UNIVERSITY OF LIVERPOOL · DEPARTMENT OF PHYSICS

First-Year PhD Report

Hadronization in the Nuclear Medium with GENIE

Constraining the formation zone using HERMES electron–nucleus data

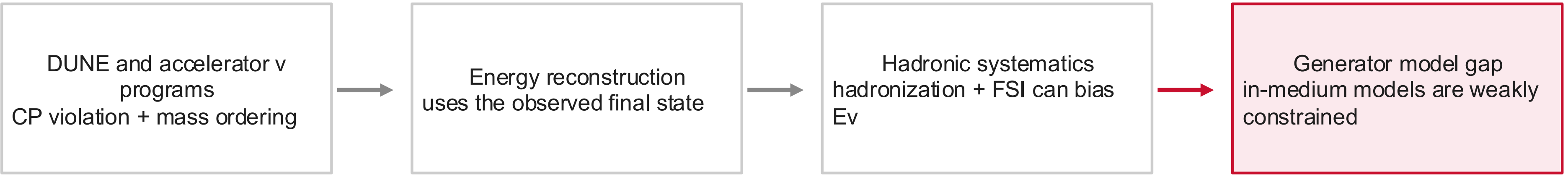
Qianying Yu (ID 201876661)

Supervisors: Prof. Costas Andreopoulos · Dr. Marco Roda · Dr. Julia Tena Vidal

June 2026

Why this study is needed

Neutrino precision depends on the reconstructed final state



DUNE and accelerator v programs
CP violation + mass ordering

Energy reconstruction
uses the observed final state

Hadronic systematics
hadronization + FSI can bias Ev

Generator model gap
in-medium models are weakly constrained

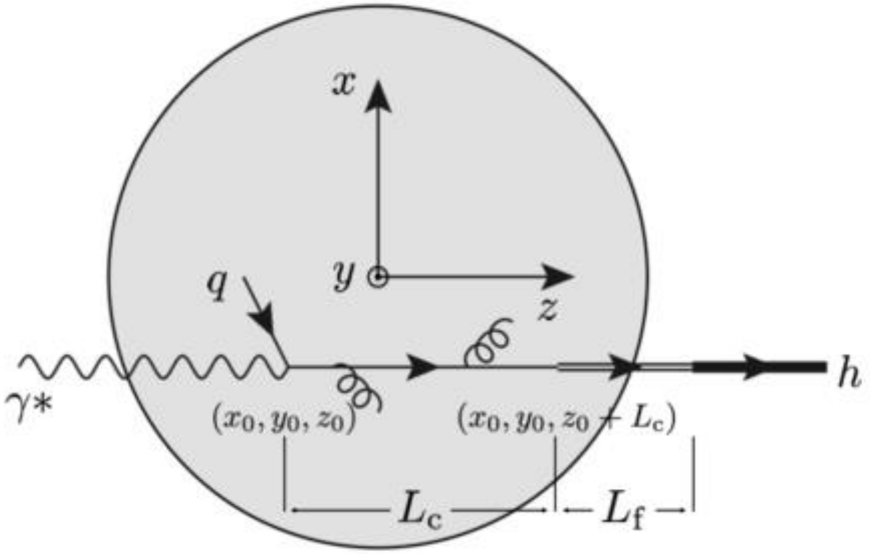
Why electron data?
Electron beams give known incident energy, clean lepton kinematics and high-statistics identified-hadron yields. They isolate the nuclear response after the primary interaction, so the same hadronization/FSI components can be stress-tested before being used in neutrino analyses.

What is GENIE?
GENIE is a Monte Carlo event generator used to simulate lepton–nucleus interactions. In this work the key point is not “can software do it?”, but what model ingredients must be exposed, validated and tuned so electron–nucleus constraints can improve neutrino simulations.

This work asks: what needs to happen in GENIE for electron–nucleus data to constrain hadron formation and final-state interactions?

Aim: validate and tune the shared hadronization + FSI chain with controlled electron–nucleus data.

What controls hadron attenuation?



Report Fig. 2.1 · nSIDIS space–time geometry

Colour length Lc (τc)
Distance travelled by the struck coloured object before colour neutralisation. It is mainly probed through transverse-momentum broadening.

