Hadronization models in GENIE

http://www.genie-mc.org

Costas Andreopoulos[1,2]

[1]University of Liverpool, Department of Physics

[2]U.K. Research and Innovation, Science and Technology Factilities Council,
Rutherford Appleton Laboratory, Particle Physics Department

NuSTEC Workshop on Shallow- and Deep-Inelastic Scattering
Gran Sasso Science Institute (GSSI), L'Aquila, Italy - 11-13 October 2018

October 13, 2018

Hadronization modelling

Models of exclusive single-meson production (single-π, single-K)

- Hadronic distributions calculable

But, no model of exclusive hadronic multiparticle production.

- Make predictions by stitching together modelling elements
- **Hadronization modelling** an important ingredient

Hadronization modelling

Functionally, it is a process that receives a minimal set of inputs:

- The neutrino ID ($\nu/\bar{\nu}$)
- The hit nucleon and hit quark IDs
- The interaction type (CC/NC)
- The hadronic invariant mass W

and it generates **hadronic showers**!

In the context of the GENIE empirical model, hadronic showers are produced by answering the following 3 questions:

- How many hadrons are produced?
- What are the hadron IDs?
- What are the hadron 4-momenta?

Significance

Hadronization modelling impacts:

- **Neutrino energy reconstruction**

  If the detector responds differently to different hadrons, then you need to know the exact mixture of hadrons in your showers!

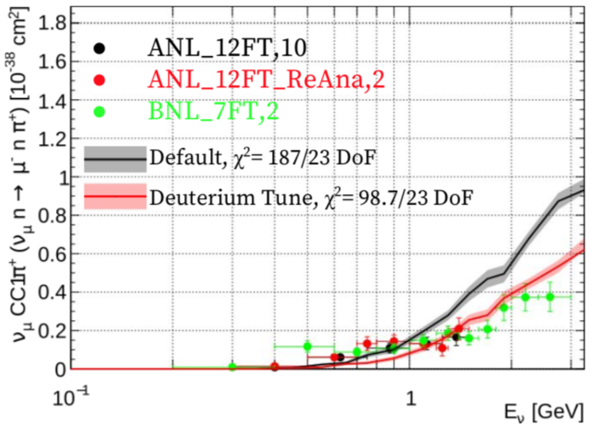
- **Efficiency calculations for event identification**

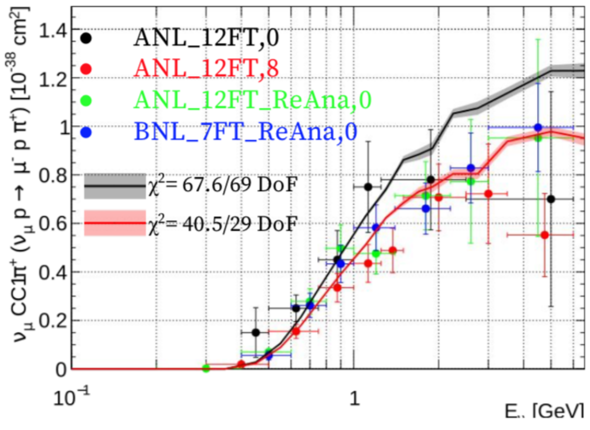
  For example, hadronic showers (from NC or high-y ν_μCC events) with large EM component could be misclassified as ν_eCC events. Shower mismodelling can impact the estimation of backgrounds.

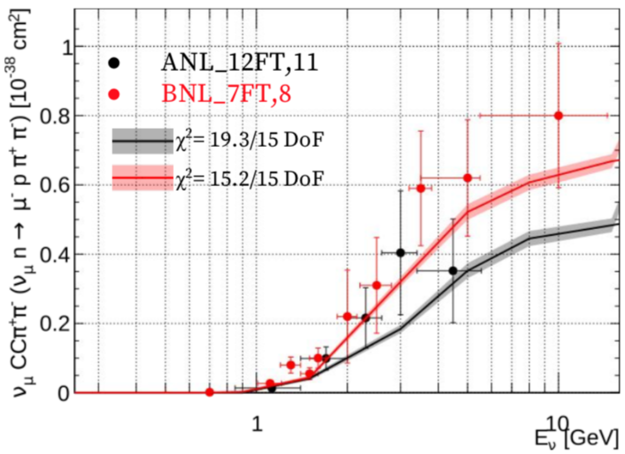
Significance

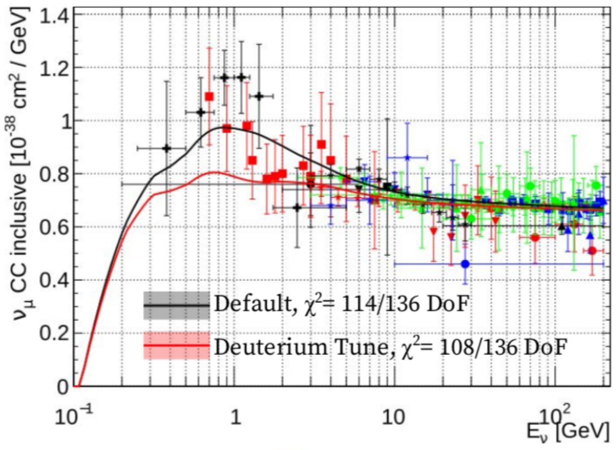
Hadronization modelling is not just about hadron shower shapes.

String coupling with cross-section modelling (even for 1-π): Decomposes the (computed) incl. cross-section into nearly all excl. cross-sections.









Hadronization modelling in GENIE

Three main elements:

- PYTHIA6, valid at higher W
- Empirical model, valid for SIS/DIS at W $<$ 3 GeV
- Empirical model, specialised for DIS charm production

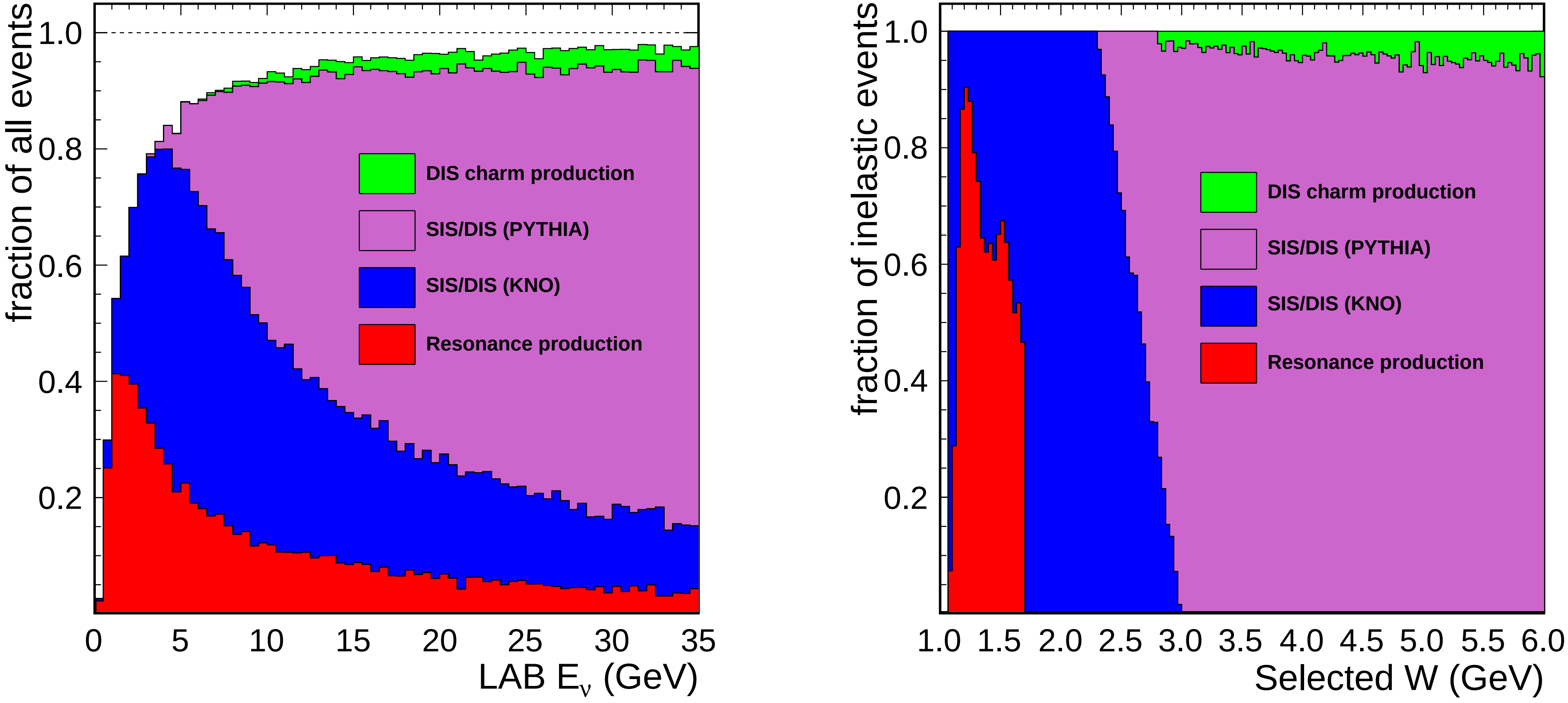
The above are not all the places in GENIE that produce hadrons (pre-FSI):

- Resonance single-pion production
- Other states from baryon resonance decays
- Single-Kaon production

These modelling elements are not thought of as parts of the family of GENIE *fragmentation models* and will not be covered.
However, it should be recognised that the distinction can be somewhat arbitrary (e.g. mixture of resonance / non-resonance contributions)

Hadronization modelling

Fractions of GENIE events generated by each hadronization model:

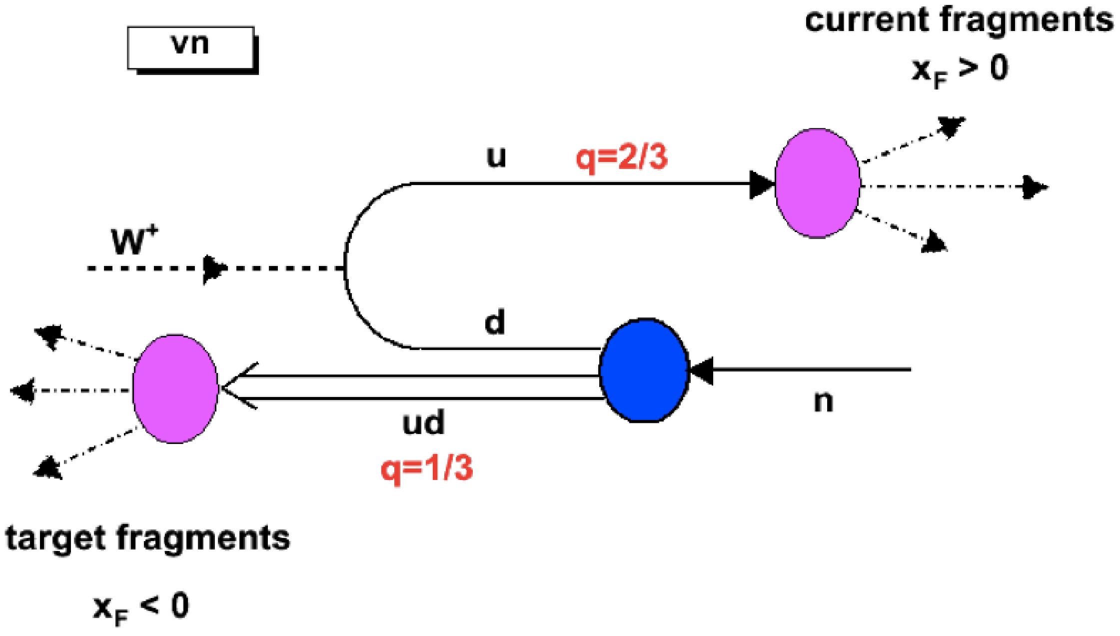


Basic picture of hadronization

In the simple quark-parton model, the lepton interacts via W/Z exchange with one of the (anti)quarks in the nucleon:

- Struck quark hadronizes (current fragments, predominantly $x_F > 0$)
- Target remnant also hadronizes (target fragments, $x_F < 0$)

- 2 correlated hadron jets.
- Smooth transition through a central rapidity region.
- At low W, the two fragmentation regions largely overlap.



Data on neutrino-induced hadron shower characteristics

Several pieces of data exist.

- Average charged and neutral particle multiplicities
- Forward and backward hemisphere average multiplicities
- Multiplicity dispersion as function of avg multiplicity and W
- Multiplicity correlations (e.g. charged hadrons - π^0)
- Fragmentation functions (z distributions)
- x_F distributions
- p_T^2 distributions
- x_F - p_T^2 distributions

However, coverage of the low - W range (W $<$ 4-5 GeV is poor)

PYTHIA6

- LUND string fragmentation model
- Uses the assumption of linear confinement as a starting point.
- As partons move apart, their colour flux tube gets stretched.
- Stored potential energy increases linearly with distance of colour charges.
- You can think of the "string" as the axis of the flux tube.
- The string constant is ~ 1 GeV/fm.
- As the potential energy increases, the string may break producing a $q\bar{q}$ pair.
- String breaks causally disconnected; simulated in a convenient order.
- A break typically creates a meson.
- Baryons also produced; A string can break by antidiquark-diquark production, or baryons can be produced using a 'popcorn' model.
- With every break, a produced hadron takes away a fraction of the available energy/momentum.
- Continuing till some cut-off point.

Driving PYTHIA6 from GENIE

Some amount of monkey business in making **quark** + **diquark** assignments most certainly due to our own unfamiliarity with PYTHIA. Luckily, overall generation outcomes not sensitive to choices made.

Init state	Hit quark	Leading quark	Remnant system	PYTHIA6 assignment		Weirdness level
ν + p CC	d valence	(d $\rightarrow$) u	uu	u	uu	
ν + p CC	d sea	(d $\rightarrow$) u	$\bar{d}$ + uud	u	uu	*
ν + p CC	s sea	(s $\rightarrow$) u	$\bar{s}$ + uud	u	uu	**
ν + p CC	$\bar{u}$ sea	($\bar{u}$ $\rightarrow$) $\bar{d}$	u + uud	u	uu	***
ν + n CC	d valence	(d $\rightarrow$) u	ud	u	ud	
ν + n CC	d sea	(d $\rightarrow$) u	$\bar{d}$ + udd	u	ud	*
ν + n CC	s sea	(s $\rightarrow$) u	$\bar{s}$ + udd	u	ud	**
ν + n CC	$\bar{u}$ sea	($\bar{u}$ $\rightarrow$) $\bar{d}$	u + uud	u	ud	***
...	...	...	...	...	...	...
...	...	...	...	...	...	...

