Highly granular way to a new level of validation of simulations

Marina Chadeeva

LPI / MEPhI

September 10, 2021

Outline

1. Motivation, samples and event selection
2. MC-truth variables and calorimetric observables
3. Relationship between parameters of secondaries in a shower and calorimetric observables

Motivation and approach

Validation of Geant4 hadronic models

Geant4 uses material properties from thin-sample measurements and theor./phenomenological models. Additional tuning is performed using test beam data. Geant4 simulations provide:

- percent-level precision for simulations of ionisation losses and electromagnetic shower properties;
- good prediction of main calorimetric observables for hadrons (e.g. measured hadron energy) though discrepancies increase with energy.

Particle Flow reconstruction relies also on topological observables in highly granular calorimeters:

- position of first inelastic interaction for clustering
- shower radius to disentangle close-by showers
- shower profiles for particle identification and separation

Data-MC discrepancies in topological characteristics are larger than for traditional calorimetric observables. Reconstruction of intrinsic shower parameters might help to understand the source of disagreement.

Machine-learning-based approach to study properties of secondaries in a shower

- select calorimetric observables correlated with parameters of secondaries at generation level (MC truth)
- use machine learning technique to train regression model with calorimetric input and predict MC truth
- apply the trained model to data to estimate the characteristic/parameter under study

Samples and event selection

Data from beam tests of technological CALICE AHCAL at CERN as of June 2018

- single negative pions, 10–80 GeV
- reconstruction software v04-14
- official PID (Vladimir's BDT)

Run numbers and pion energies

10 GeV	20 GeV	30 GeV	40 GeV	60 GeV	80 GeV
61265	61272	61378	61275	61262	61279

MC samples (many thanks to DESY group)

- centrally generated negative pion samples, 10–80 GeV, about 500 kevt / sample
- Geant4 v10.3, physics lists: **FTFP_BERT_HP** and **QGSP_BERT_HP**
- official digitisation, no PID

Reconstruction and event selection

- official reconstruction chain, 0.5 MIP cut for hits
- official shower start finder algorithm
- for analysis: only events with found shower start at 3–6 AHCAL layers
- no other constraints, no clustering

$\sim$20 kevt / data sample after selections; to compare with data, MC samples are truncated accordingly.

MC-truth variables

Parameters of secondaries at generation level are extracted from MCParticle collection.

In this context "secondaries" means also tertiaries, etc.
All parameters are kept on an event-by-event basis.
Parameters of primary particle (type, energy and direction/position) are set as Geant4 input.

Two types of secondaries are considered, which are produced in a hadronic shower:

Neutral pions

are counted independently of their parents (some of them might be from η mesons)

- **Number of neutral pions in an event**, N_{π^0}
- **Sum of the energies of neutral pions**, E_{π^0}