Formation length Lf (τf)
Distance over which the prehadron becomes a physical hadron. Before full formation, its effective FSI cross section can be reduced.

Observable used later: nuclear multiplicity ratio

$R^h(A)$ = hadron yield per DIS event in nucleus A / deuterium

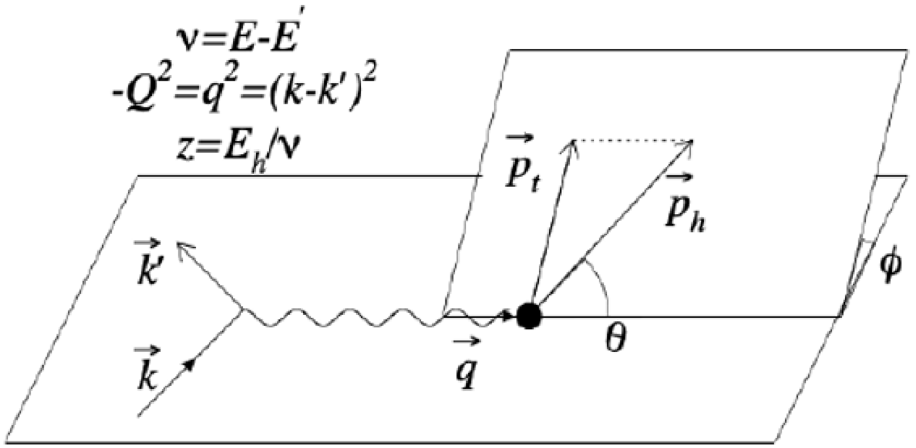
$R^h < 1$ means attenuation.

A: nuclear target; D: deuterium reference; h: identified hadron species.

Longer formation length ⇒ less re-scattering inside the nucleus ⇒ larger R^{Ah}.

The HERMES experiment

Report Fig. 2.2 · SIDIS kinematics used by HERMES



$$R_A^h(\nu,Q^2,z,p_T^2) = \frac{\left[N^h(\nu,Q^2,z,p_T^2)\,/\,N^e(\nu,Q^2)\right]_A}{\left[N^h(\nu,Q^2,z,p_T^2)\,/\,N^e(\nu,Q^2)\right]_D},$$

N^h: number of semi-inclusive hadrons of type h in a given (ν, Q^2, z, p_t^2) bin.
N^e: number of inclusive DIS leptons in the corresponding (ν, Q^2) bin.
ν is the virtual-photon energy;
Q^2 is its negative four-momentum squared;
$z = E_h/\nu$ is the hadron energy fraction;
p_t is the hadron momentum transverse to the virtual-photon direction.
The HERMES definition integrates over the azimuthal angle between the lepton and hadron planes.