PYTHIA6 tuning

- NOMAD (NUX) PYTHIA6 tuning was adopted in 2007.
- Some PYTHIA6 defaults were restored in later GENIE re-tune (2010).

	PYTHIA default	NUX 2001	GENIE 2010 re-tune
$s\bar{s}$ production suppression	0.30	0.21	0.30
$< p_T^2 > (GeV^2)$	0.36	0.44	0.44
Non-gaussian p_T tail parameterization	0.01	0.01	0.01
Fragmentation cut-off energy (GeV)	0.80	0.20	0.20

Main issues with GENIE PYTHIA studies (circa 2007!).

- Could not find enough knobs to influence predictions.
    - How to express uncertainty?
- Could not understand change in behaviour below W $\approx$ 2.5 - 3.0 GeV.
    - Limits of validity range?

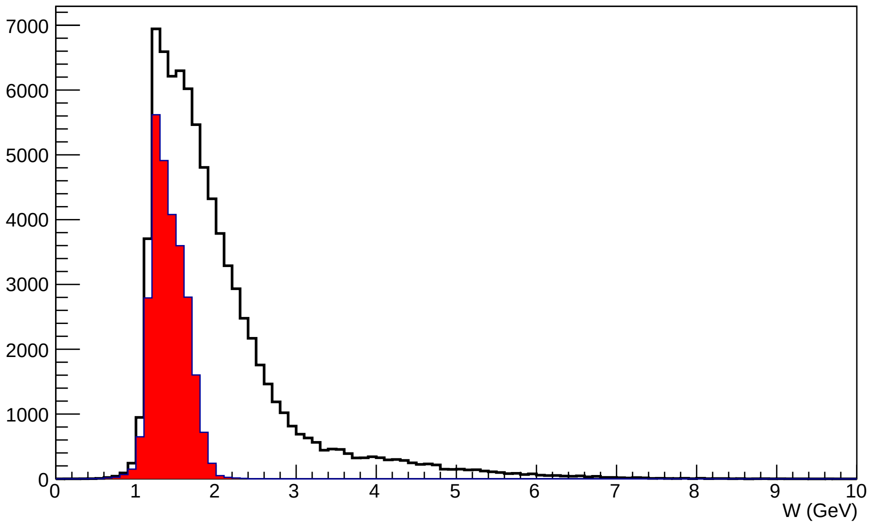
Extending the validity of GENIE model to lower W

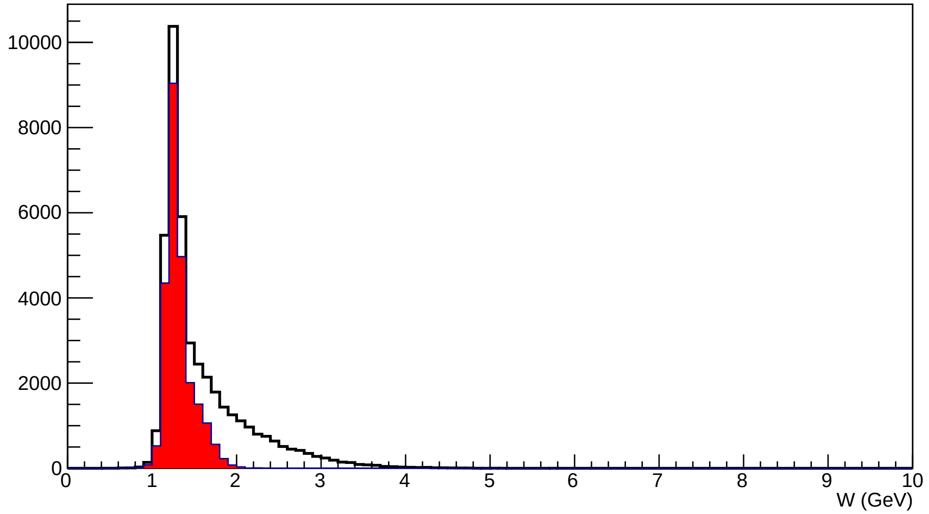
On the right, the invariant mass distributions for inelastic events. (Distribution is smeared due to Fermi momentum.)

Up plot: DUNE, Down: HK
Red component: Resonances

Kinematic area **below 2.5 - 3.0 GeV** in invariant mass is **critically important**.

Augment PYTHIA with an **empirical GENIE model**, anchored to data and **valid in the area below 3 GeV**. Install handles to express uncertainty.





Empirical low-W model

- An **effective KNO-based hadronization model was built** (T.Yang, H.Gallagher, P.Kehayias, C.Andreopoulos - circa 2007) for low W and was "*integrated*" with PYTHIA to cover the full kinematic space (AGKY model, Eur.Phys.J.C63:1-10,2009)
- The model was **anchored on several pieces of bubble chamber data and captures several observations on the characteristics of neutrino-induced hadron showers** (for an excellent description, see Norbert Schmitz, Adv.Ser.Direct.High Energy Phys. 2 (1988) 3-56)
- A similar, KNO-inspired model pre-existed (in neugen3). Several model improvements were installed in 2007-2008,
- Several caveats were recognized over time; few improvements were made (e.g strange baryon production by K.Hoffmann, H.Gallagher)

Empirical low-W model: How many hadrons are produced?

First order of business is to calculate the hadronic multiplicity.

First, we answer that question on average.

Average charged hadron multiplicities $< n_{ch} >$ are well described by:

$$< n_{ch} >= a + b \cdot ln(W^2/GeV^2)$$

The values of a,b were measured in several experiments.

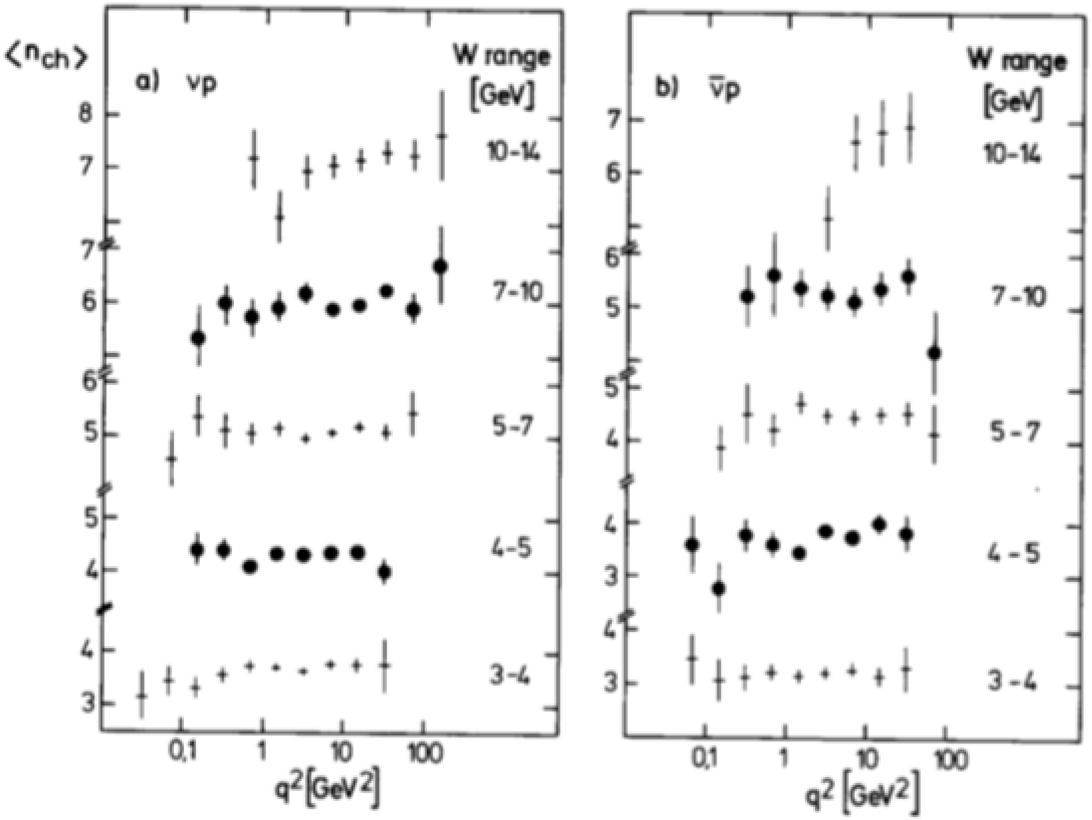
GENIE uses:

	νp	νn	$\bar{\nu} p$	$\bar{\nu} n$
a	0.40	-0.20	0.02	0.80
b	1.42	1.42	1.28	0.95

Empirical low-W model: How many hadrons are produced?

Average charged hadron multiplicities $< n_{ch} >$ could, more generally, have an additional Q^2 dependence:

$$< n_{ch} >= a + b \cdot ln(W^2/GeV^2) + b' ln(Q^2/GeV^2)$$

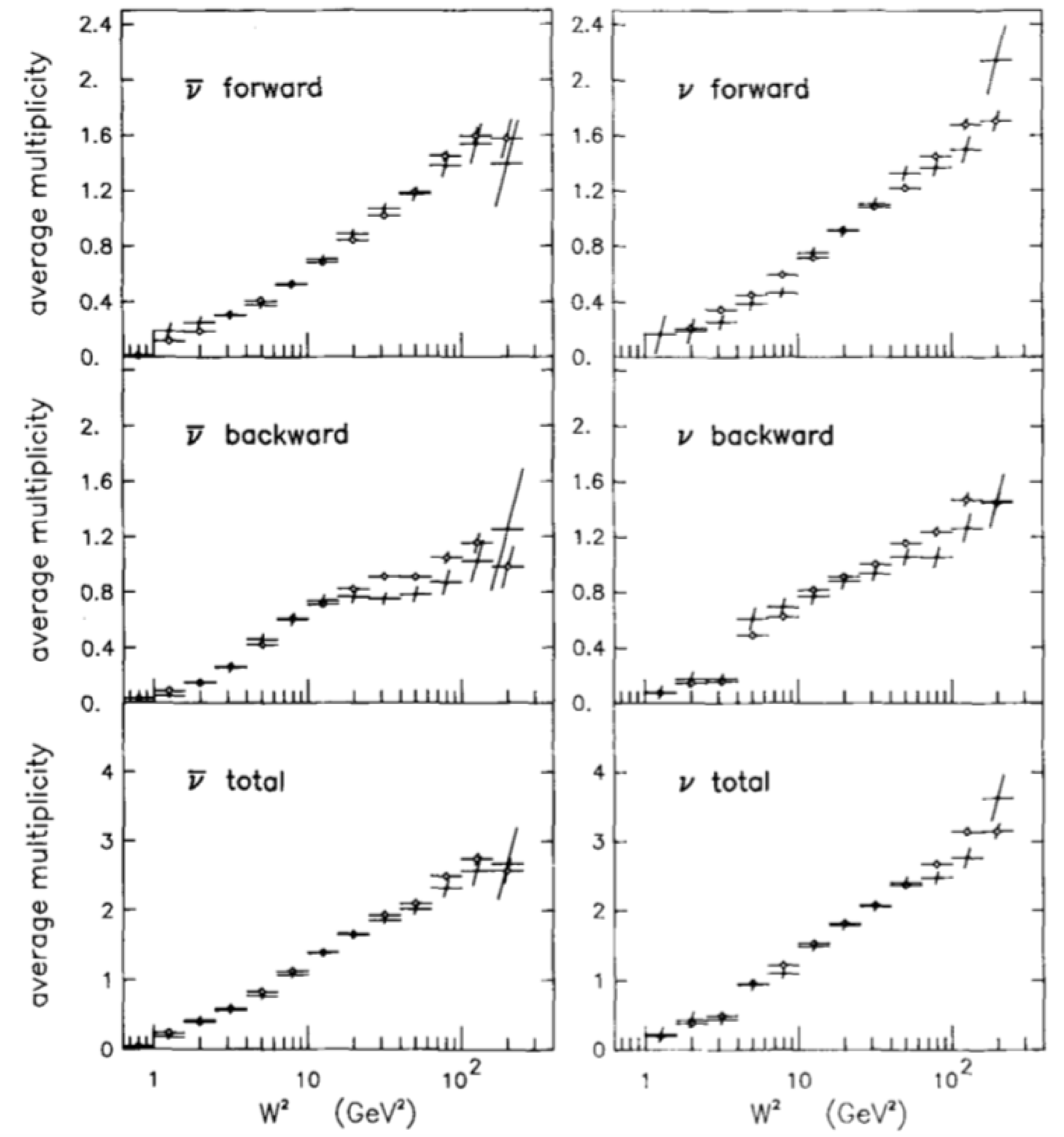


No Q^2 dependence has been observed in $\nu/\bar{\nu}$ scattering [H. Grassler et al., Nucl. Phys., **B223**, 269 (1983)].

Values of b' are 0.04 ± 0.02 for νp and 0.05 ± 0.04 for $\bar{\nu}$p

In GENIE, $b' = 0$ for all channels.

Empirical low-W model: How many hadrons are produced?



