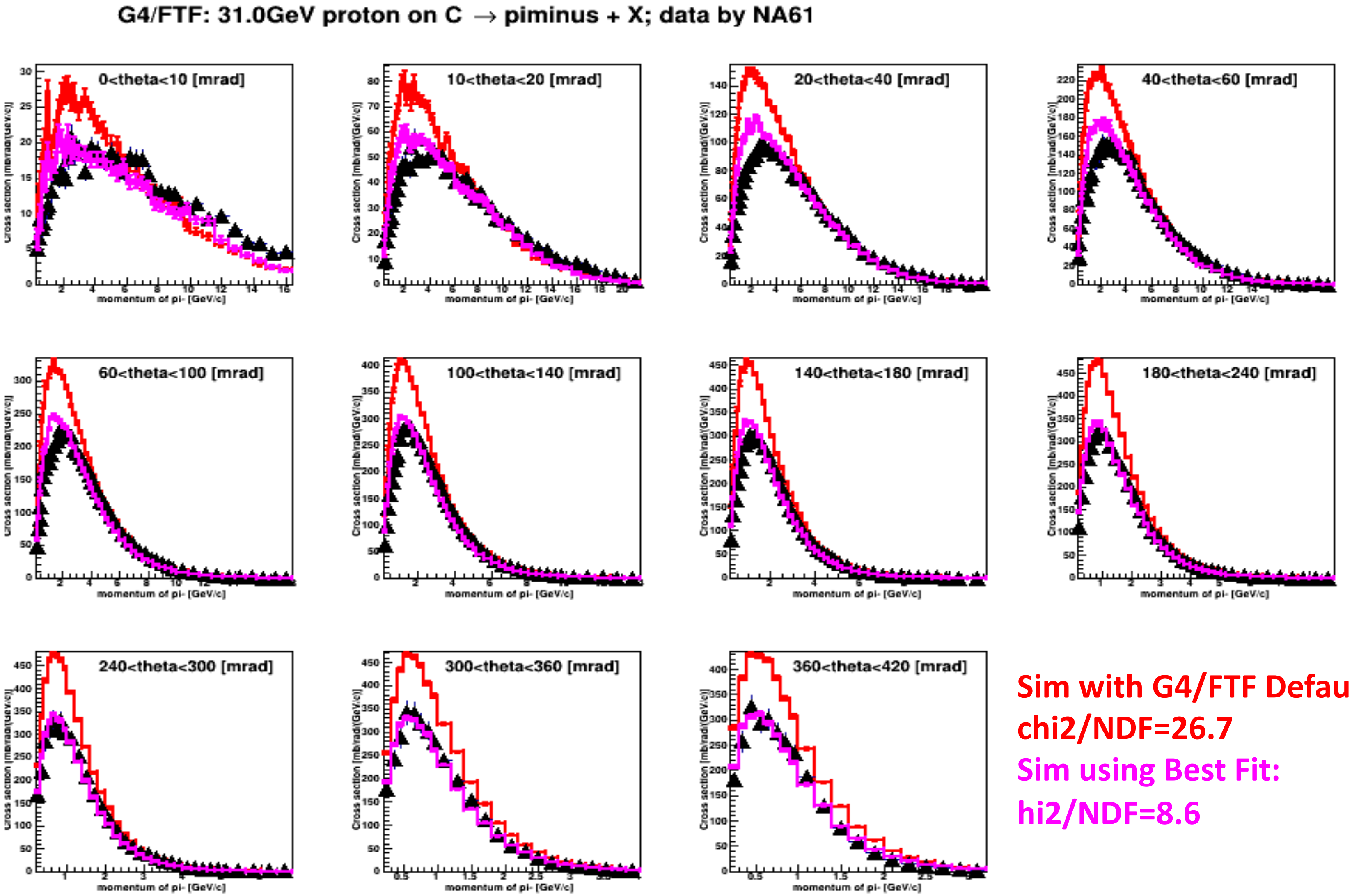
- Probability of quark exchange process without excitation of participants (Fig. 44*(b)*); *(Proc# 0)*
- Probability of quark exchange process with excitation of participants (Fig. 44*(c)*); *(Proc# 1)*
- Probability of projectile diffraction dissociation; *(Proc# 2)*
- Probability of target diffraction dissociation. *(Proc# 3)*

All these probabilities have the same functional form:

$$P_p = A_1\, e^{-B_1\, y} + A_2\, e^{-B_2\, y} + A_3,$$

where y is the projectile rapidity in the target rest frame.