HERMES experimental overview

27.6 GeV e^+/e^- beam at DESY

SIDIS measurements on D, He, Ne, Kr and Xe targets

Data collected from 1997–2005

Identified $\pi^\pm$, π^0, $K^\pm$, p and $\bar{p}$ final states

Measured R^{Mh} and $\Delta\langle p_t^2\rangle$ vs ν, z, Q^2 and p_t^2

Why HERMES is ideal for GENIE validation

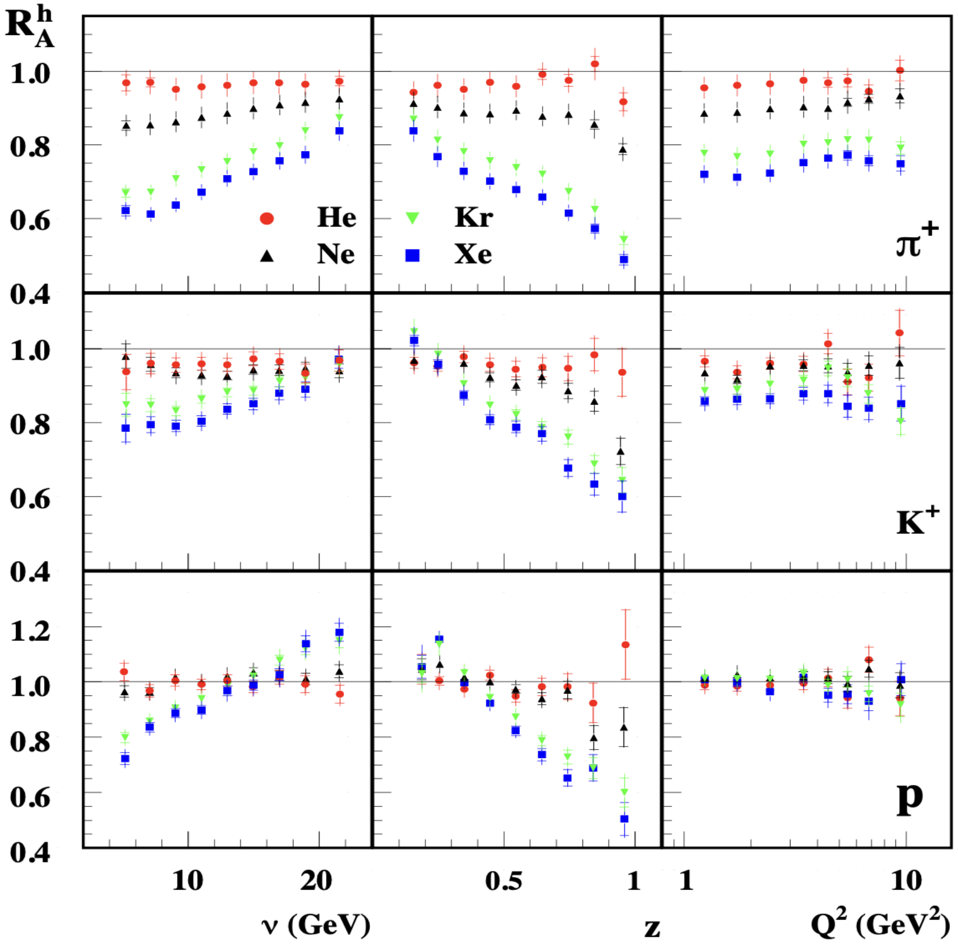
Electron mode in GENIE
GENIE can simulate electron–nucleus scattering as well as neutrino scattering, so the same framework can be validated against e–A data.

Shared model components
Hadronization, formation-zone treatment and INTRANUKE FSI are common pieces of the event chain that affect the final-state hadrons.

Direct validation target
HERMES gives identified-hadron attenuation ratios on multiple nuclei, providing a clean benchmark before applying the models to neutrino analyses.

Kr/D is used first because the attenuation signal is large and the published ratio directly compares a heavy target with the deuterium reference.

HERMES observables & the nuclear mechanism



HERMES multiplicity ratios R_M^h for identified hadrons.

Y-axis: R_M^h (1 = no attenuation). X-axes: ν, z, pT² and Q².

How to read the plot

Each point is a hadron yield ratio on a nuclear target relative to deuterium. Values below 1 mean absorption or energy loss in the nucleus, means attenuation.

Main trends to reproduce

R^h falls for heavier nuclei and high z, rises toward unity at larger ν, and differs by hadron species. These are shape constraints, not just a normalization test.

Origin of the shape

Low ν → shorter formation length and longer in-medium path, giving stronger rescattering and lower R^h.

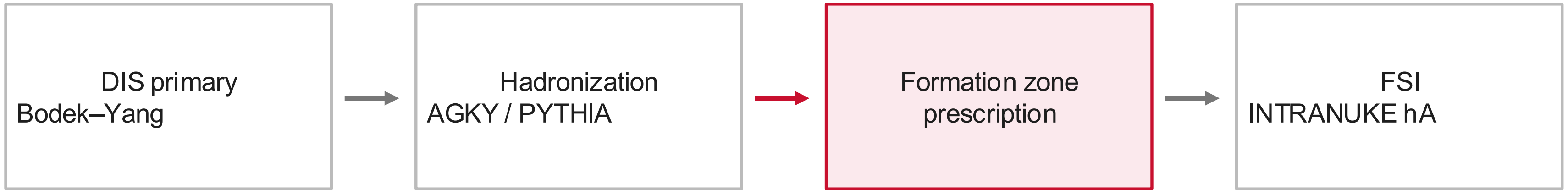
High z → the leading hadron carries most of the photon energy, so energy loss or absorption produces a larger suppression.