The average neutral pion multiplicity was found to be:

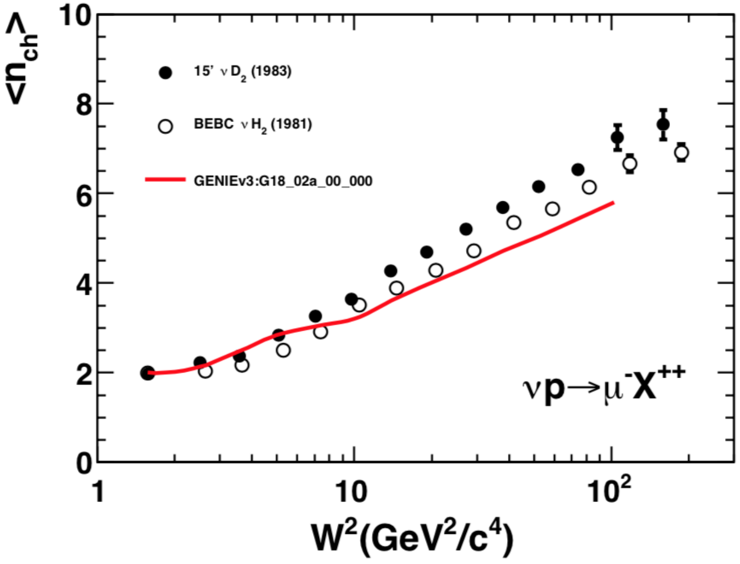
$$\frac{2 < n_{\pi^0} >}{< n_{\pi^+} > + < n_{\pi^+} >} \approx 1$$

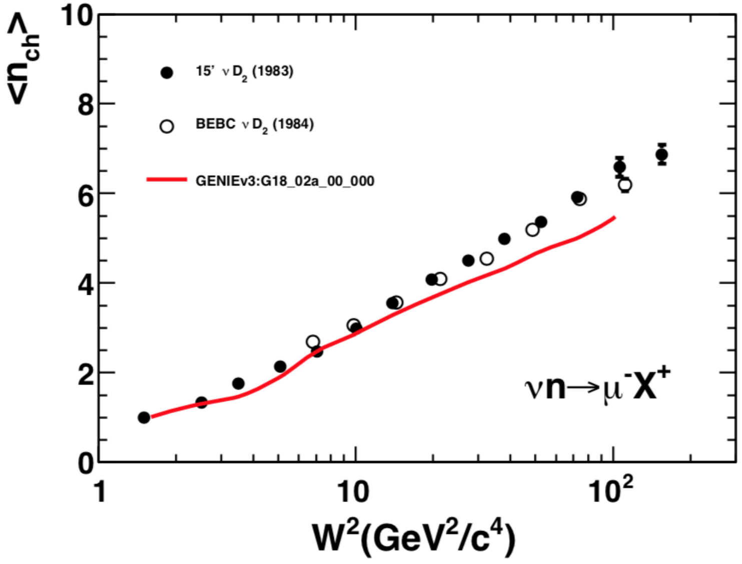
Therefore, we can write:

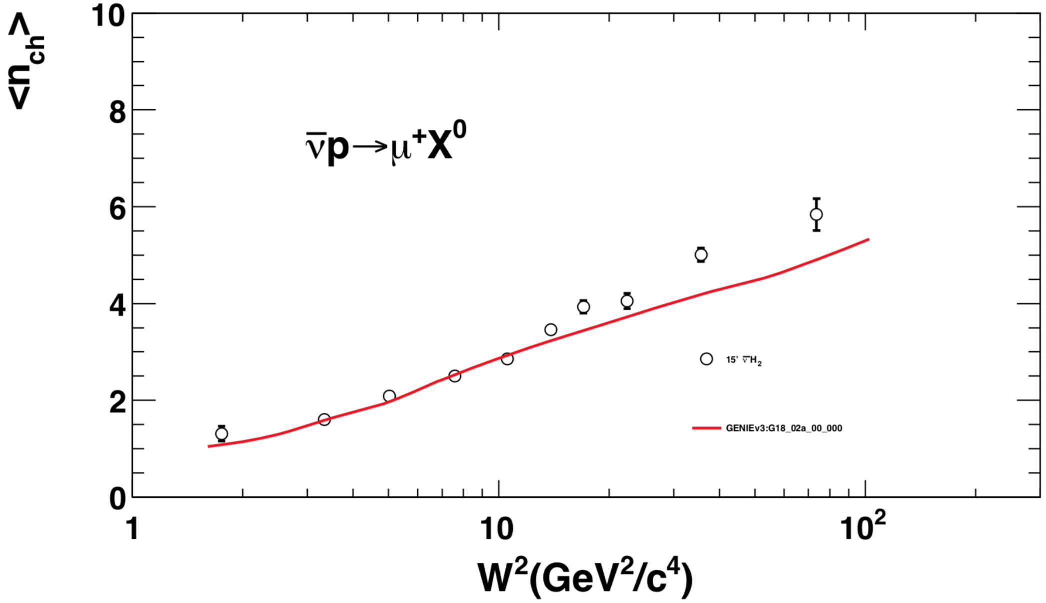
$$< n_{tot} > \approx 1.5 < n_{ch} >$$

On the left: Avg. π^0 multiplicity (crosses) and half the sum of the avg. charged π multiplicities (squares) [W. Wittek et al., Z. Phys. C **40**, 231 (1988).

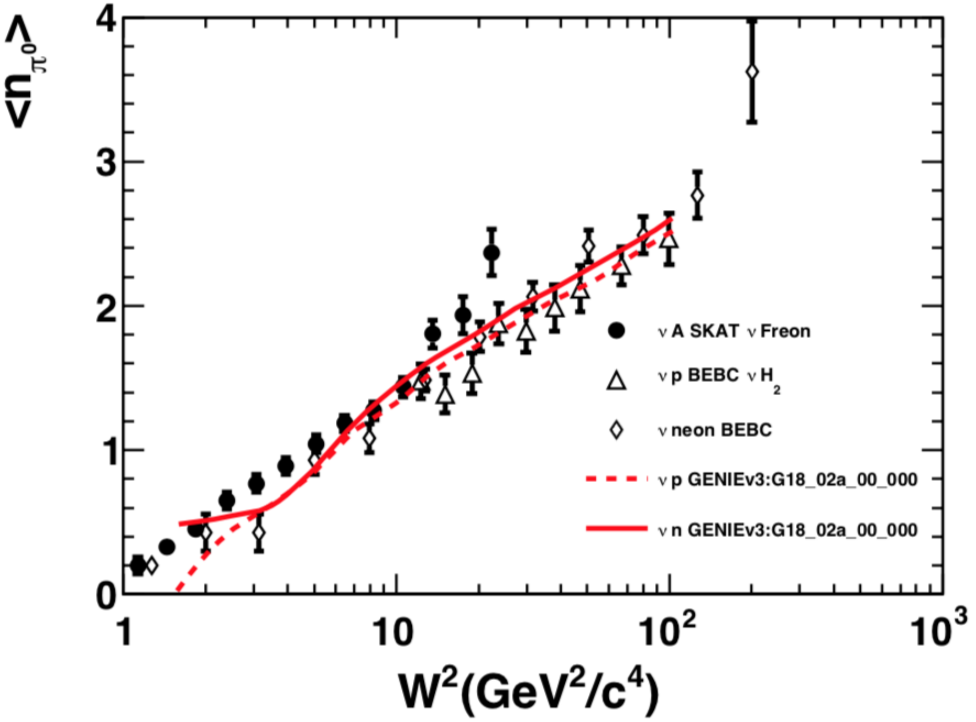
GENIE comparisons with average charged multiplicity data

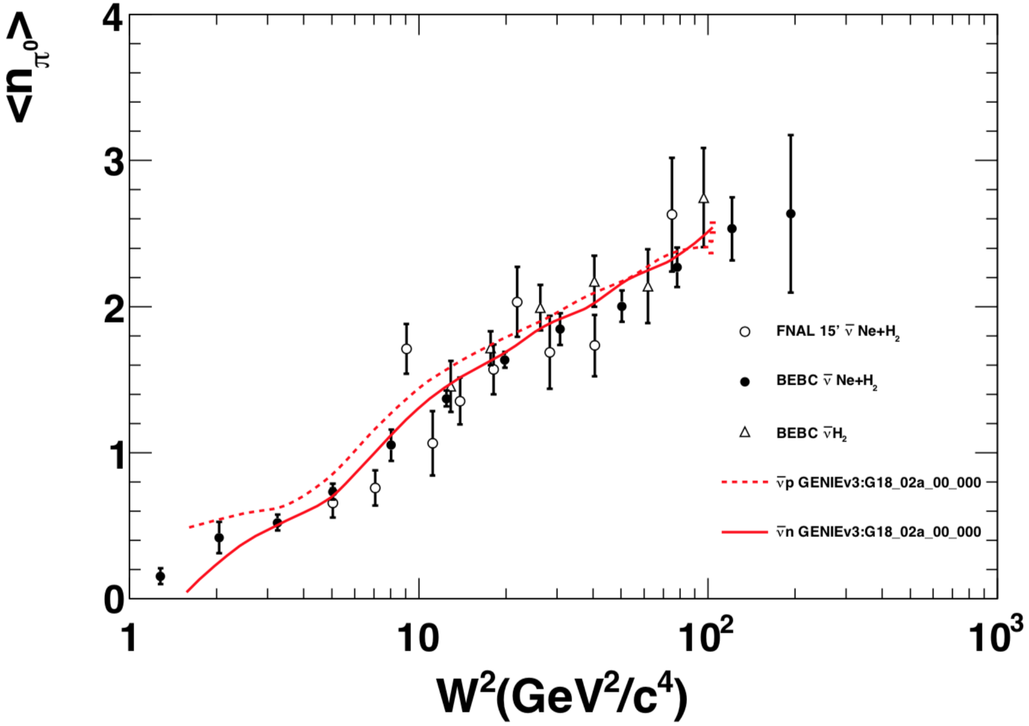






GENIE comparisons with average neutral multiplicity data





Empirical low-W model: How many hadrons are produced?

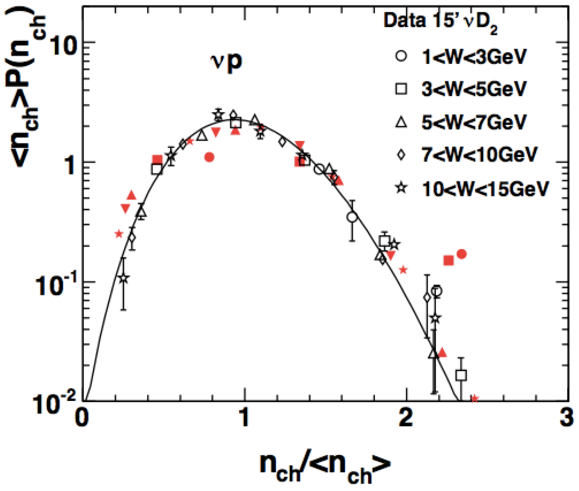
From average → to actual multiplicities on an event-by-event basis?

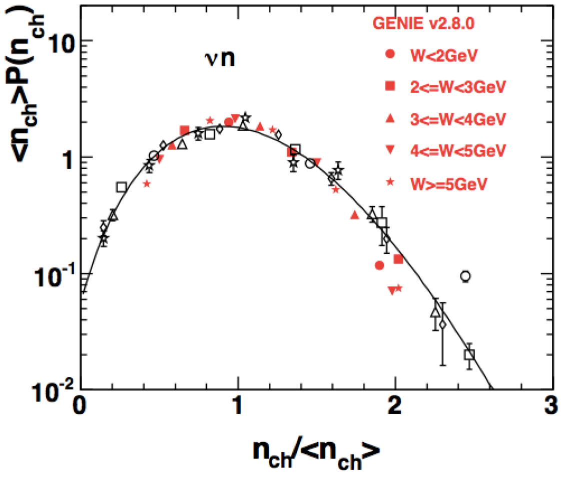
- Need more information than just $< n_{tot} >$!
- Require the probability distribution of n_{tot}, $P(n_{tot})$.

- Draw actual multiplicities from a Poisson distribution with given average?
- The particles **are not independently produced** and the actual multiplicity is not Poisson-distributed.
- **KNO scaling** to the rescue!

Empirical low-W model: KNO scaling

KNO scaling: $< n > P(n) = f(n/ < n >)$ is independent of W [Z.Koba, H.B.Nielsen, P.Olesen, Nucl.Phys.B40,317(1972)]





The function $f(z = n/ < n >)$ is parameterized using the Levy function with parameter c:

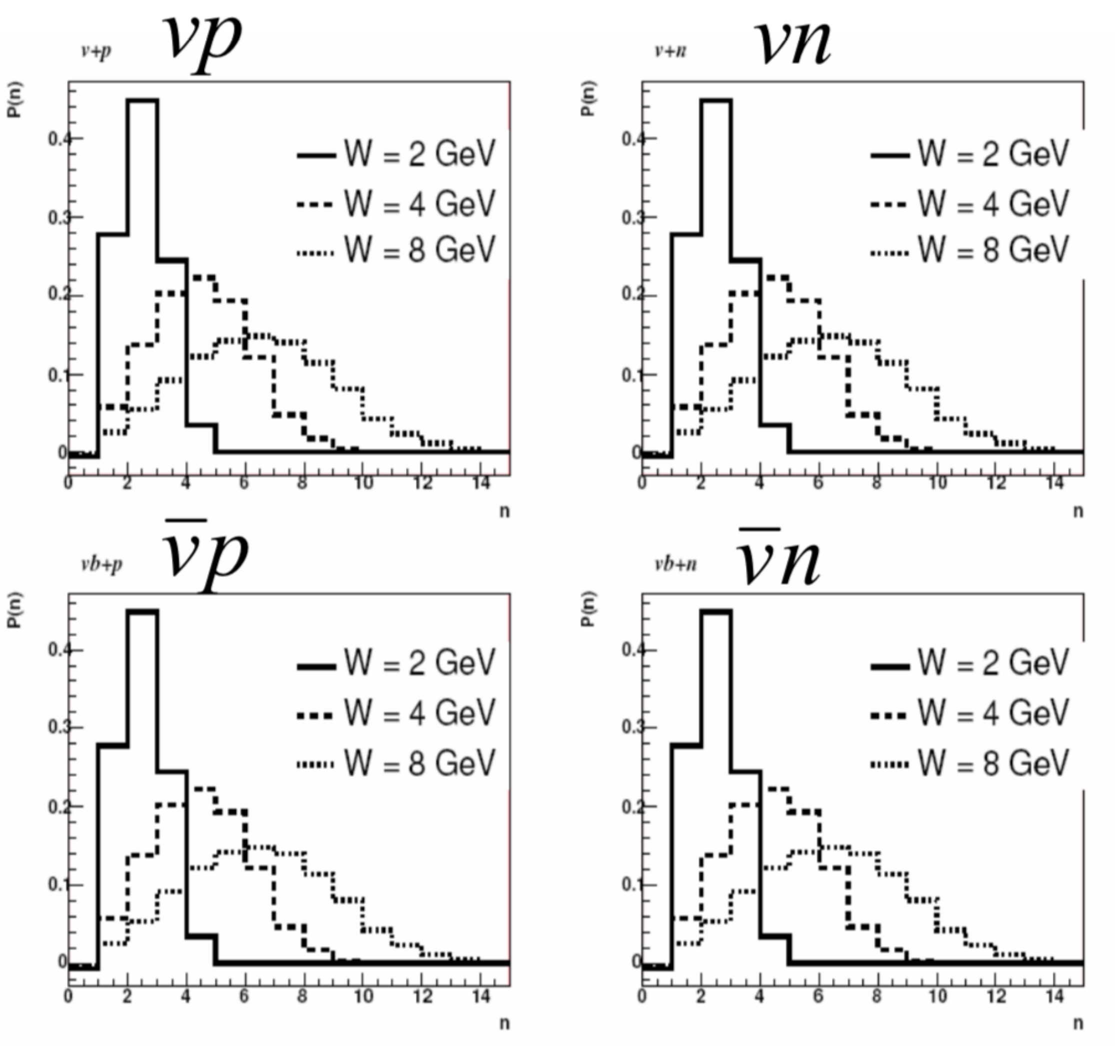
$$L(z; c) = \frac{2e^{-c}c^{cz+1}}{\Gamma(cz+1)}$$

The following parameters c were determined by a GENIE fit to data:

	νp	νn	$\bar{\nu} p$	$\bar{\nu} n$
c	7.93	5.22	as in νn	as in νp

Empirical low-W model: How many hadrons are produced?