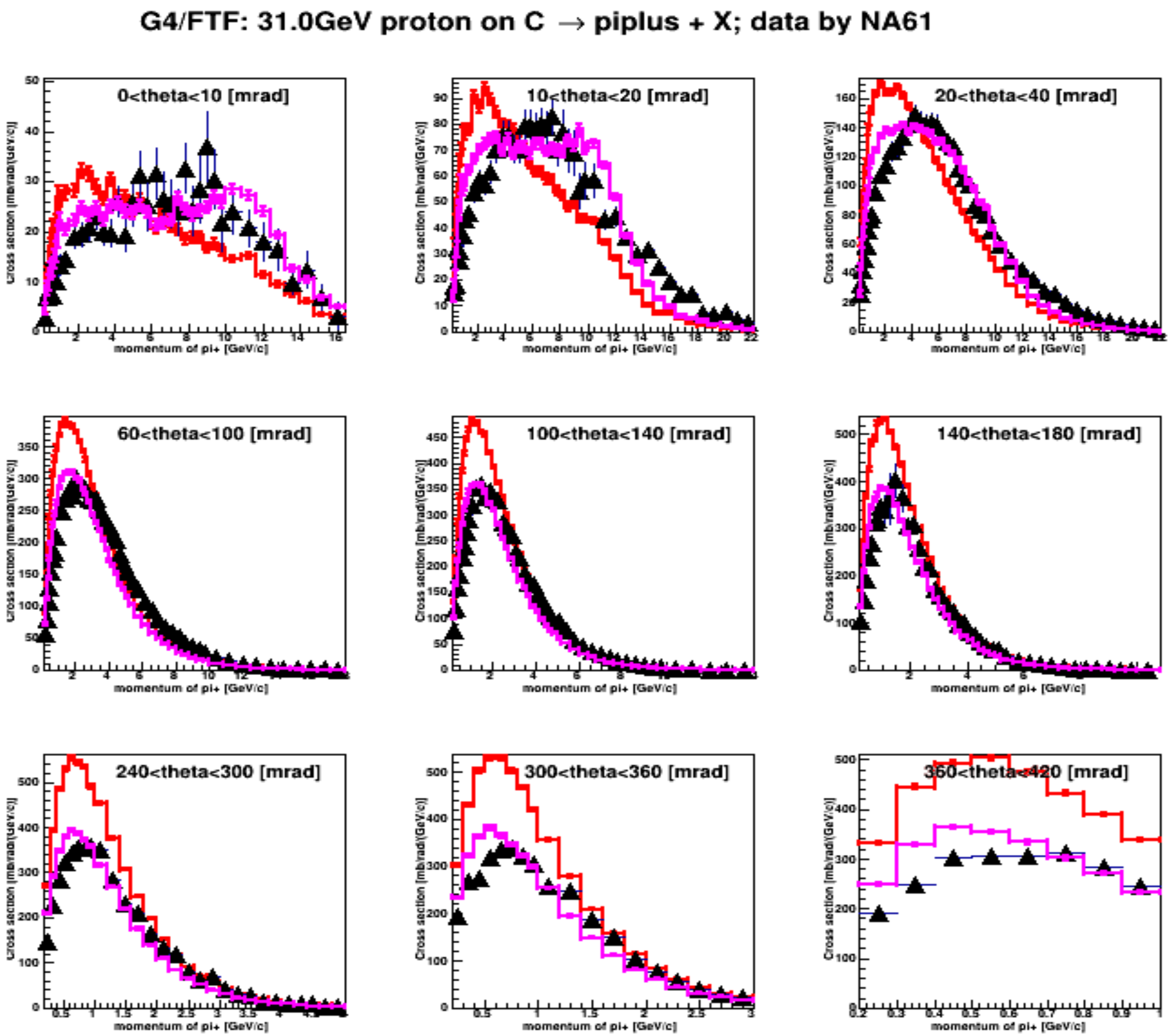
Quark Exchange with Excitation, Proton Projectile (I)