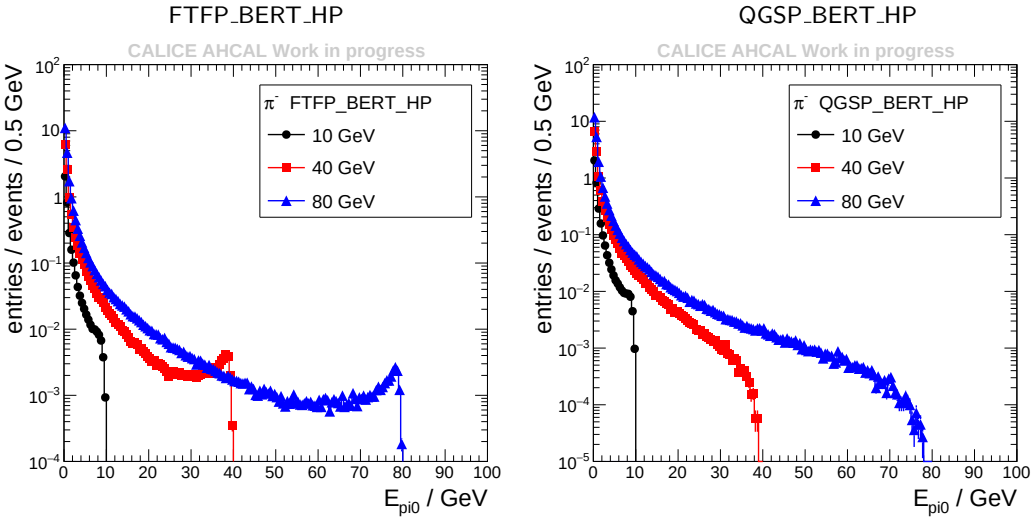
Neutrons

are counted except for those that have one parent only that is also neutron (to avoid double counting)

- **Number of neutrons from interactions**, $N_{neutron}$
- **Sum of kinetic energies of neutrons from interactions**, $T_{neutron}$

Energy spectra of neutral pions

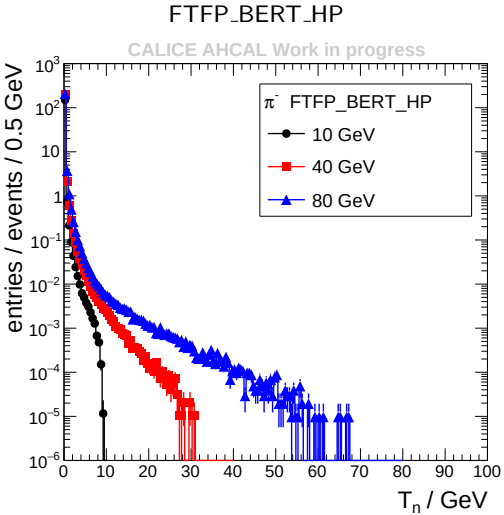
Legend: **10 GeV**, **40 GeV**, **80 GeV** (full MC sample, ∼100 kevt / sample after selections)

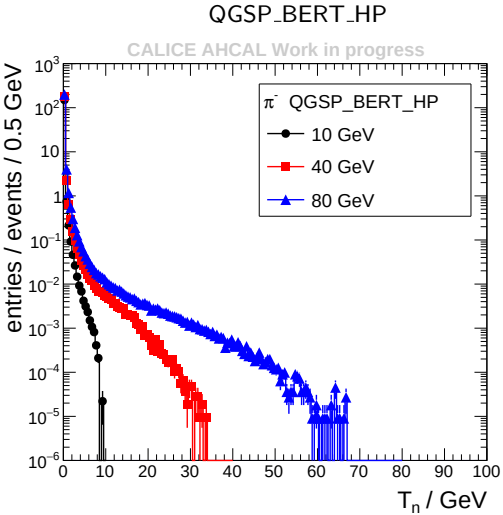


Very different spectra from FTFP and QGSP models above 10 GeV.
Similar behaviour at 10 GeV due to the same Bertini model (BERT) in this energy range.

Spectra of neutron kinetic energy

Legend: **10 GeV**, **40 GeV**, **80 GeV** (full MC sample, ∼100 kevt / sample after selections)





Different shape of high energy tails from FTFP and QGSP models. Neutron energy cut is set to 1 MeV.
Similar behaviour at 10 GeV due to the same Bertini model (BERT) in this energy range.

Calorimetric observables

Values are calculated from hits beyond the found shower start layer, except for N_{hits} and E_{reco}.

Counting observables

- **Total number of hits**, N_{hits}
- **Number of isolated hits**, N_{iso} [isolation – 0 neighbours in a cube of 3×3×3 cells around the hit]
- **Number of track hits**, N_{trk} [defined as having 2 in-line neighbours and MIP-like deposition]

Amplitude observables (e_i - energy of hit with coordinates x_i, y_i, z_i; N_{sh} - number of shower hits)