From average → to actual multiplicities on an event-by-event basis?

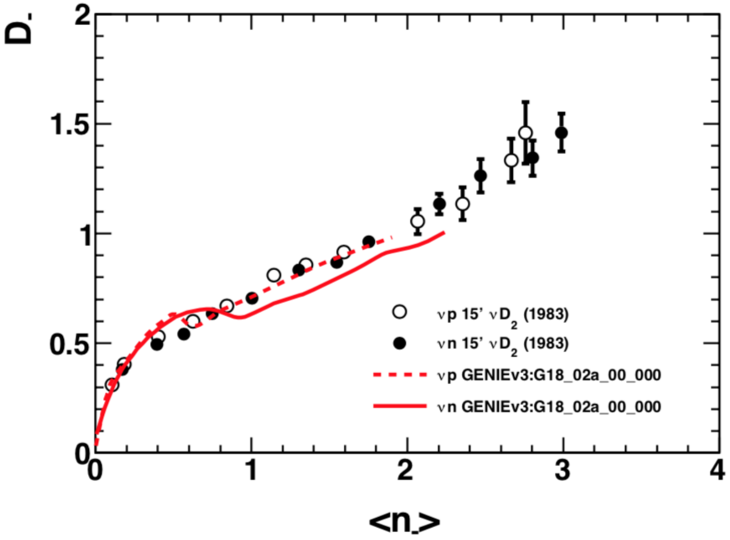


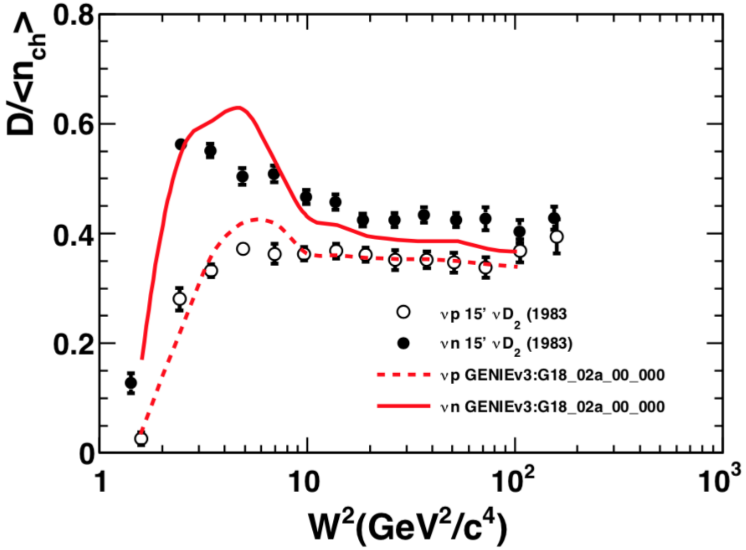
We now have

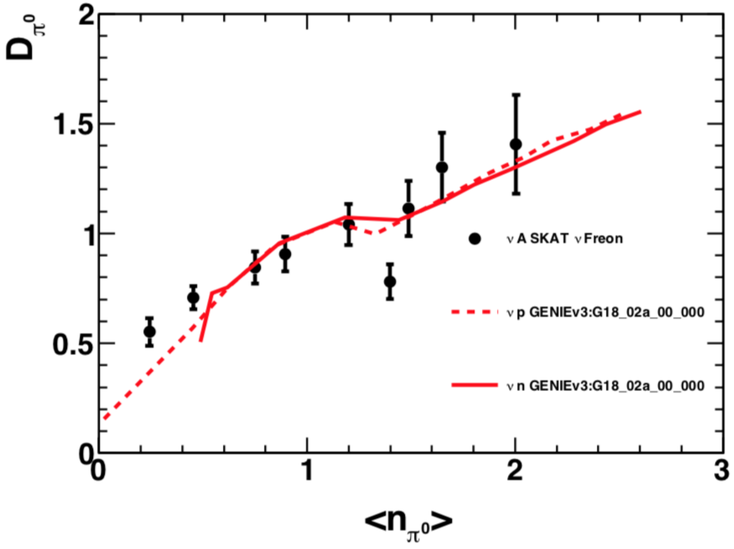
- an expression for $< n_{tot} > P(n_{tot}) = f(n_{tot}/ < n_{tot} >)$, and
- a value for $< n_{tot} >$

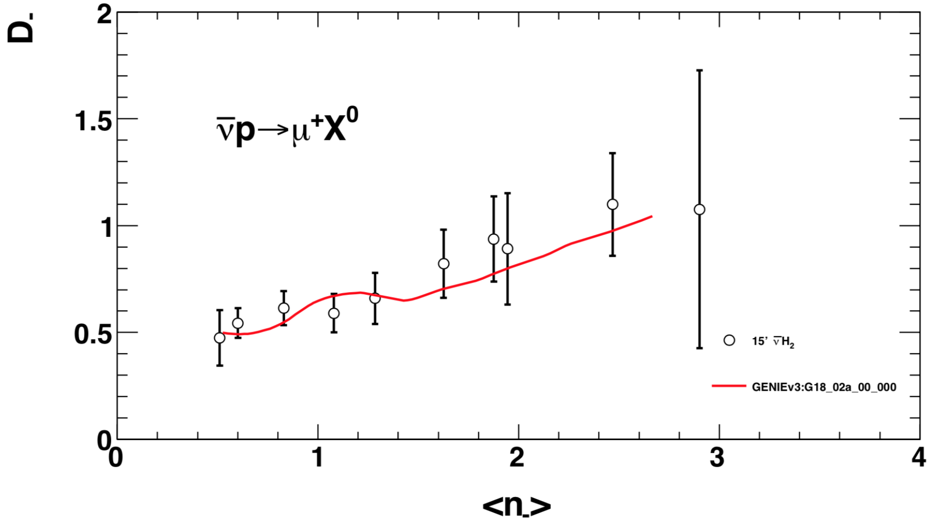
We can combine the two into an expression for the multiplicity probability distribution $P(n_{tot})$

GENIE comparisons with multiplicity dispersion data

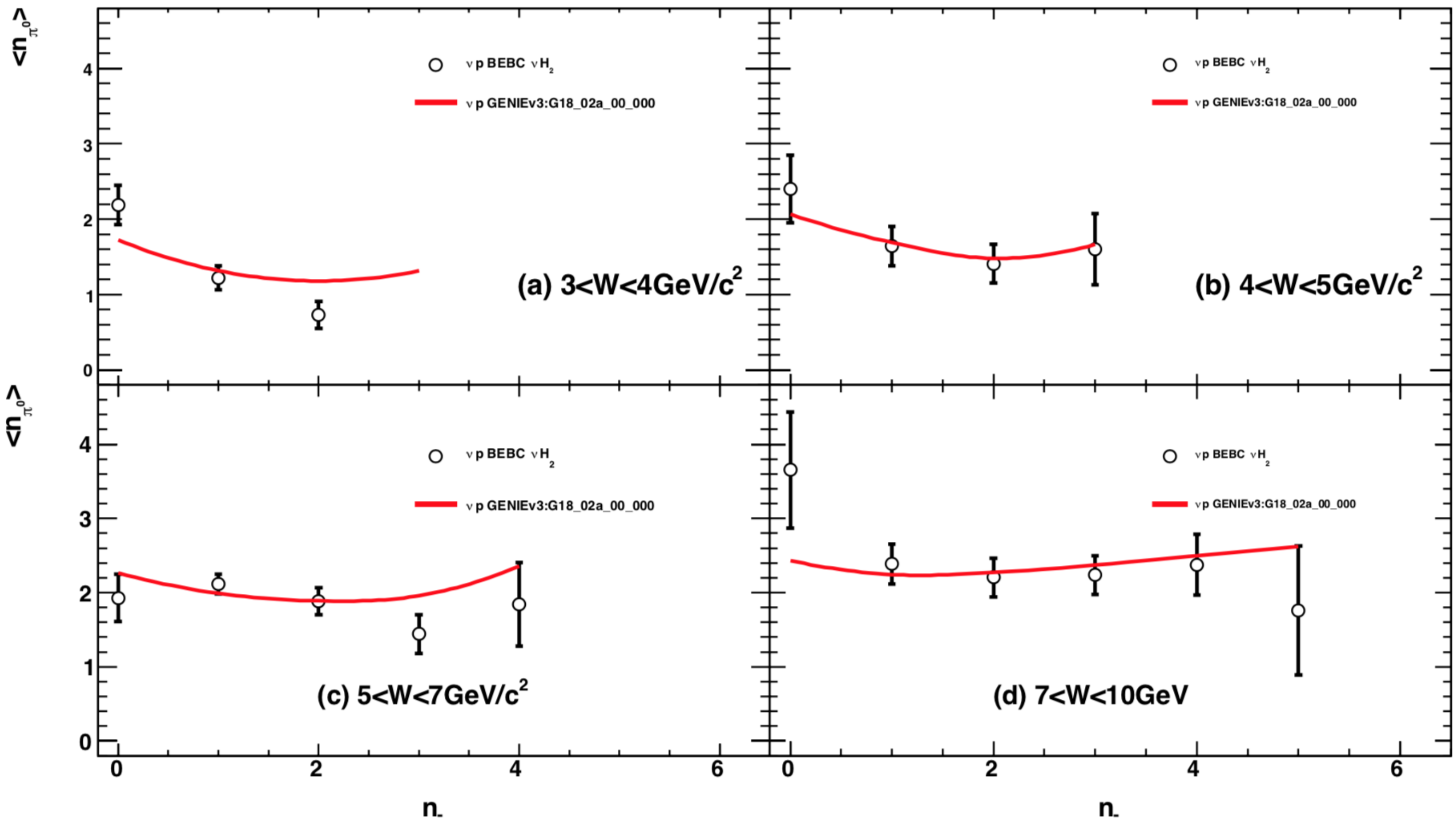




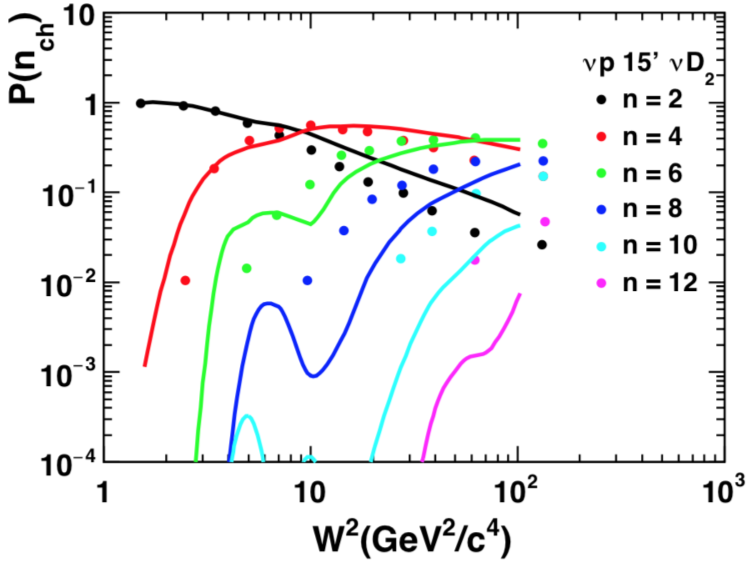


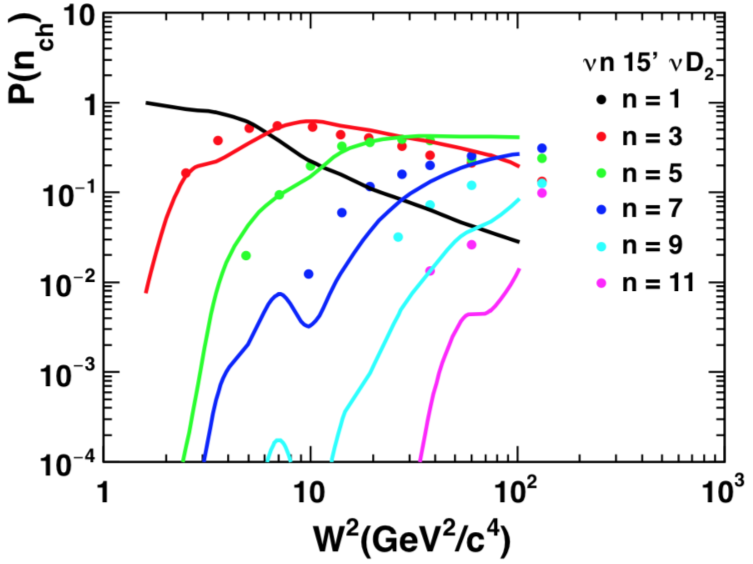


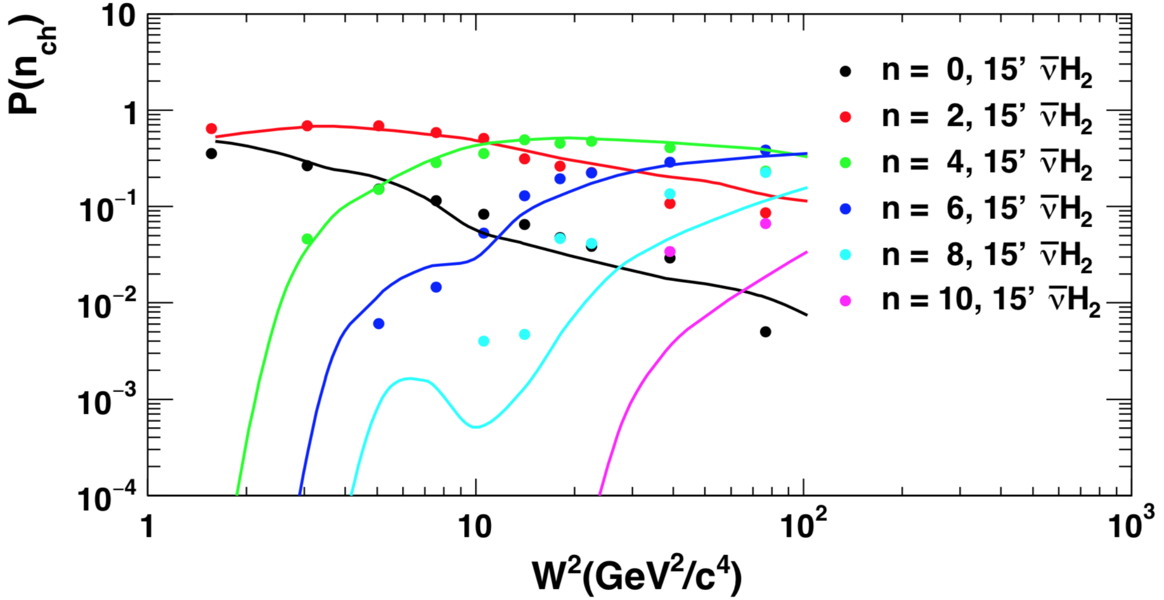
GENIE comparisons with multiplicity correlation data