Sim with G4/FTF Default: chi2/NDF=26.7

Sim using Best Fit: hi2/NDF=8.6

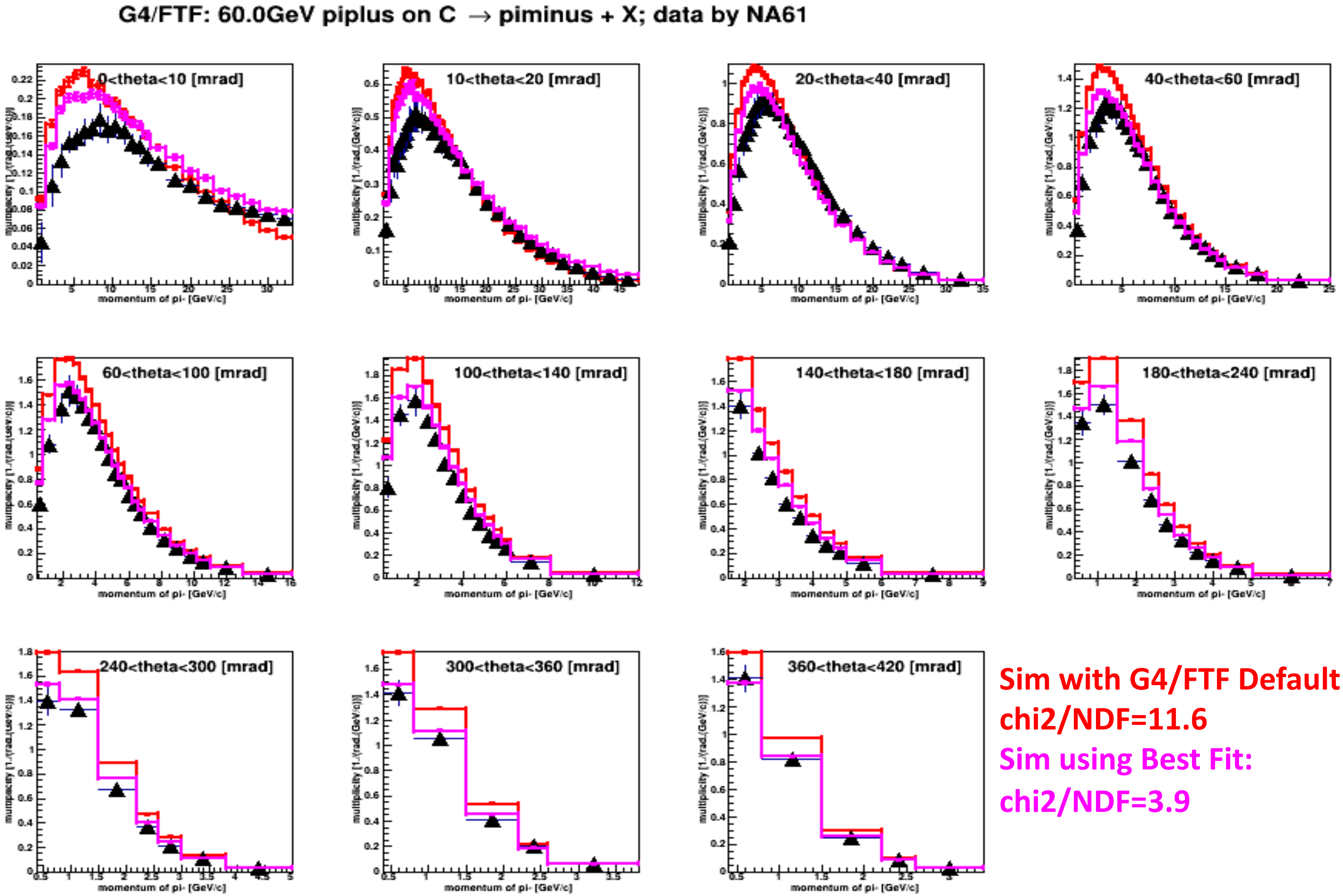
Quark Exchange with Excitation, Proton Projectile (II)