Q^2 trend → weaker because the attenuation is mainly controlled by post-scattering formation and FSI rather than by the hard scale itself.

GENIE should describe both the size of the attenuation and its dependence on ν, z, pT² and Q².

Qianying Yu · University of Liverpool · Hadronization in the Nuclear Medium with GENIE

GENIE chain used in the report



DIS primary
Bodek–Yang

Hadronization
AGKY / PYTHIA

Formation zone
prescription

FSI
INTRANUKE hA

Simulation setup
GENIE v3.06.02
Configuration/tune:
GEM21_11a_00_000
Beam: 27.6 GeV electrons
Targets: D and Kr-84
Species: $\pi^\pm$, π^0, $K^\pm$, p, $\bar{p}$
Observable: R_M^h with
HERMES cuts

formation-zone prescription
GENIE default option that delays when newly produced hadrons can undergo full final-state interactions.

GENIE tune / configuration
A named, reproducible model-and-parameter set. A GENIE tune is a predefined set of physics-model choices and parameter values. Here, GEM21_11a_00_000 fixes the baseline DIS, hadronization, formation-zone and FSI settings used before any dedicated HERMES tuning.

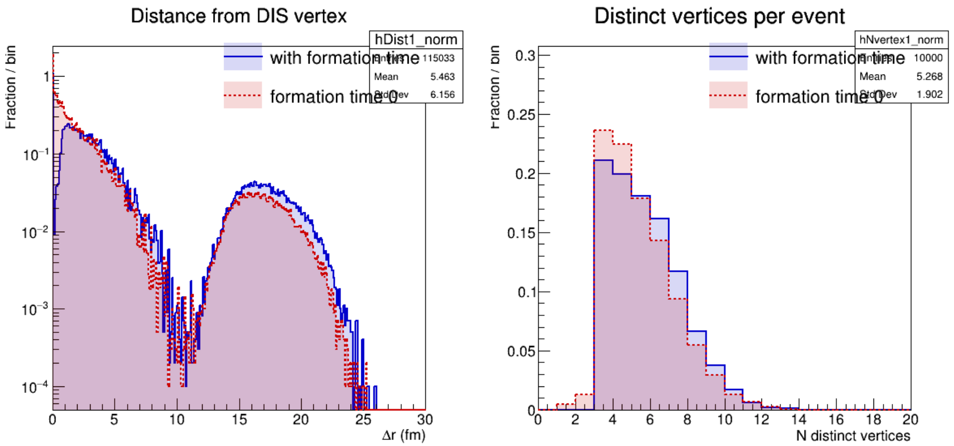
INTRANUKE hA
Effective intranuclear-cascade model for final-state interactions of produced hadrons inside the nucleus.

Why cuts matter
Apply the same DIS and hadron acceptance cuts to data and simulation so R^h reflects a nuclear effect rather than phase-space bias.

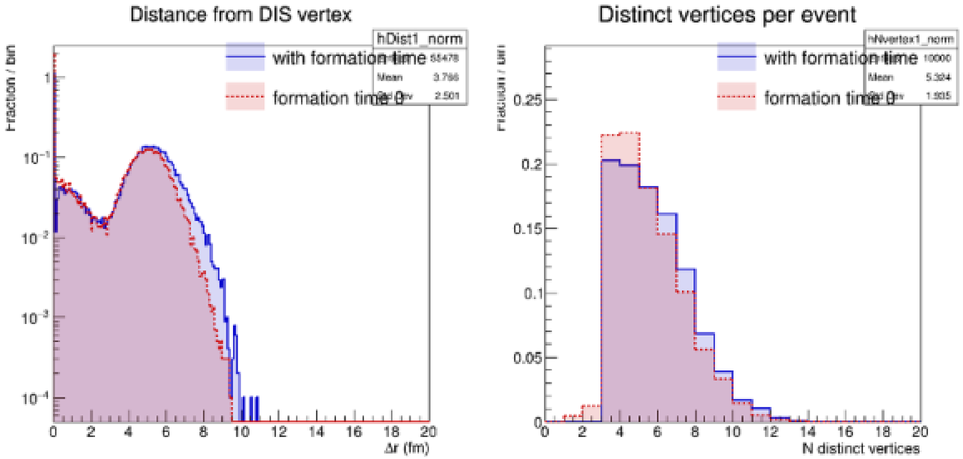
The formation-zone model controls when hadrons start to re-interact; the hA model controls what those re-interactions do.