GENIE comparisons with topological cross-section data







Empirical low-W model: Generating the particle spectrum

Because of kinematical constraints, it is assumed that the shower contains only 1 baryon. We decide between a p or n, with probabilities P_p and P_n (=1-P_p):

	n_{tot}	νp	νn	$\bar{\nu} p$	$\bar{\nu} n$
P_p	2	1.00	0.33	0.67	0.
	>2	0.67	0.50	0.50	0.33

Subsequently, one of those will be converted to a strange baryon (for ν interactions: p $\rightarrow$ Σ^+ and n $\rightarrow$ Λ; for $\bar{\nu}$ interactions: p $\rightarrow$ Λ and n $\rightarrow$ Σ^-) The probability for generating a strange baryon is given by:

$$< n_{hyperon} >= a_{hyperon} + b_{hyperon} \cdot log(W^2)$$

where

	νp	νn	$\bar{\nu} p$	$\bar{\nu} n$
$a_{hyperon}$	0.022	0.022	0.022	0.022
$b_{hyperon}$	0.042	0.042	0.042	0.042

Empirical low-W model: Generating the particle spectrum

Once a baryon (p, n or hyperon) is generated, there are n_{tot}-1 remaining particles. They all assigned meson IDs, with the following procedure:

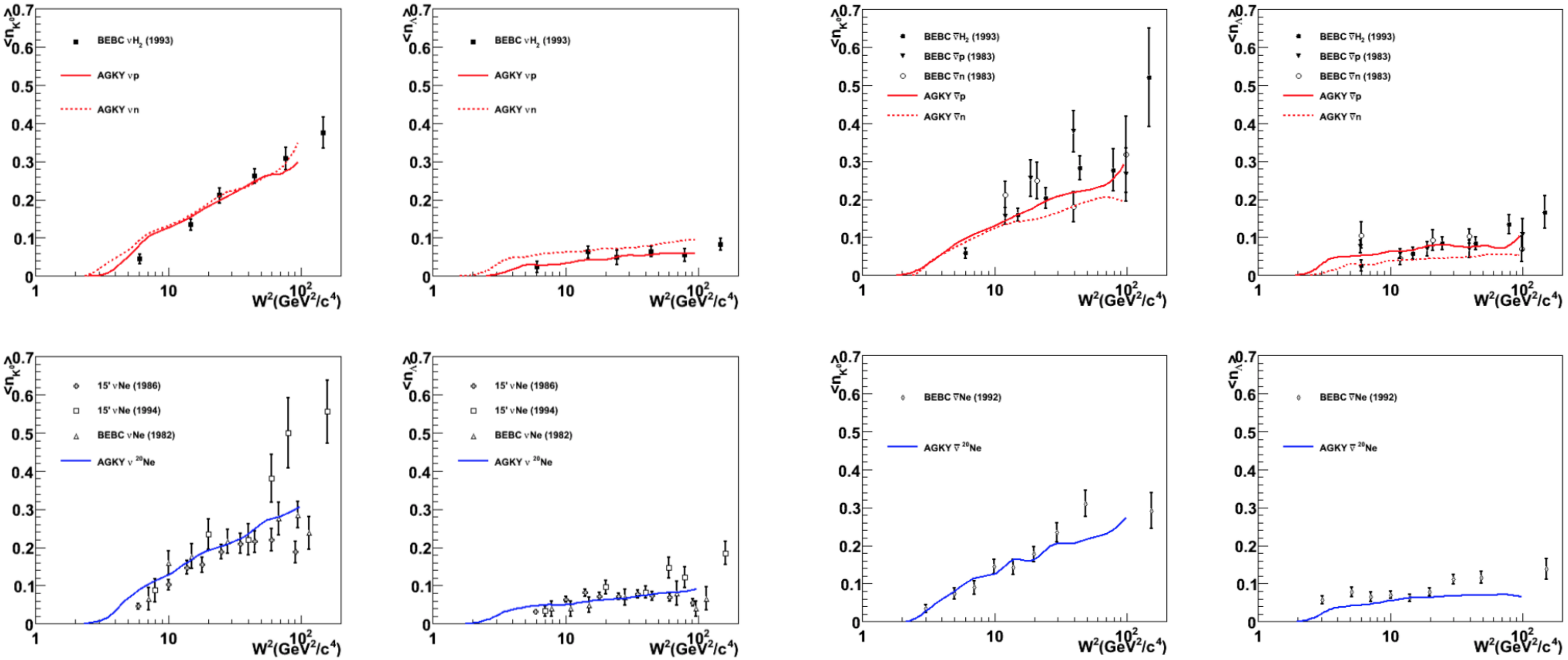
Conserve strangeness:
If a hyperon was produced, add a strange meson to conserve strangeness (no ΔS=1 production in hadronization; this is added separately).

Conserve charge:
Keep on adding π^+'s or π^-'s in the hadron shower till charge is balanced.

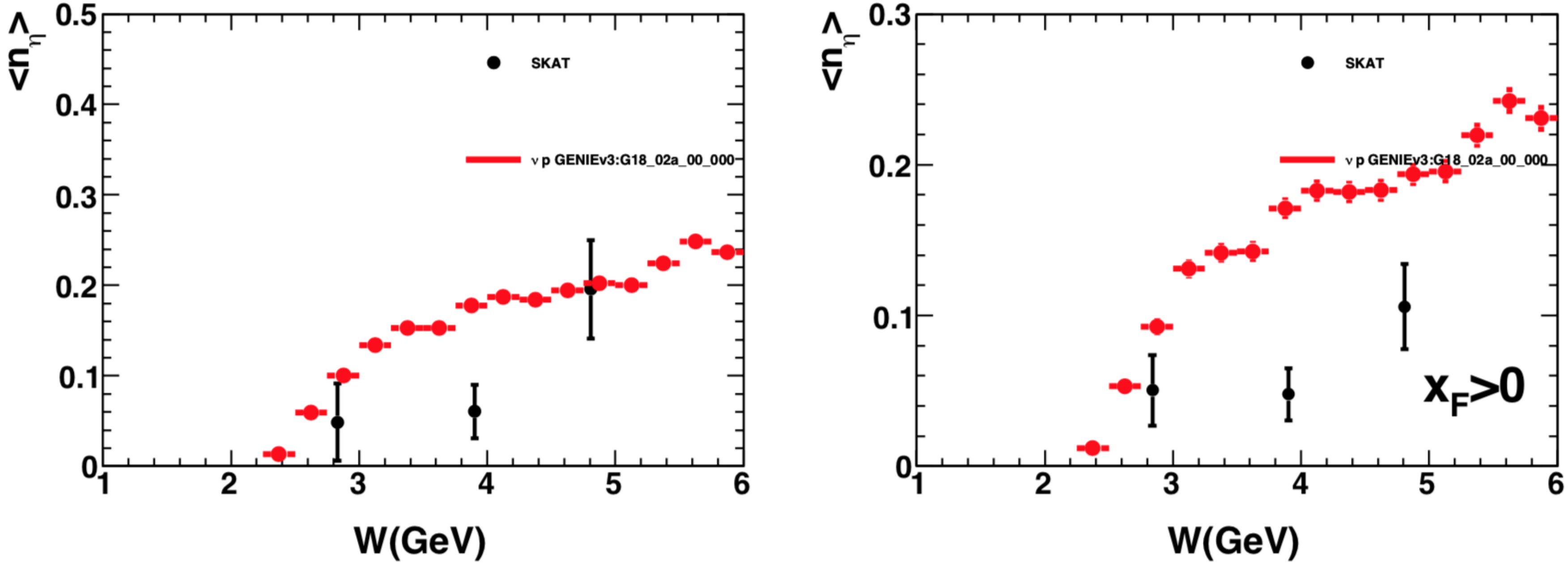
Fill-up:
Add particles in pairs with zero net strangeness and charge till all n_{tot} particle codes have been assigned.
Particle pairs are added with the following probabilities:

$\pi^0\pi^0$	$\pi^+\pi^-$	$K^0\bar{K}^0$	K^+K^-	$\pi^0\eta$	$\eta\eta$
0.3133	0.6267	0.03	0.03	0.	0.

GENIE comparisons with K^0 and Λ production data



GENIE comparisons with η production data (ν)

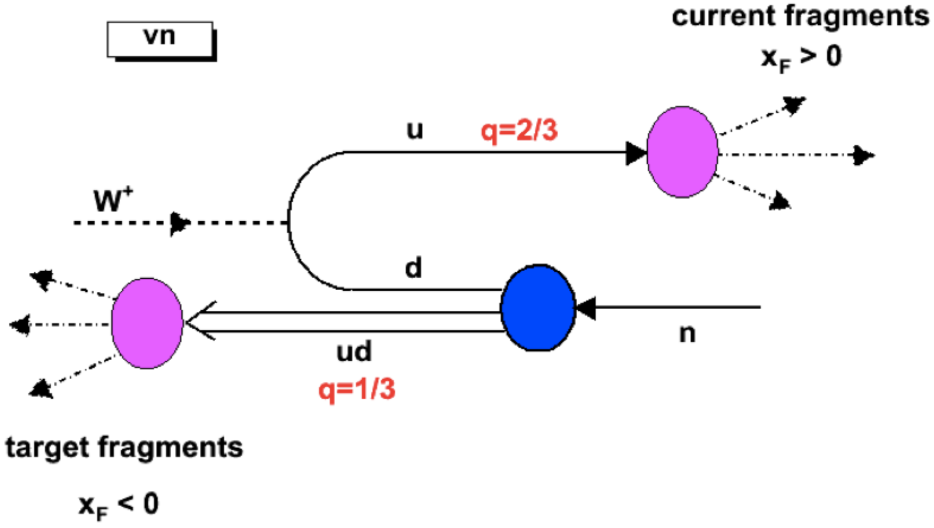


Empirical low-W model: Generating particle momenta

Early versions of the model generated momenta using phase space decays.

This fails to account for key features in the data!

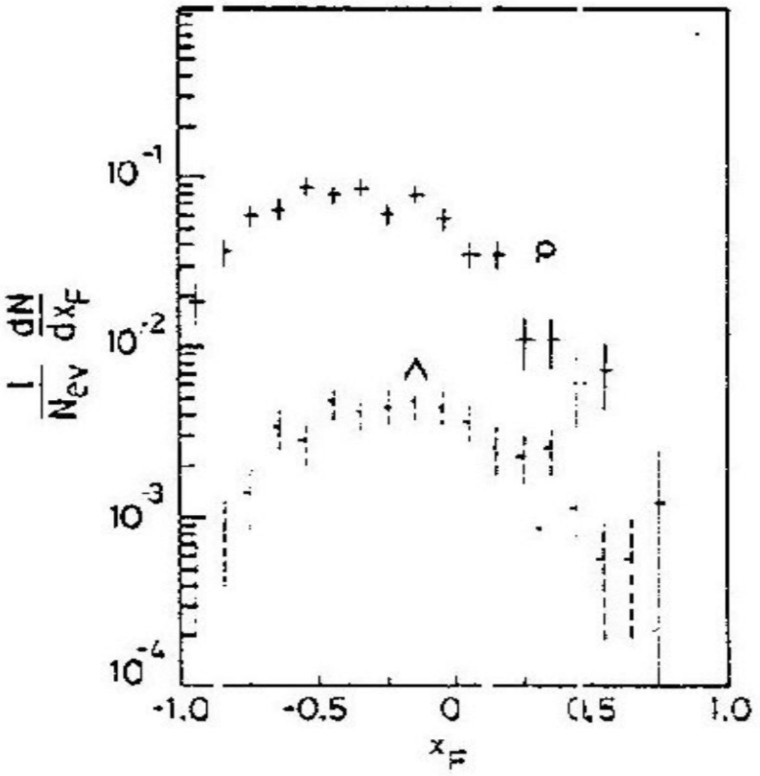
- Final-state nucleon most likely associated with target fragments.
  A heavy, predominantly backwards-going particle creates a multipicity imbalance in the two x_F hemispheres.
- The transverse momentum (p_T) and longitudinal momentum ($\propto x_F$) can not come from the same distribution.