Sim with G4/FTF Default: chi2/NDF=31.1

Sim using Best Fit: chi2/NDF=16.6

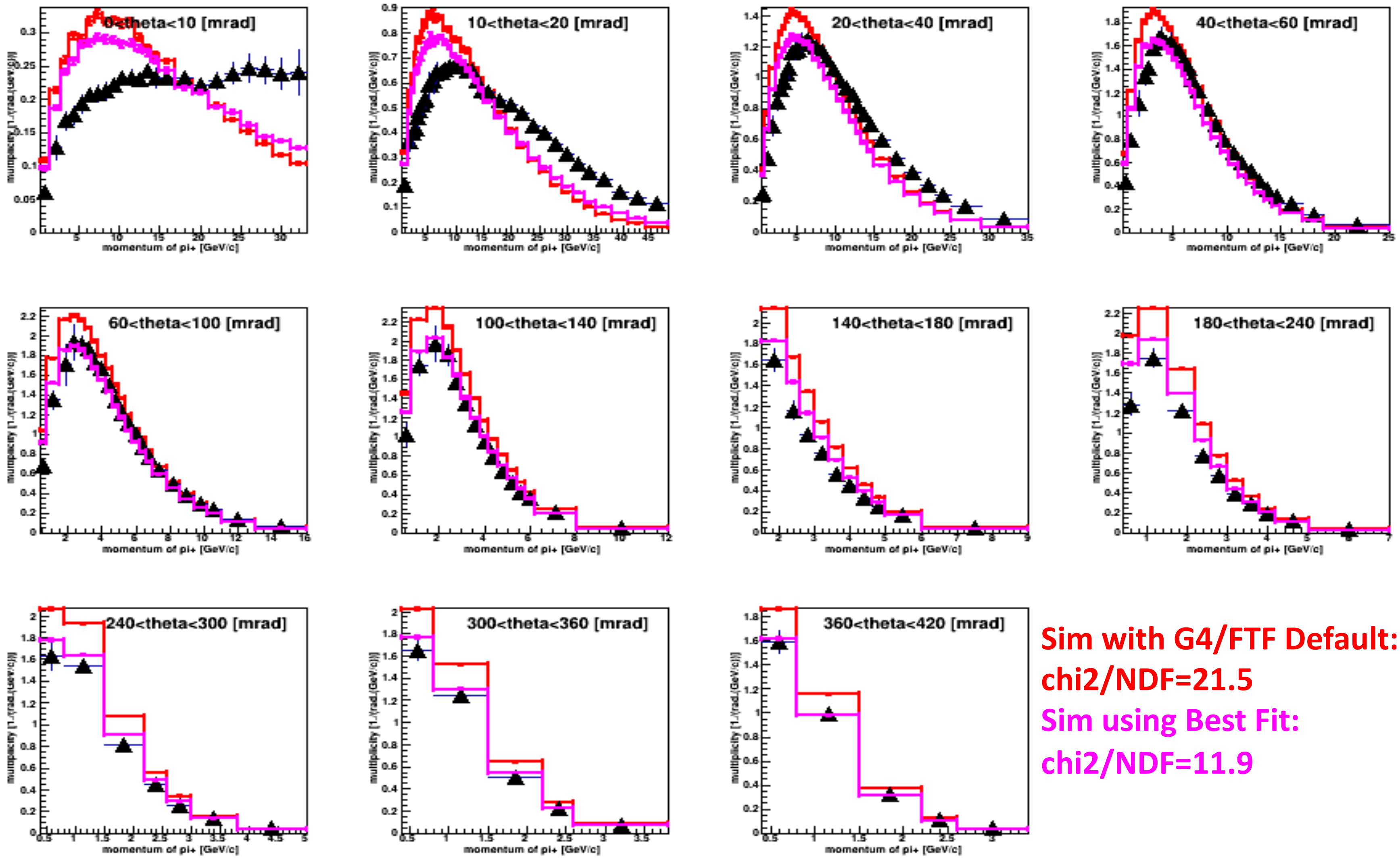
Quark Exchange with Excitation, Pion Projectile (I)



Sim with G4/FTF Default:
chi2/NDF=11.6
Sim using Best Fit:
chi2/NDF=3.9

Quark Exchange with Excitation, Pion Projectile (II)

G4/FTF: 60.0GeV piplus on C $\rightarrow$ piplus + X; data by NA61



Sim with G4/FTF Default: chi2/NDF=21.5

Sim using Best Fit: chi2/NDF=11.9

The Idea of Introducing Tunes in Geant4

- Although the work remains largely in progress, we have decided to introduce a possibility to set alternative (as compared to defaults) group of selected parameters, aka tunes, similar to e.g., Pythia8
- **Preliminary** tunes for FTF become available in release 11.1 and reflect the two ongoing study cases, for baryon or pion projectile
  - In the future - for kaons, hyperons, etc.
  - In the future – maybe also tunes for different energy ranges
- At present
  - The feature is meant for internal tests and further study/development
  - The required infrastructure is in the early phase of development
- In the future
  - Similar tunes can be introduced for other Geant4 Hadronic models
  - When properly mature, tunes may be offered to users for certain applications

Challenges : Experimental Data for Fits/Benchmark

- We are fortunate to be able to employ an extensive collection of experimental data
- However, it has certain limitations and leaves quite a few corners of the phase space uncovered
- Of course, there are more data out there that we need to get more familiar with
  - Suggestions on this matter are most welcome !
- We are also looking forward to the new data, e.g., recent data from NA61 (we are currently incorporating some) and/or upcoming data from EMPHATIC (we are in communication with the team)
- However, even if one has the richest collection of experimental data at hands, selecting the sets that are right/best for specific studies is a challenging task

Challenges: Fitting Tool(s)