Result 1: spatial structure of FSI

Formation zone pushes FSI to the nuclear surface



Report Fig. 3.3 · e–Kr vertex-distance distribution



Report Fig. 3.2 · e–D comparison

displacement ⟨Δr⟩ from DIS vertex

Interpretation

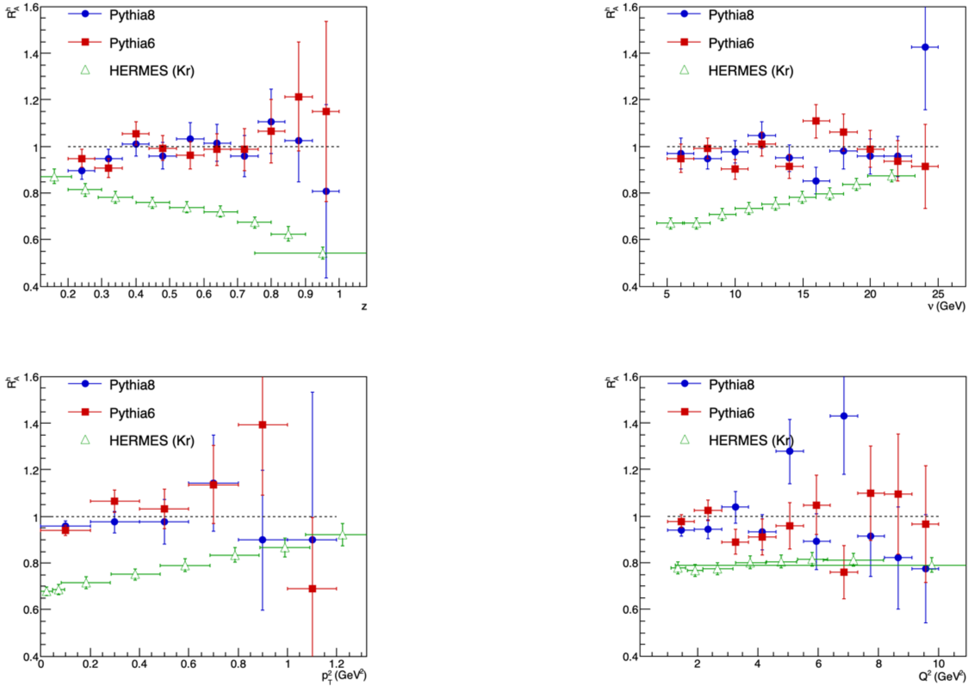
Default formation time: hadrons travel before full FSI can occur. In Kr, ⟨Δr⟩ ≈ 5.09 fm, close to the nuclear radius (~5.3 fm), so re-scattering is pushed towards the surface and the nucleus becomes too transparent.

Formation time = 0: hadrons interact immediately after the DIS vertex. In Kr, ⟨Δr⟩ drops to ≈ 3.19 fm, so FSI occurs deeper inside the nucleus and the effective in-medium path length is larger.

The two cases bracket the data: the default delay is too long, while zero formation time is too absorptive.

Formation time changes where interactions occur, not how many distinct vertices appear.