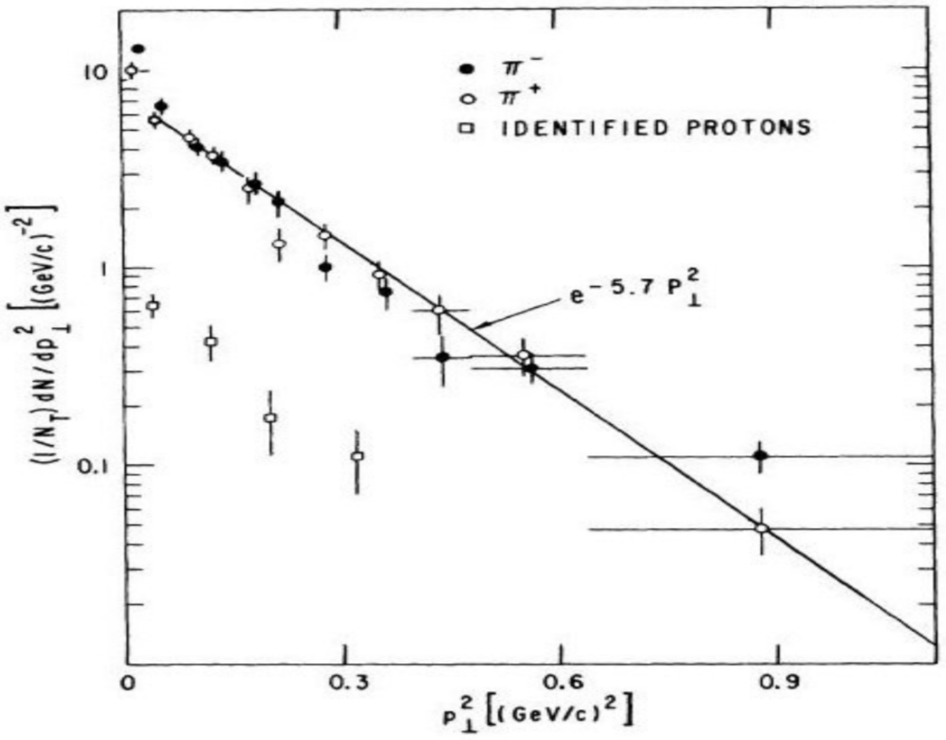


Empirical low-W model: Generating particle momenta

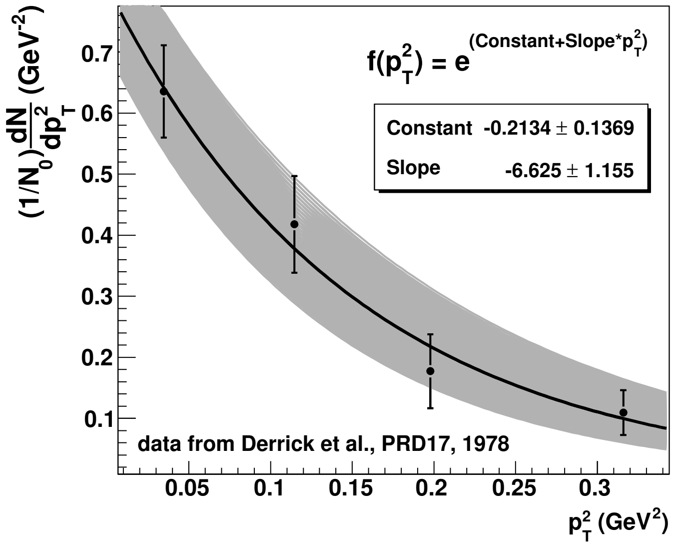
How to introduce an x_F asymmetry and limit p_T (less spherical hadron showers in the HCM frame)?

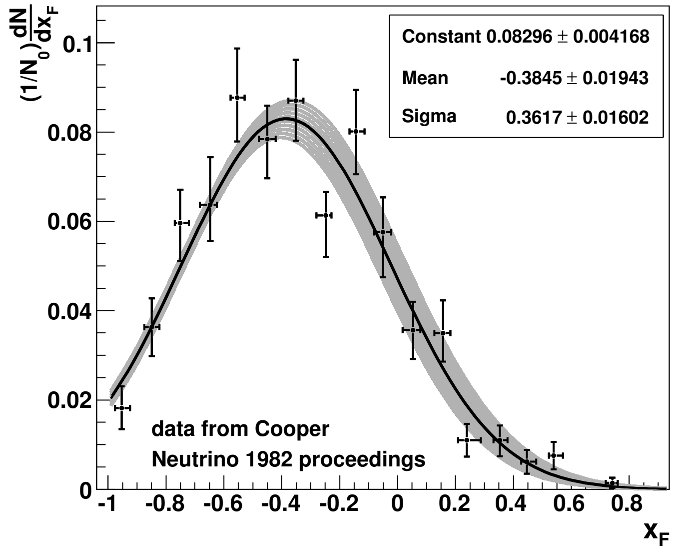
Looking for guidance in data we found the following in Neutrino 1982 proceedings (Cooper) - Never published (to our knowledge) but a reasonable enough starting point for our empirical model.





Empirical low-W model: Generating particle momenta





Baryon p_T and x_F distributions were paraneterized from the previous data:

$$f(p_T^2) \propto e^{-6.625 p_T^2/GeV^2}$$

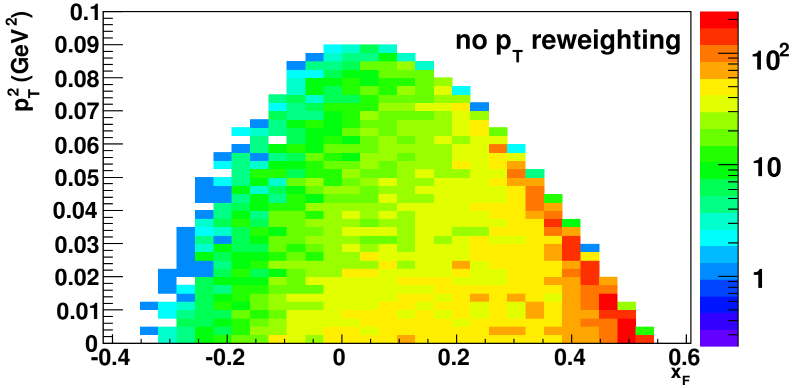
$$f(x_F) \propto e^{-3.817(x+0.385)^2}$$

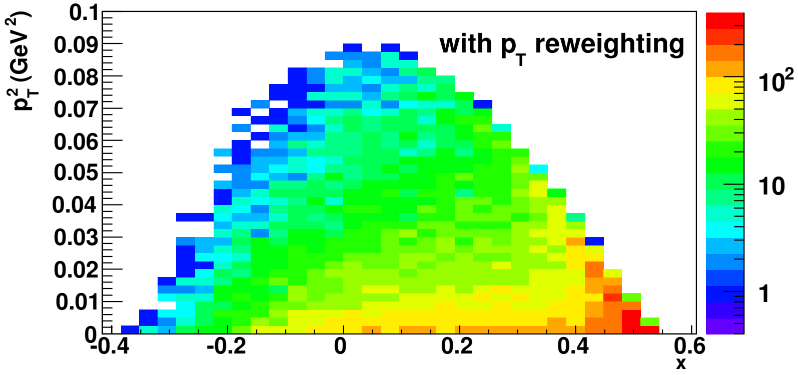
This has **important consequences** for the shower shape.

There is now some evidence that the shower shapes should have been more fwd/bkw-symmetric at lower W.

Future plan is to start shifting the x_F p.d.f. closer to 0 for lower W.

Empirical low-W model: Generating particle momenta





p_T is the momentum perpendicular to the direction of $W^{\pm}/Z^0$ in the HCM frame.

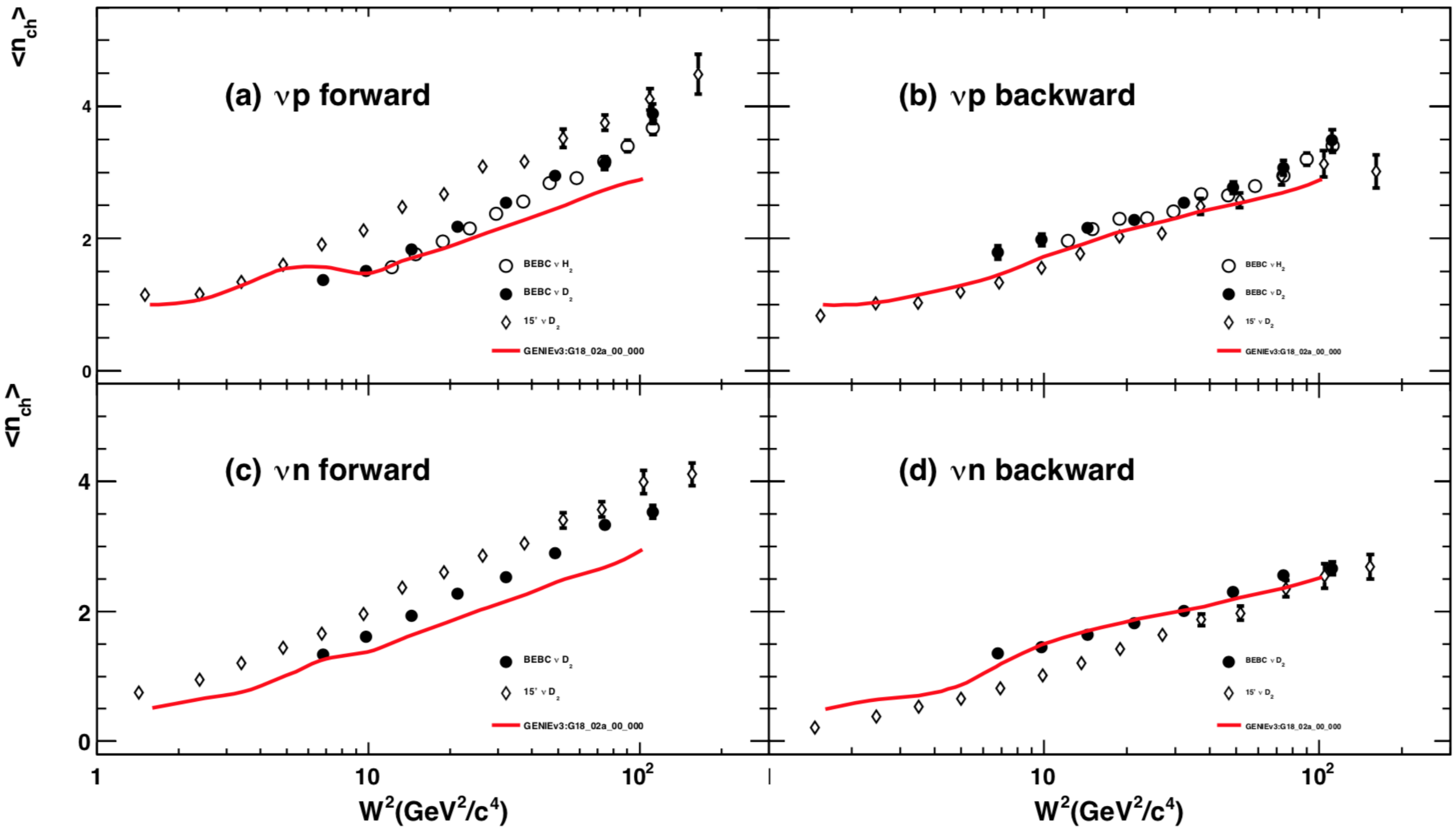
Momenta for the rest of the hadronic system (mass = W - M_{baryon}) is generated with a phase space decay.

The likelihood of higher p_T values is decreased by assigning a weight w_i to each hadron i in the decay:

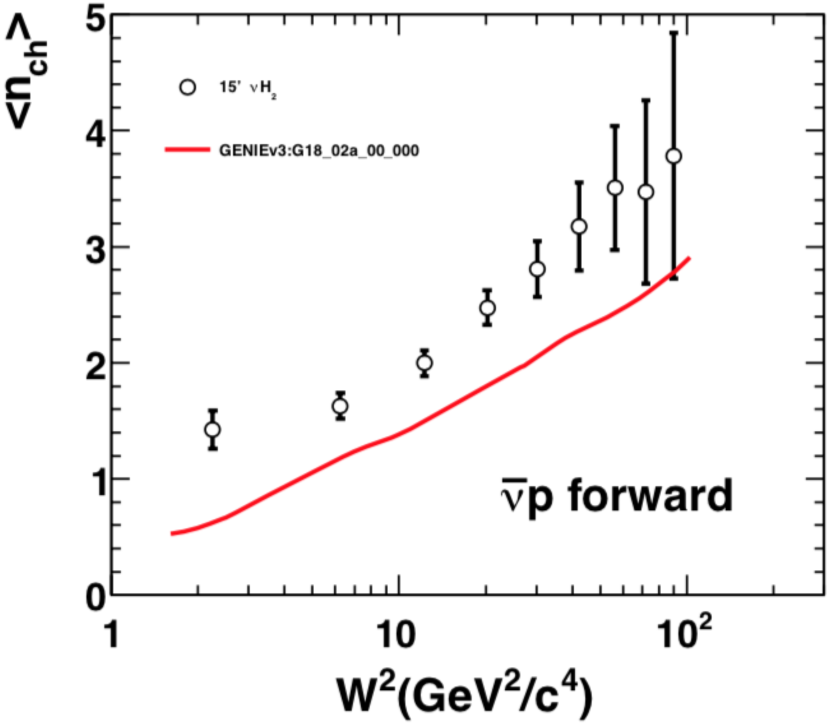
$$w_i = e^{-A \cdot p_T^i} \; (A \; = \; 3.5 \; GeV^{-1})$$

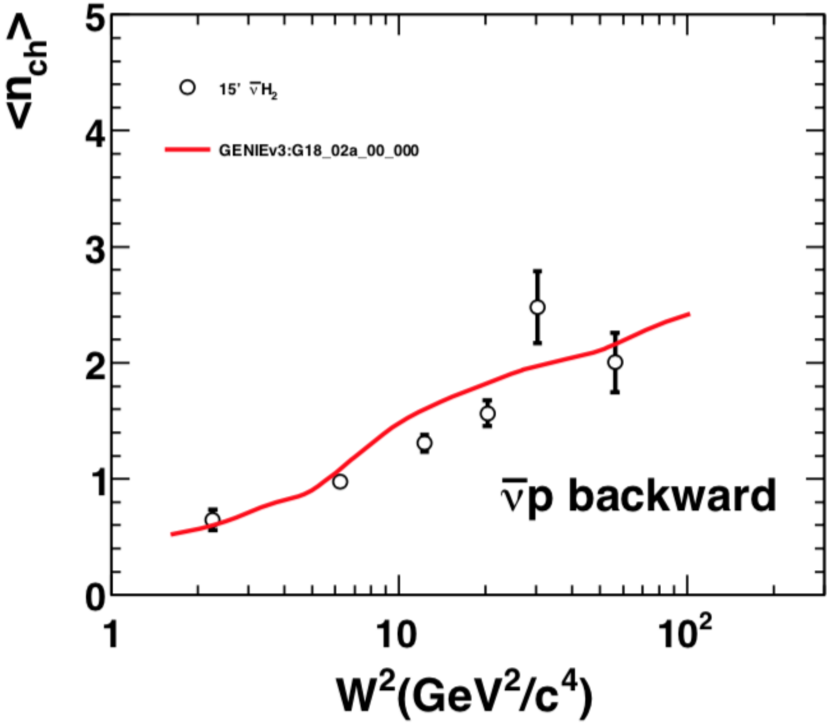
[Clegg and Donnachie, Description of Jet Structure by pt-limited Phase Space, Z. Phys. C 13: 71 (1982)]

