MINERvA Technical Note No. 045
DocDB document 9474

Pion calorimetry from MINERvA testbeam data

Rik Gran, University of Minnesota Duluth

May 5, 2014

The testbeam pion calorimetry data between 0.35 and 1.95 GeV from the 2010 20ECAL20HCAL run is well described by the simulation, leading to the conclusion that we have constrained the pion shower energy response uncertainty to within 4%. Despite this agreement, there is indication of a Geant4 modeling error equivalent to having a too-long mean free path in the detector (e.g. too small a reaction cross section) for pions with energy greater than 0.9 GeV. With systematic uncertainties from 2.6% to 3.6% for data/mc comparisons and 3.3% to 4.2% absolute, depending on energy and species, but with evidence of a discrepancy of at least 4%, a larger 4% uncertainty is a balanced way to propagate the energy scale to full MINERvA analysis. Other comparisons of the calorimetric response between this dataset and simulation are not sensitive to aspects of the fates of pions that interact. Some parameters of the simulation here are optimized differently than the Resurrection-era simulation for the full MINERvA detector, which require additional uncertainty or bias correction in the calorimetric response.

Contents

1 Introduction

One of the primary deliverables of the MINERvA testbeam experiment is to demonstrate that the Monte Carlo, the detector model as well as the Geant4 physics models, describes the calorimetric response to hadrons. The energy response and resolution are vital for the accurate reconstruction of the hadronic energy in the neutrino events, as well as the total neutrino energy. We have constrained the accuracy of the simulation and have a bound on the single-particle hadronic energy bias. In this case, there is mild a discrepancy between data and MC, these data might be used to obtain a correction to reduce the energy scale bias.

2 Calibration, event selection, and simulation

After selection, the data set from the 2010 run with the 20 ECAL + 20 HCAL (20E20H) configuration has 8,229 positive pions and 10,239 negative pions. This is compared to a post-selection MC sample that is 37 or 36 times as large. The size of the data set and the nature of the analysis gives 0.5% statistical uncertainty on the measurement, so systematic uncertainties dominate the result.

2.1 calibration

The detector calibrations are described in the technical note TN018 in docdb:8686. An important additional calibration is the best tuned Birks' parameter TN037 docdb:9131 and by default there is no non-linearity correction to the data nor non-linearity applied to the MC. The beamline calibrations are discussed in TN017 docdb:8547. In the end, the beamline pion energy scale is accurate to 1% at most energies, rising to 2% accuracy at 1.8 GeV. The detector energy scale and mass model combine to give an uncertainty of 1.5%. There are additional uncertainties in the analysis coming from the event selection and backgrounds.

Most of the pions analyzed here undergo an interaction. In some cases they were tracked before the interaction occurred, and in a few cases the daughter particles are tracked afterward. This analysis does not directly use any tracking quantities for selection or analysis. However, if some part of the event was tracked, then the 3D information is used to apply attenuation corrections to the clusters in both data and MC, and that corrected energy is used in the analysis.

The plot-dump files associated with this document include discussion of low level hit occupancy, containment, and position in the detector, all of which seem stable, well calibrated, and well modeled.

2.2 event selection

The event selection criteria are primarily designed to ensure that an accurately measured pion and no other particles enter the detector. To a large extent, the careful selection described below is successful, however systematics related to the event selection remain a major systematic, and will be discussed later.

- Particle passes beamline fit quality checks. We have two versions; the loose beamline cuts are the default, and tight beamline cuts are available as a systematic alternate.
- Event passes beamline detector event match, which also includes a rejection if there appears to be a second particle entering the front. This is run on both data and MC.
- To reduce detrimental effects of pileup, events with another reconstructed time-slice 400 ns before or 200 ns after are rejected.
- The beamline reconstructed momentum and TOF (and mass calculated from those) is used to select pions. Because the pion sample is near the physical speed of light boundary for the TOF, the selection is not as simple as the 20% window around the actual particle mass, as used for kaons, protons, and deuterons.
- The above are incorporated into pass-all and PDG code variables in the tuple for convenience, and are used to generate a collection of actual data particles to seed the data-driven MC simulation. At analysis time they can be rebuilt in order to allow looser or tighter cuts.
- Pion available energy (kinetic + mass) is between 350 and 1950 MeV, higher energy pions are thrown out. This value is taken from the beamline momentum fit, but corrected by about 7.5 MeV using an energy dependent, Bethe-Bloch estimate for average pion energy loss in the air, scintillator, and wire chamber material between WC3 and the first plane of the detector for a pion of that momentum.
- The data is binned in 100 MeV bins starting with 350-450 MeV and going up, with a trio of bins 200 MeV wide, and a final bin 300 MeV wide from 1650 to 1950 MeV.
- There must be hits > 0.5 MeV in three of the first four planes of the detector and less than 1 MeV combined in the last four planes.
- There is less than 8ns of uncertainty on setting the start time “Tzero” corresponding to the hit time for the main activity for the event.