- We have been using Professor tuning package, and we appreciate that with it we were able to obtain some encouraging results
- However, development of Professor stopped some time ago, and at present the package is out of support
- We are lucky to be still able to use it but in the longer run it may come to an end
- We are exploring what is new on the market that can do similar (or better ?) job
- We are testing Apprentice tuning package (https://arxiv.org/pdf/2103.05748.pdf)
  - We have shared our feedback with the developers, and we are in communication with the them regarding our observations
- We are also aware that there is at least one ML-based tuning package on the market, MCNNTUNES (https://arxiv.org/abs/2010.02213), but we do not have much experience with it (yet)

Summary

- Geant4 collaboration is developing and expanding a configuration interface to models, including the Hadronic ones
- This allows to fit simulated distributions to experimental datasets and to extract optimal values of the model parameters
- We have preliminarily explored such avenue with regards to such Hadronic models as PreCompound, Bertini cascade, and Fritiof
- We are currently concentrating on the Fritiof model, and have demonstrated that certain model parameters can be optimized, through fitting techniques, to reach better agreement between MonteCarlo results and experimental thin target data
- There is still plenty of work ahead or us, and we plan to continue the studies
- We also plan to gradually expand the efforts to other Geant4 Hadronic models
- In the future, when properly tested and mature, we plan on offering alternative tunes for FTF and other models, for certain study cases

BACKUP SLIDES

Datasets (so far)

- IAEA – 3 GeV proton on C, Fe, Pb
  - K.Ishibashi et al., J.Nucl.Sci.Tech. Vol.34 N.6 1997
- HARP -- 3-12 GeV/c proton or pion on various nuclear targets
  - M. Apollonio et al., Nucl. Phys. A821 118, 2009; Phys.Rev.C80 065207, 2009; Phys.Rev.C80 035208, 2009; Phys.Rev.C82 045208, 2010
  - M.G. Catanesi et al., Phys.Rev.C77 055207, 2008
- ITEP771 – 5-7.5 GeV/c proton or 5 GeV/c pion on various nuclear targets
  - Yu. D. Bayukov et al., Preprints ITEP-148-1983; ITEP-172-1983; Sov.J.Nucl.Phys. 42 116, 1985
- NA61 – 31 GeV/c proton or 60 GeV/c pi+ on Be, C
  - N. Abgrall et al. , Eur.Phys.J.C 76, 2016 (proton beam)
  - A. Aduszkiewicz et al. , Phys.Rev.D100 112004, 2019 (pion beam, only data on C used so far)
- SAS M6E – 100 GeV/c proton or pi+ on C, Cu, Pb (at present, not used in fits but is used in validation)
  - D.S. Barton et al., Phys. Rev. D27, 2580 (1983)
- NA49 – 158 GeV/c on C (at present, not used in fits but is used in validation)
  - http://spshadrons.web.cern.ch/spshadrons/

The Fitting Package

- **Professor**: http://professor.hepforge.org
  - It was **popular at the time** we started out
  - … and has been working quite well until now
- "Fundamentally, the idea of Professor is to reduce the exponentially expensive process of brute-force tuning to a scaling closer to a power law in the number of parameters, while allowing for massive parallelization and systematically improving the scan results by use of a deterministic parameterization of the generator's response to changes in the steering parameters." – from Professor's web site
- A set of parameters Pi={xi , yi , zi ,…} is a "point" in the multi-parameter space
- Randomly sample multi-parameter space
- For each Pi simulate data combinatorics: beam × energy × target …
- Derived quantities are histograms
- Each simulated (histogram) bin content is f(Pi) - polynomial approximation
- 3rd order polynomial is a default
- Fit experimental data with f(Pi) to explore sensitivity and coupling of parameters

Number of parameters vs polynomial order vs number of "points" in the parameter space

```
int numCoeffs(int dim, int order) {
  int ntok = 1;
  int r = min(order, dim);
  for (int i = 0; i < r; ++i) {
    ntok = ntok*(dim+order-i)/(i+1);
  }
return ntok;
}
```

3 dimensional parameter space:

Polynomial order	Minimum samples
0	1
1	4
2	10
3	20
4	35
5	56
6	84
7	120
8	165
9	220
10	286

20 dimensional parameter space:

Polynomial order	Minimum samples
0	1
1	21
2	231
3	1771
4	10626
5	53130
6	230230
7	888030
8	3108105
9	10015005
10	30045015

50 dimensional parameter space:

Polynomial order	Minimum samples
0	1
1	51
2	1326
3	23426
4	316251
5	3478761
6	32468436
7	264385836