GENIE comparisons with avg. fwd/bkw multiplicity data

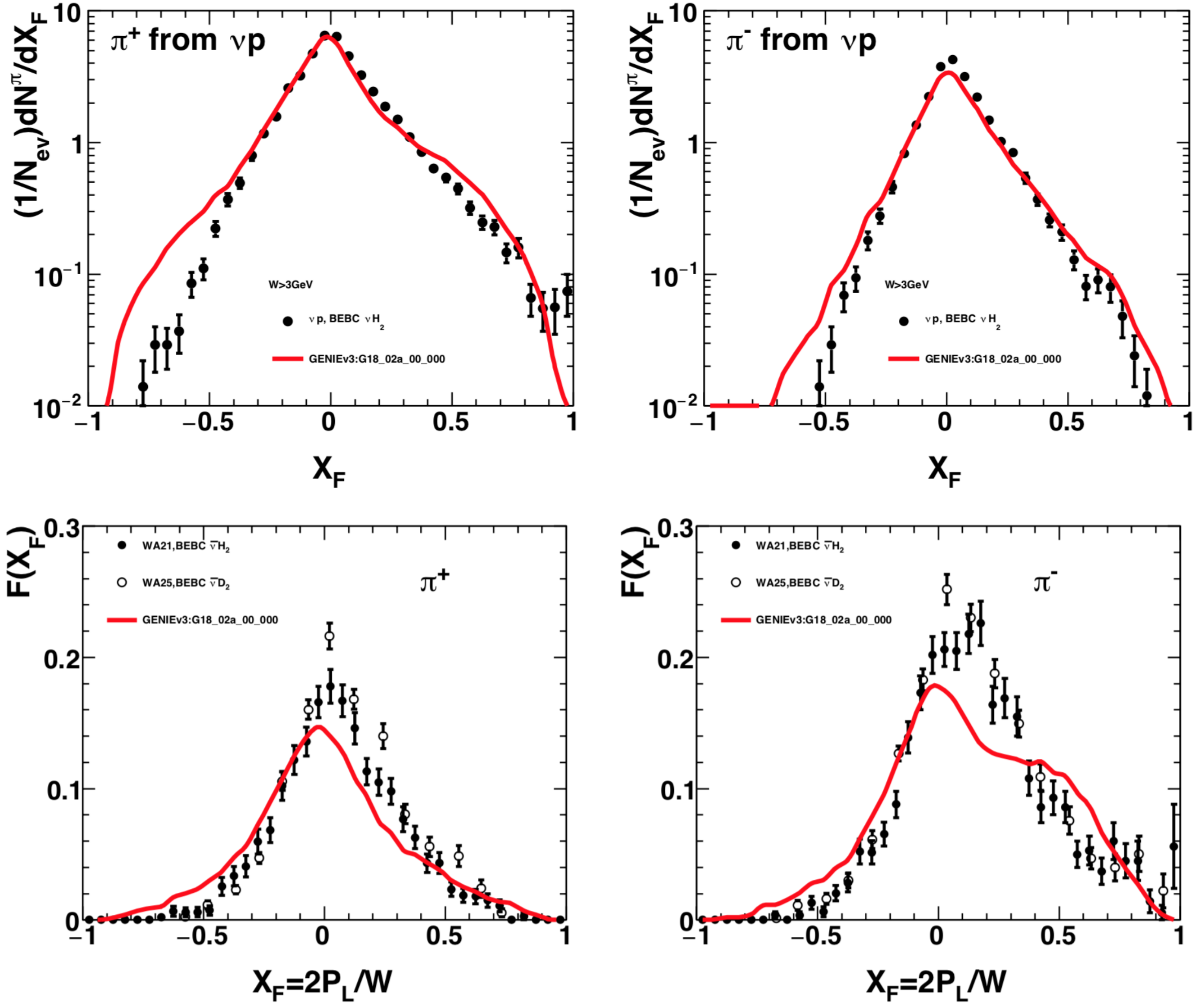


GENIE comparisons with avg. fwd/bkw multiplicity data

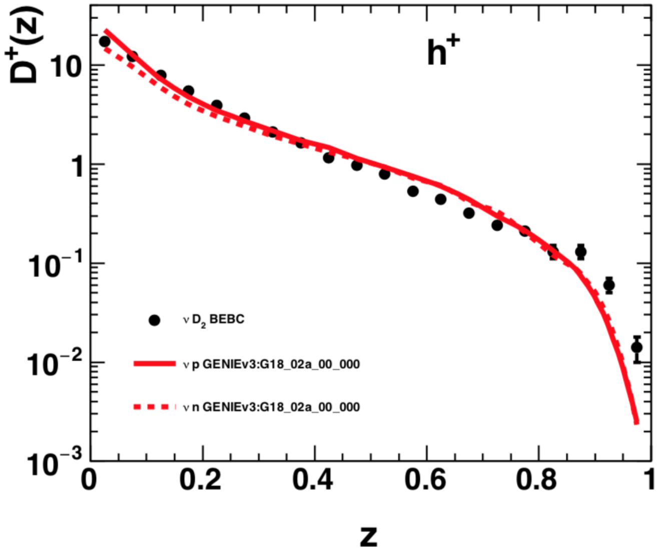


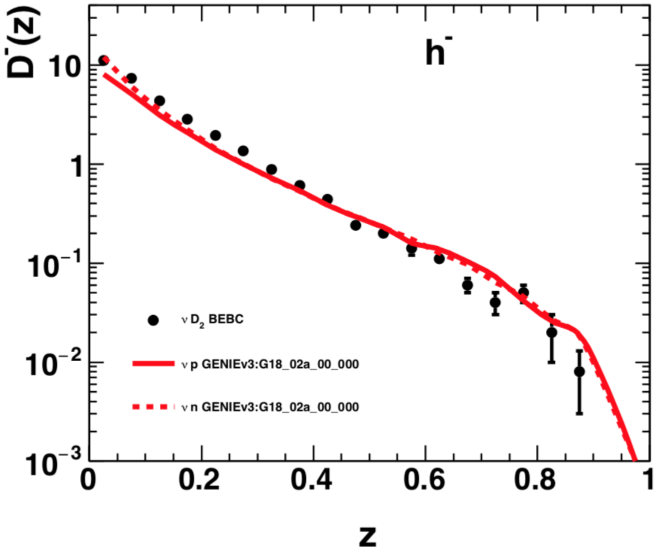


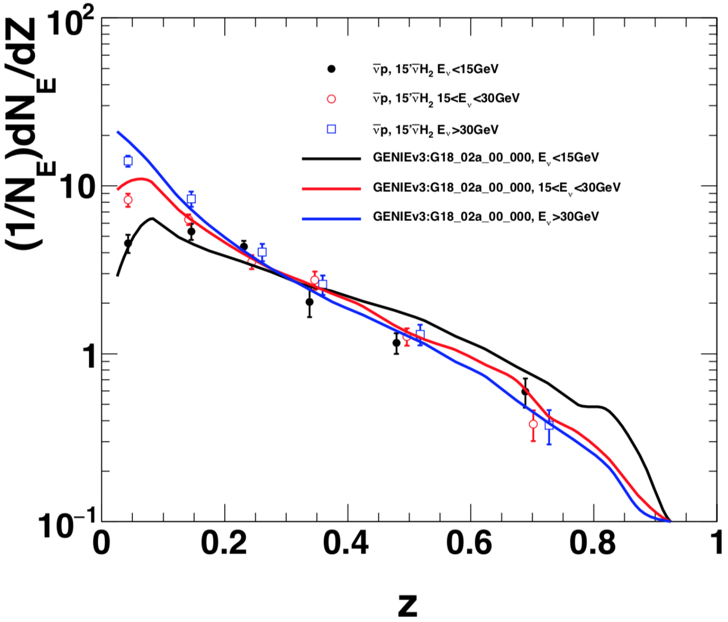
GENIE comparisons with x_F distributions



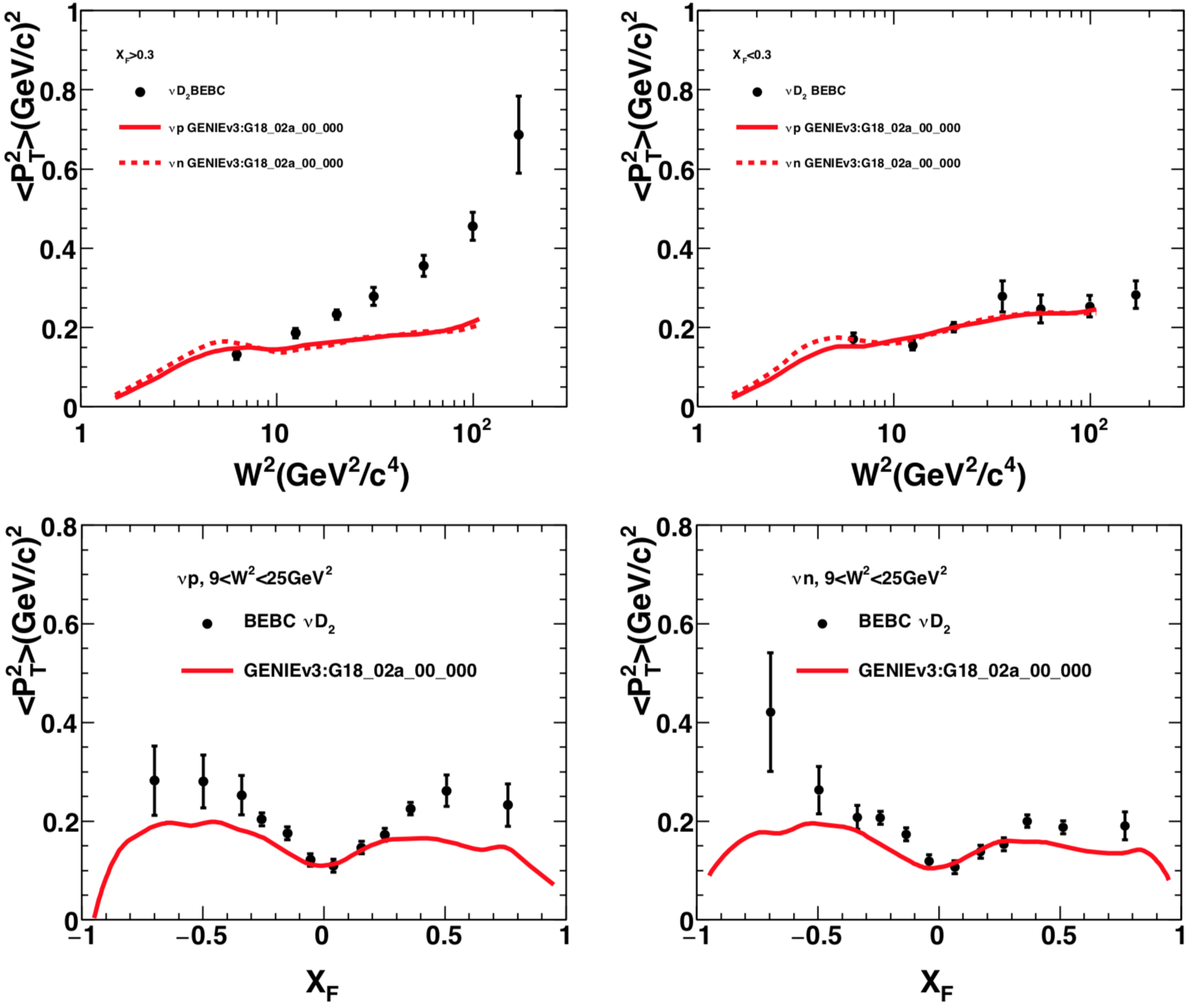
GENIE comparisons with z distributions







GENIE comparisons with p_T distributions



Empirical low-W model: Main caveats

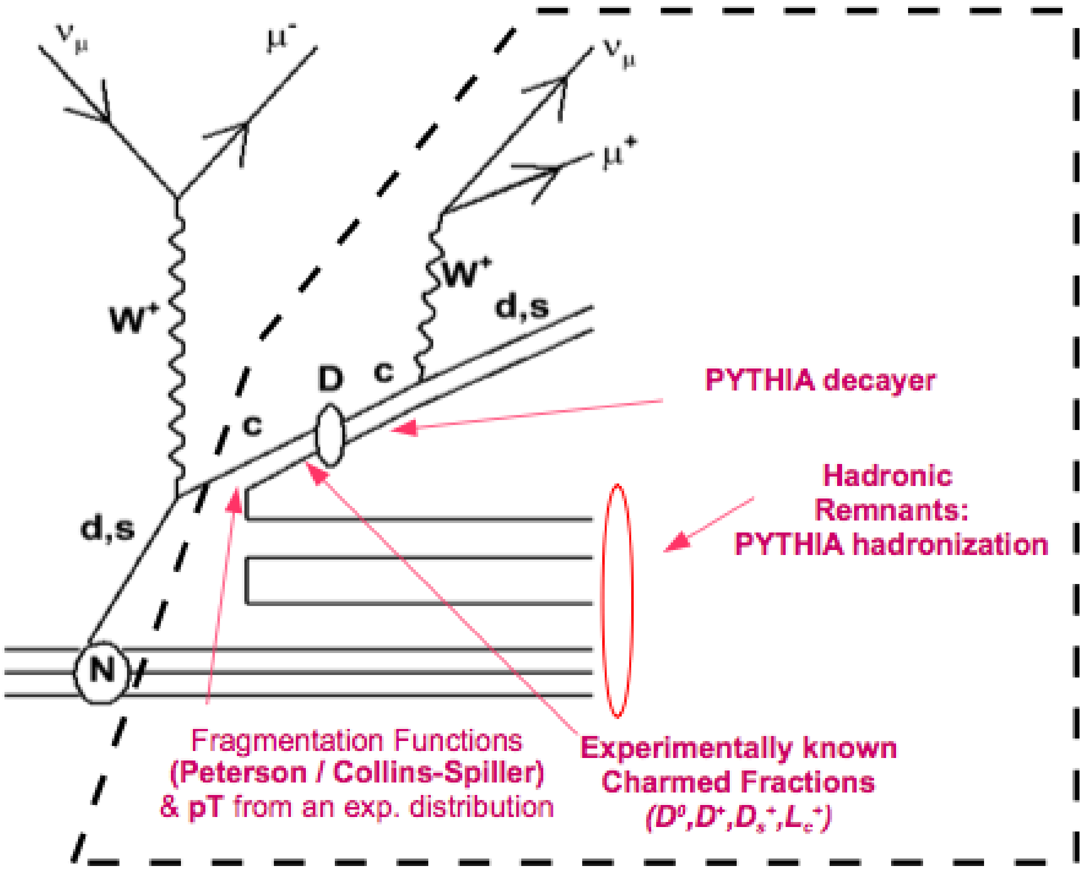
- Experimental contraint for neutral/charged pion multiplicity

  $$\frac{2 < n_{\pi^0} >}{< n_{\pi^+} > + < n_{\pi^+} >} \approx 1$$

  used as constraint on overall neutral/charged multiplicity.
- KNO scaling is applied for the total multiplicity distribution, rather for the charged multiplicity distribution.
- Too strong separation of target and current fragments at low W

DIS charm production

For DIS charm production, there is a separate hadronization model.