- **Reconstructed energy**, E_{reco} (sum of hit energies)
- **Mean shower hit energy**, $< e_{\mathrm{hit}} >$
- **Shower radius** $R_{\mathrm{sh}} = \frac{\sum_{i=1}^{N_{\mathrm{sh}}} e_i \cdot r_i}{\sum_{i=1}^{N_{\mathrm{sh}}} e_i}$, , $r_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$ - hit radial distance from shower axis (x_0,y_0)
- **Longitudinal shower centre of gravity** $Z_{\mathrm{CoG}} = \frac{\sum_{i=1}^{N_{\mathrm{sh}}} e_i \cdot (z_i - z_{\mathrm{start}})}{\sum_{i=1}^{N_{\mathrm{sh}}} e_i}$,
  z_i - hit longitudinal coordinate, z_{start} - longitudinal coordinate of shower start

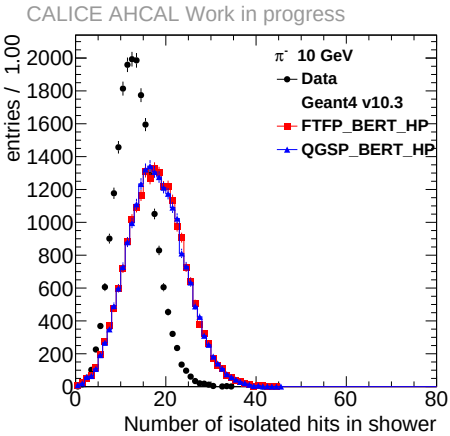
Additional "ring" observables (integrated over longitudinal depth)

3-cm wide rings around shower axis, 12 rings in total

- **number of hits in a ring**, $N_{\mathrm{all}}^{\mathrm{ring}}$; **number of isolated hits in a ring**, $N_{\mathrm{iso}}^{\mathrm{ring}}$
- **energy sum in a ring**, $E_{\mathrm{all}}^{\mathrm{ring}}$; **energy of isolated hits in a ring**,$E_{\mathrm{iso}}^{\mathrm{ring}}$

Number of isolated hits vs. number of neutrons at 10 GeV

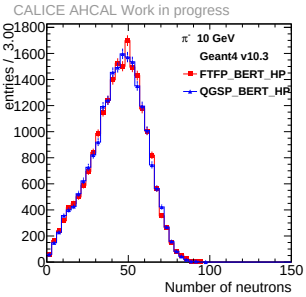
Data, FTFP_BERT_HP, QGSP_BERT_HP

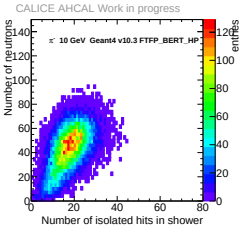


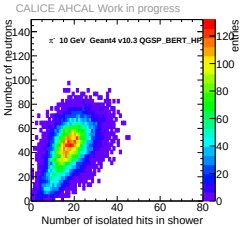
Larger number of isolated hits in simulations

Good agreement between physics lists as expected

MC truth: number of neutrons

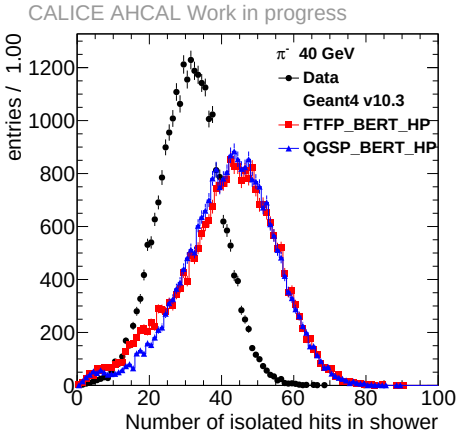




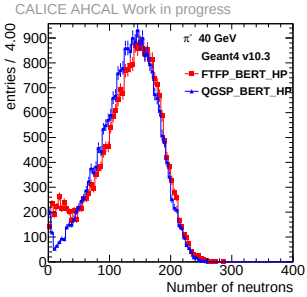


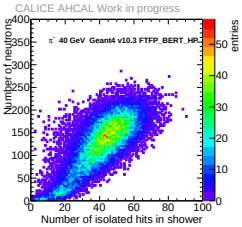
Number of isolated hits vs. number of neutrons at 40 GeV