Hadronization model choice matters — but it is not enough



Report Fig. 3.5 · π^+ Kr/D, PYTHIA6 vs PYTHIA8 vs HERMES

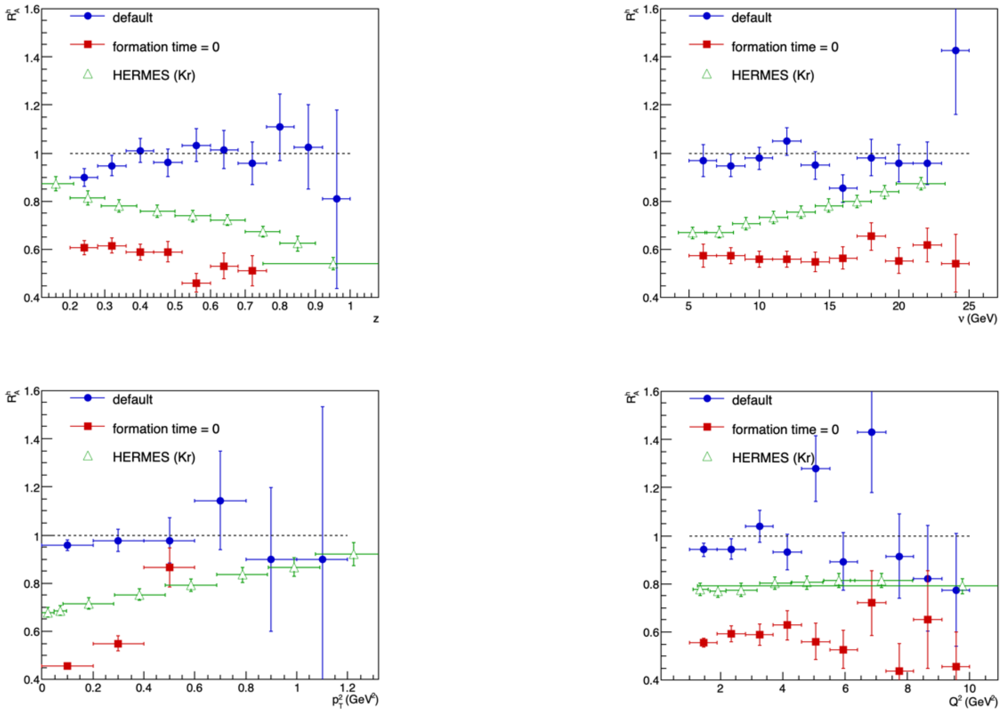
PYTHIA8
predicts stronger attenuation and is used as the baseline for the first GENIE comparison.

Both models
remain too close to $R^h = 1$ and miss much of the 10–30% HERMES suppression.

What this means
Hadronization model choice is a visible systematic, but changing PYTHIA alone does not solve the missing nuclear attenuation.

The disagreement is not only normalization: the kinematic dependence is also too weak.

Formation time is the dominant control parameter



Report Fig. 3.7 · π^+ default vs zero formation time, compared with HERMES

Default formation time
too transparent: R_M^h ≈ 0.90–0.95

formation time = 0
too absorptive: downward shift ≈ 0.30–0.35

Next model
Replace constant formation time with a momentum-dependent formation length, with τ_f tuned to data rather than fixed by default.

Next: tune HERMES + CLAS, extend to He/Ne/Xe, and reduce DUNE hadronization + FSI systematics.

Main conclusion: HERMES lies between the two limits — τ_f is non-zero, but shorter than the GENIE default.

The formation zone is the priority to fix

What we found

- **GENIE default is too transparent.** It underestimates the nuclear attenuation measured by HERMES.
- **The formation-zone delay is not suitable.** FSI are active, but the default delay pushes many interactions near the Kr surface; zero formation time shifts them inward and becomes too absorptive.
- **The data define a useful bracket.** HERMES lies between the default and zero-formation-time limits, so formation time should be made tunable and constrained with e–A data.

Possible plan: physical formation length

Replace fixed formation time with kinematics-dependent formation length

- **1. Implement / expose:** make formation-zone parameters configurable in GENIE electron mode.
- **2. Tune with e–A data:** fit HERMES R^h and $\Delta\langle p_t^2\rangle$ with the same DIS and hadron cuts.
- **3. Cross-check:** test the tuned model on CLAS and on He, Ne, Xe, not only Kr/D.
- **4. Propagate uncertainty:** provide GENIE systematic variations for neutrino analyses.

Goal: a HERMES-constrained formation-zone model that reduces hadronization + FSI systematics for neutrino simulations.