PYTHIA decayer

Hadronic Remnants: PYTHIA hadronization

Fragmentation Functions (Peterson / Collins-Spiller) & pT from an exp. distribution

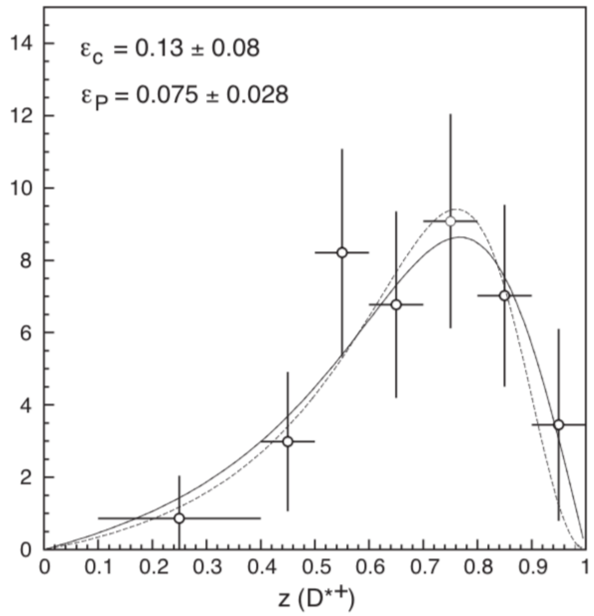
Experimentally known Charmed Fractions (D^0,D^+,D_s^+,L_c^+)

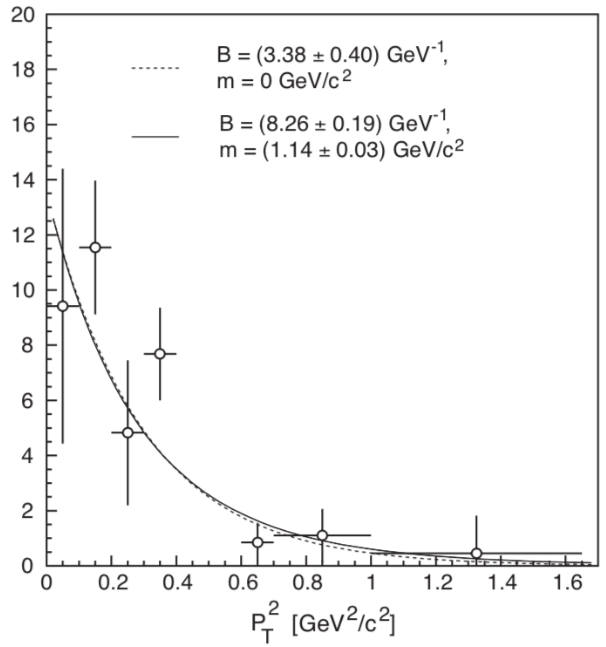
- Place the main **emphasis in the simulation of the charm hadron**, anchoring it to data.
- Generate the charm hadron energy using the Peterson or Collins-Spiller fragmentation functions
- Generate the charm hadron ID (D^0, $D^\pm$, $D_s^\pm$,Λ_c^+) using experimentally measured charm fractions
- Use PYTHIA for the remnant hadronic system

DIS charm production

G. De Lellis et al., Phys. Reports 399 (2004)

E_ν (GeV)	f_{D^0}	f_{D^+}	$f_{D_s^+}$	$f_{\Lambda_c^+}$
5–20	0.38 ± 0.13	0.05 ± 0.06	0.17 ± 0.09	0.41 ± 0.11
20–40	0.56 ± 0.09	0.09 ± 0.07	0.19 ± 0.07	0.16 ± 0.06
40–80	0.69 ± 0.09	0.19 ± 0.08	0.08 ± 0.07	0.04 ± 0.01
> 80	0.66 ± 0.12	0.26 ± 0.10	0.00 ± 0.06	0.08 ± 0.07
> 40	0.67 ± 0.07	0.23 ± 0.03	0.03 ± 0.01	0.06 ± 0.02
> 30	0.68 ± 0.07	0.19 ± 0.06	0.07 ± 0.05	0.07 ± 0.04
> 20	0.63 ± 0.06	0.18 ± 0.04	0.10 ± 0.04	0.10 ± 0.04
> 5	0.59 ± 0.06	0.14 ± 0.04	0.11 ± 0.04	0.15 ± 0.04





DIS charm production

PYTHIA6 (di)quark assignments for the hadronization of the remnant system:

Init state	Pre-fragm quarks	Selected charm hadron	Final state charm hadron + [PYTHIA pre-fragm inputs]
νp CC	uu(d$\rightarrow$)c	D^0	c$\bar{u}$ (D^0) + [uu + u]
νp CC	uu(d$\rightarrow$)c	D^+	c$\bar{d}$ (D^+) + [uu + d]
νp CC	uu(d$\rightarrow$)c	D_s^+	c$\bar{s}$ (D_s^+) + [uu + s]
νp CC	uu(d$\rightarrow$)c	Λ_c^+	cud(Λ_c^+) + [uu + $\bar{u}\bar{d}$]
νn CC	ud(d$\rightarrow$)c	D^0	c$\bar{u}$ (D^0) + [ud + u]
νn CC	ud(d$\rightarrow$)c	D^+	c$\bar{d}$ (D^+) + [ud + d]
νn CC	ud(d$\rightarrow$)c	D_s^+	c$\bar{s}$ (D_s^+) + [ud + s]
νn CC	ud(d$\rightarrow$)c	Λ_c^+	cud(Λ_c^+) + [ud + $\bar{u}\bar{d}$]
νp CC	uud + $\bar{d}$(d$\rightarrow$)c	D^0	c$\bar{u}$ (D^0) + [uud + $\bar{d}$ + u = uu + u]
νp CC	uud + $\bar{d}$(d$\rightarrow$)c	D^+	c$\bar{d}$ (D^+) + [uud + $\bar{d}$ + d = uu + d]
...	...	...	...
...	...	...	...

In-medium effects to hadronization

- QCD version of the Landau-Pomeranchuck-Migdal effect

  The QED effect [A.Migdal,Phys.Rev.103,1811(1956)] was first observed at SLAC/E-146 [P.Anthony et al, Phys.Rev.Lett.75, 1949 (1995)]

  Leads to: **Smaller parton cross section within a nuclear medium**

- Cronin effect

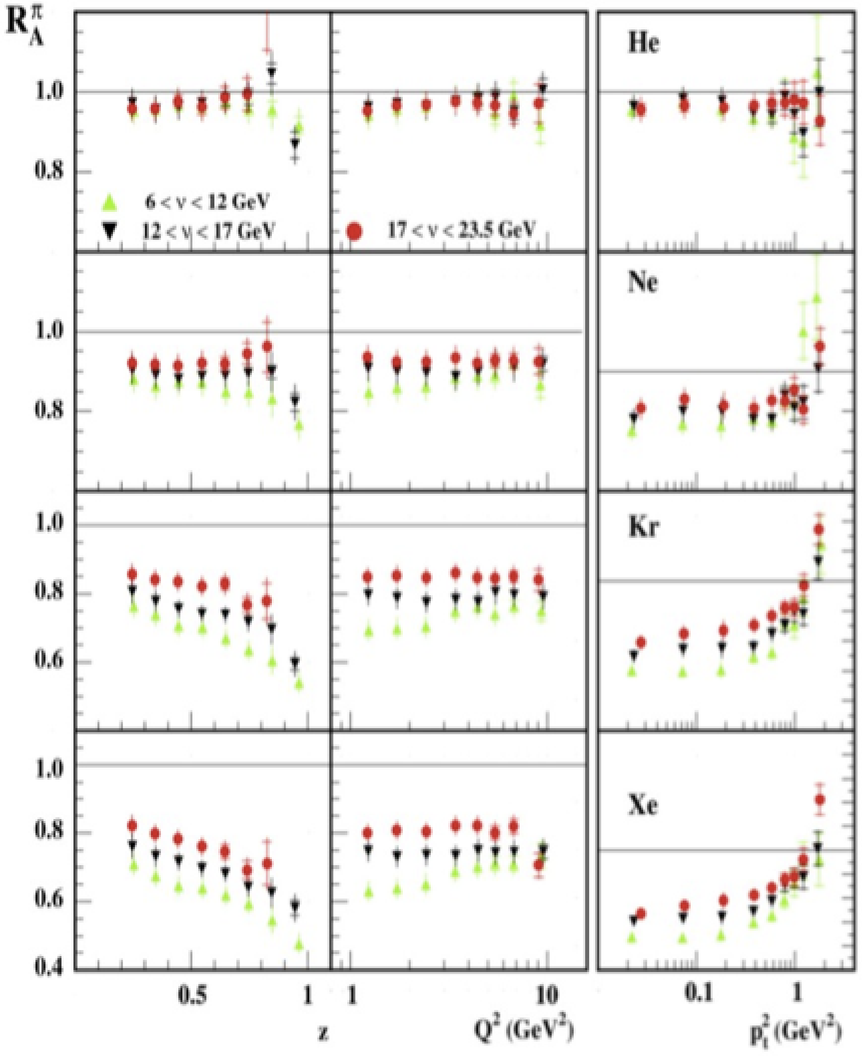
  First observed at pA interactions at Fermilab [J.W.Cronin et al.,Phys.Rev.D 11, 3105 (1975)]

  Leads to: **Broader pT spectrum**

In-medium effects to hadronization

- First studied using e- DIS at SLAC
  [L.S.Osborne et al., Phys.Rev.Lett. 40, 1624 (1978)]
- Similar experiments with muon beams at CERN (EMC)
  [J. Ashman et al, Z.Phys.C52:361-388 (1991)]
- ... and Fermilab (E665)
  [M.R.Adams et al, Phys.Rev.D50, 1836 (1994)]
- Recent data (e-, e+) by HERMES at DESY
  [A.Airapetian et al, Euro.Phys.J.C.20,479 (2001)]
- ... and (e-) by CLAS at JLAB
  [K.Hafidi, hep-ex/0609005]

In-medium effects to hadronization

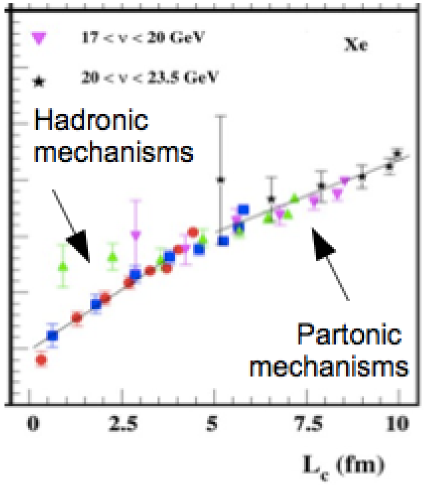


Airapetian et al., Nucl.Phys.B 780 (2007) 1-27,

A number of very interesting, quantitative results of direct interest to ν MC

Airapetian et al., Nucl.Phys.B 780 (2007) 1-27, 2007
Harlem et al., arXiv: 0704.3712

- Attenuation increases with increasing A
- Attenuation decreases with increasing ν
- Attenuation increases with increasing z
- Attenuation increases slightly with Q2 (pions)
- Attenuation almost independent of pT
    - except for large pT's (Cronin effect)

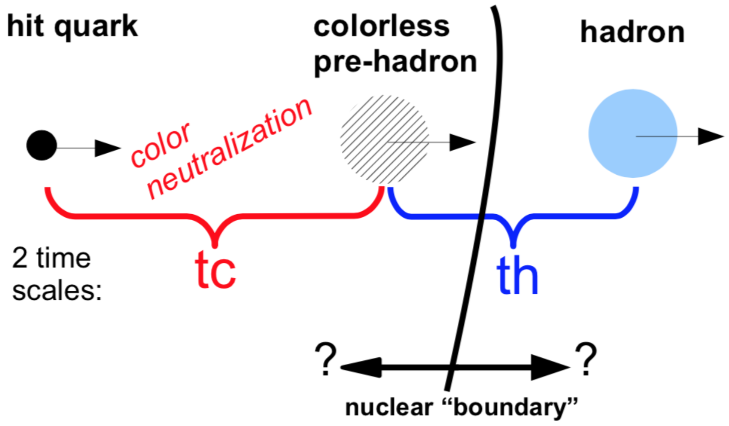


<u>Scaling !</u>

$$L_c = z^{0.35}(1 - z)\nu/\kappa$$

In-medium effects to hadronization

Crucial point:
Hadronization time scales?



Several models:

- B.Andresson et al., Phys.Rep.97, 31 (1983)
- A.Bialas, Acta Phys Pol B 11, 475 (1980)
- M.Gyulassy et al., Nucl. Phys.B 346, 1 (1990)
- J.Cryzewski et al., Z.Phys.C56, 493 (1992)
- N.Akopov et al., Eur.Phys.J C44, 219 (2005)
- X.F.Guo et al., Phys.Rev.Lett.85, 3591 (2000)
- E.Wang et al., Phys.Rev.Lett.89, 162301 (2002)
  F.Arleo, Eur.Phys.J C30, 213 (2003)
- A.Accardi et al., Nucl.Phys.A 761, 67 (2005)
- B.Z.Kopeliovich, Nucl.Phys.A740, 211 (2004)
- T.Falter et al., Phys.Rev.C70, 054609 (2004)

In GENIE, in-medium effects to handronization are included using a single formation time of 0.342 fm/c.

We would welcome any collaborative effort to go beyond that!

Next steps for hadronization modelling in GENIE

- Showed lots of data and empirical models with lots of parameters!
- GENIE now has a very capable **global analysis of neutrino data**. (Exercised in a retune of the free-nucleon x-section model, and in a global fit of nuclear 0π and 1π x-section data)
  - Relying GENIE/Comparisons product and its interfaces with Professor.
  - **Reduces computational complexity of brute-force tuning** and allows for **massive parallelisation**.
  - Can handle a **broad range of (non-reweightable) model uncertainties**.
- Understanding systematics of data shown on this talk and incorporating them in our global analysis is one of our next priorities.
- Obtain: Improved model parameters, and a **handle on model systematics** / data-driven **parameter correlations**.

Summary

- Hadronization model in GENIE
  - Simple and with several caveats, but...
  - As evidenced from data/MC comparisons, it is quite robust
  - Get's right lots of hadronic shower features
- Would love to hear inputs for the improvement of the model

- GENIE 3 was released earlier this week
  - First major new release in more than a decade!
  - An overhaul - touched all modelling aspects (except hadronization!)
  - Implemented a powerful new GENIE global analysis/fit of scattering data (and used it to inform new GENIE tunes)
- Hadronization modelling and tuning become next GENIE priorities.