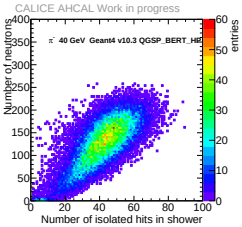
Data, FTFP_BERT_HP, QGSP_BERT_HP



MC truth: number of neutrons





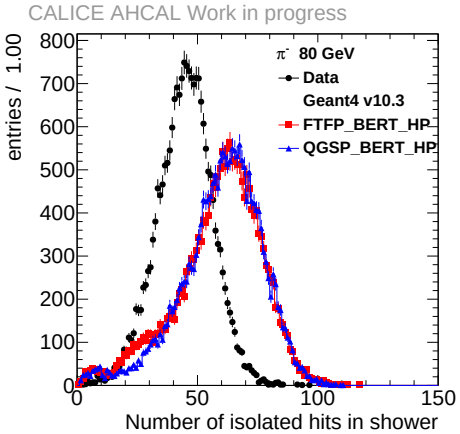


QGSP_BERT_HP: smooth distribution

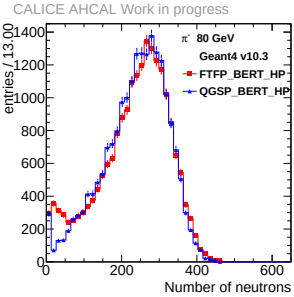
FTFP_BERT_HP: excess of low N_n and iso hits

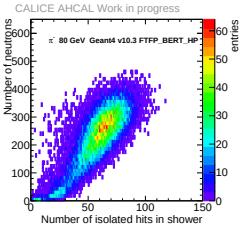
Number of isolated hits vs. number of neutrons at 80 GeV

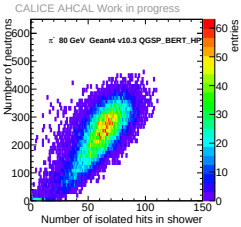
Data, FTFP_BERT_HP, QGSP_BERT_HP



MC truth: number of neutrons





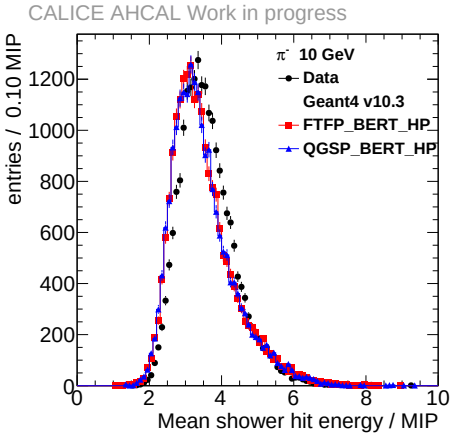


QGSP_BERT_HP: smooth distribution

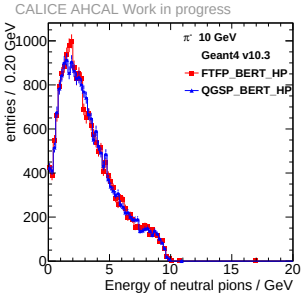
FTFP_BERT_HP: excess of low N_n and iso hits

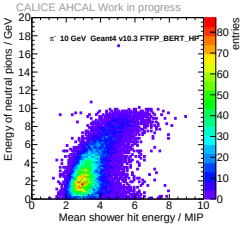
Mean shower hit energy vs. energy of π^0s at 10 GeV

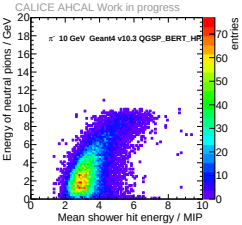
Data, FTFP_BERT_HP, QGSP_BERT_HP



MC truth: energy of π^0s



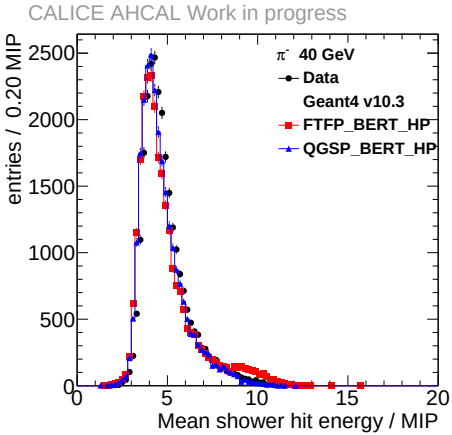




Good agreement between physics lists as expected
Slightly higher mean hit energy in data

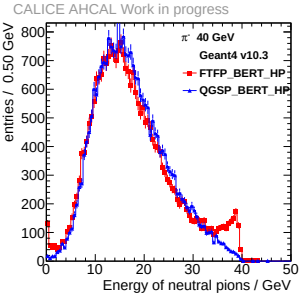
Mean shower hit energy vs. energy of π^0s at 40 GeV

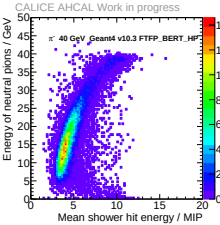
Data, FTFP_BERT_HP, QGSP_BERT_HP

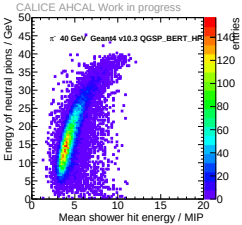


Good agreement between physics lists as expected
Slightly higher mean hit energy in data

MC truth: energy of π^0s

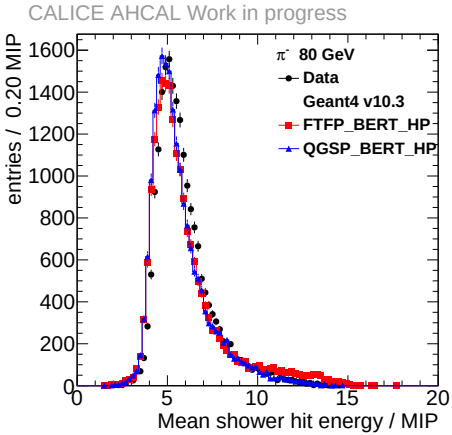






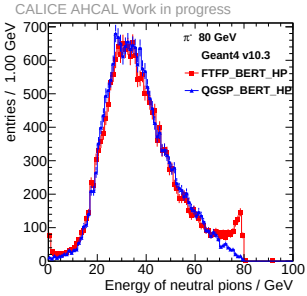
Mean shower hit energy vs. energy of π^0s at 80 GeV

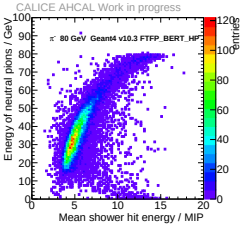
Data, FTFP_BERT_HP, QGSP_BERT_HP

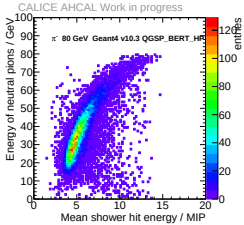


Good agreement between physics lists as expected
Slightly higher mean hit energy in data

MC truth: energy of π^0s







QGSP_BERT_HP: correlation maps

G4 10.3 QGSP_BERT_HP: π^-, 10 GeV in CALICE AHCAL

	N_{hits}	E_{reco}	N_{iso}	N_{trk}	$\langle e_{hit} \rangle$	R_{sh}	Z_{CoG}
$T_{neutron}$	0.15	-0.34	0.47	0.29	-0.41	0.52	0.31
$N_{neutron}$	0.31	-0.34	0.55	0.39	-0.53	0.56	0.32
E_{π^0}	-0.29	0.42	-0.46	-0.49	0.59	-0.62	-0.38
N_{π^0}	0.28	0.07	0.12	0.10	-0.16	0.06	0.05

G4 10.3 QGSP_BERT_HP: π^-, 80 GeV in CALICE AHCAL

	N_{hits}	E_{reco}	N_{iso}	N_{trk}	$\langle e_{hit} \rangle$	R_{sh}	Z_{CoG}
$T_{neutron}$	0.45	-0.14	0.48	0.45	-0.52	0.33	0.37
$N_{neutron}$	0.76	-0.07	0.78	0.72	-0.73	0.38	0.55
E_{π^0}	-0.37	0.43	-0.41	-0.44	0.71	-0.49	-0.37
N_{π^0}	0.68	-0.04	0.61	0.60	-0.66	0.33	0.49

Sronger correlation or **stronger anticorrelation**

G4 10.3 QGSP_BERT_HP: π^-, 40 GeV in CALICE AHCAL

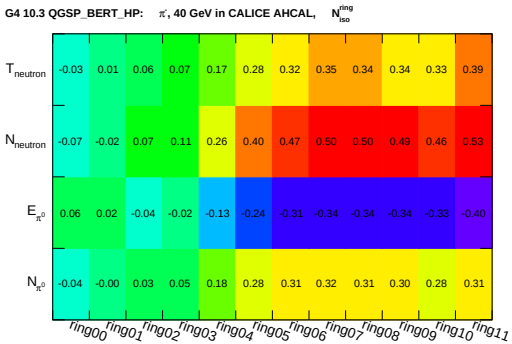
	N_{hits}	E_{reco}	N_{iso}	N_{trk}	$\langle e_{hit} \rangle$	R_{sh}	Z_{CoG}
$T_{neutron}$	0.44	-0.21	0.52	0.45	-0.54	0.44	0.37
$N_{neutron}$	0.70	-0.16	0.74	0.66	-0.73	0.52	0.51
E_{π^0}	-0.43	0.41	-0.49	-0.51	0.70	-0.56	-0.40
N_{π^0}	0.57	-0.04	0.47	0.46	-0.54	0.35	0.35

- N_{neutrons} strongly correlates with number of isolated hits and anticorrelates with mean hit energy.
- energy of neutral pions (dominate in em fraction) correlates with mean hit energy and anricorrelates with number of track hits and shower radius.
- different trends of correlations with energy

See correlations for FTFP_BERT_HP in backup.

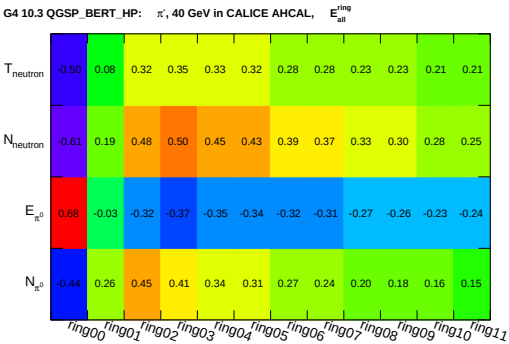
QGSP_BERT_HP: correlation maps of "ring" observables at 40 GeV

Number of isolated hits in a ring



Number of neutrons correlates with number of isolated hits in outmost rings.

Energy sum in a ring



Energy of neutral pions correlates with energy sum in the innermost ring.

(*ring*00 – innermost, *ring*11 – outmost)

Summary

Subject of the study:

- characteristics of hadronic showers induced by single pions in the CALICE AHCAL (π^- of 10–80 GeV);
- highly granular calorimetric observables: number of isolated hits, mean hit energy, shower radius, etc.;
- properties of secondaries within a shower (number of neutrons, energy of neutral pions, etc.) produced at generation level from simulations by Geant4 v10.3 with FTFP_BERT_HP and QGSP_BERT_HP physics lists.

N.B.: G4 v10.3 makes sense as proof-of-principle, while the community is interested in v10.6 and above.

Observations:

- **noticeable relationships between calorimetric observables and properties of secondaries** (many relationships are nonlinear);
- different trends of these correlations with energy;
- different intrinsic shower parameters generated by QGSP and FTFP models;
- differencies in calorimetric observables between MC and data.

In the next talk by Sergey:

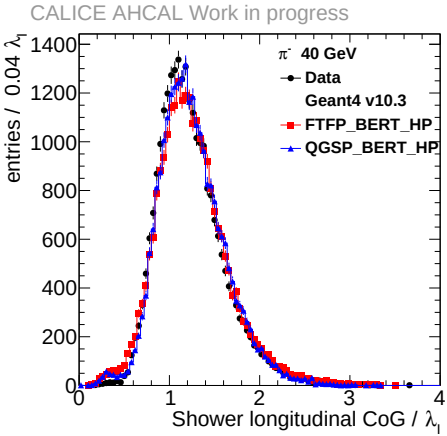
implementation of the neural network technique for prediction of properties of secondaries in pion-induced showers using calorimetric observables.

Backup:

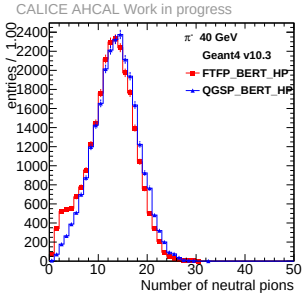
Additional slides

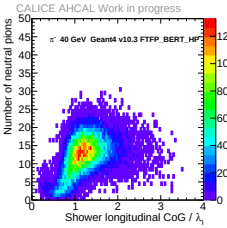
Longitudinal CoG vs. number of neutral pions at 40 GeV

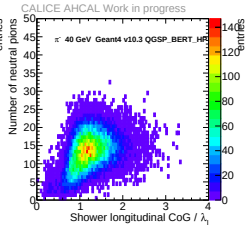
Data, FTFP_BERT_HP, QGSP_BERT_HP



MC truth: number of π^0s







Good agreement between data and simulations

QGSP_BERT_HP: smooth distribution

FTFP_BERT_HP: excess at low number of π^0s

FTFP_BERT_HP: correlation maps

G4 10.3 FTFP_BERT_HP: π^-, 10 GeV in CALICE AHCAL

	N_{hits}	E_{reco}	N_{iso}	N_{trk}	$\langle e_{hit} \rangle$	R_{sh}	Z_{CoG}
$T_{neutron}$	0.15	-0.35	0.46	0.30	-0.41	0.53	0.30
$N_{neutron}$	0.31	-0.33	0.55	0.40	-0.52	0.55	0.31
E_{π^0}	-0.29	0.41	-0.45	-0.49	0.58	-0.61	-0.38
N_{π^0}	0.25	0.06	0.12	0.07	-0.15	0.06	0.03

G4 10.3 FTFP_BERT_HP: π^-, 80 GeV in CALICE AHCAL

	N_{hits}	E_{reco}	N_{iso}	N_{trk}	$\langle e_{hit} \rangle$	R_{sh}	Z_{CoG}
$T_{neutron}$	0.57	-0.22	0.61	0.59	-0.63	0.44	0.44
$N_{neutron}$	0.77	-0.21	0.80	0.76	-0.79	0.51	0.58
E_{π^0}	-0.41	0.44	-0.46	-0.48	0.68	-0.52	-0.39
N_{π^0}	0.70	-0.20	0.67	0.66	-0.75	0.47	0.52

Sronger correlation or **stronger anticorrelation**

G4 10.3 FTFP_BERT_HP: π^-, 40 GeV in CALICE AHCAL

	N_{hits}	E_{reco}	N_{iso}	N_{trk}	$\langle e_{hit} \rangle$	R_{sh}	Z_{CoG}
$T_{neutron}$	0.56	-0.28	0.64	0.59	-0.65	0.59	0.42
$N_{neutron}$	0.72	-0.28	0.77	0.71	-0.79	0.67	0.54
E_{π^0}	-0.47	0.41	-0.52	-0.54	0.68	-0.60	-0.42
N_{π^0}	0.63	-0.17	0.56	0.55	-0.67	0.51	0.43

- N_{neutrons} strongly correlates with number of isolated hits and anticorrelates with mean hit energy.
- energy of neutral pions (dominate in em fraction) correlates with mean hit energy and anricorrelates with number of track hits and shower radius.
- different trends of correlations with energy