2.3 simulation

The standard simulation is “data driven” and uses measured beamline particles that pass the beamline loose selection, but not the additional detector cleaning cuts. The measured beamline particle has its momentum and direction randomized within its Gaussian resolution and put into the simulation 5cm upstream from WC3, the point at which the beamline fit is returning its most robust information and is close to the halfway path length through the beamline. These resolutions are estimated by the beamline on a particle-by-particle basis from the beamline fit, so they change with the quality of the fit including the natural trend with reconstructed momentum as well as any oddities of the individual event. By using different random number seeds for this procedure, the initial data selection can smoothly produce many multiples of the initial data set, typically we specify 10x or 20x samples of MC.

Because the particles are put into the simulation just before WC3, the Geant4 simulation propagates them through material representing air, wire chambers, TOF, and muon wall scintillator. Further multiple scattering, 7.5 MeV energy loss, and interactions occur in the simulation, and not all of these particles will pass the matching and clean event cuts. The efficiency for data and background is shown in Fig 1 for positive pions.

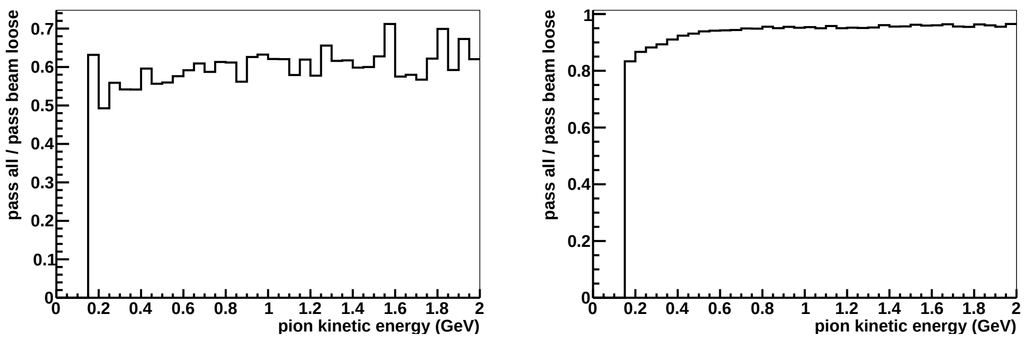

Figure 1: Efficiency (acceptance?) for a π^+ trigger that passes the loose beamline cuts to also pass the clean event cuts. Left is data, and is around 60% because of beamline pileup cuts or because the event was actually some background and not a valid trigger. Right is MC, which is at 94% due to particles that scatter out of the beamline while the simulation passes them from WC3 to the front of the detector. Except for the overall offset, the character of the efficiency is the same, and the resulting samples have well matched energy spectra. The corresponding plots for π^- have similar shapes.

The simulation explicitly does not include beamline pileup of any kind and has an overall efficiency of 94% compared to the data 60%, so for this reason also a different fraction of events pass the match and clean event cuts as well as the (-400,200)ns pileup cuts. A 20x sample during generation turns into more like an 37x sample during analysis for pions, and within statistical fluctuations is uniform across all energies. Systematics using different selection or weighting is evaluated with this same large MC sample, but there are special-run MC systematics studies discussed in this document which are evaluated with samples half this size. The corresponding plots for π^- look the same, except the data efficiency is 64% instead of 59%, likely due to running at lower beam intensity.

The reconstructed quantities that are used for analysis are the beamline actual quantities from the particle that seeded the simulation, not the smeared MC truth quantities. Making distributions of only these beamline quantities for a 37x MC sample would show identical repeated events and incorrect statistical uncertainties. Making distributions of detector reconstructed quantities based on the smeared particle put far upstream to start the detector simulation would NOT be afflicted, and will in fact demonstrate resolutions and fluctuations that match the real data. Combining relatively coarse binning for the beamline quantities with the smeared detector quantities gives the desired statistically smooth properties for the final MC sample.

The passage of particles through the detector is simulated using Geant4 version 9.4.p02. By default, the test beam simulation uses finer length scale for the MCHit aggregator. This code takes multiple MC hits and combines them before simulating Birks' suppression and photostatistics. We can test special samples with the Titan/Resurrection default and even finer step sizes. The simulation of the testbeam detector response follows the same strategy as for the MINERvA detector: individual strip response is decalibrated, simulated with photostatistics and resolution smearing, recalibrated, then made available to reconstruction algorithms.

Several low-level tests confirm that, after tuning, the different resolutions are well simulated, and are documented in TN018. Proton beamline KE vs. range shows the beamline resolution is accurate, proton visible energy per plane shows the resolution in dE due to gains, attenuations, and photostatistics is good. The observed residual location and angle at the front of the detector is good. There is a spectrum distortion from this selection which is most pronounced at the threshold momentum for each particle species, in Fig. 2, but the spectrum distortions after all selections are applied is very mild.

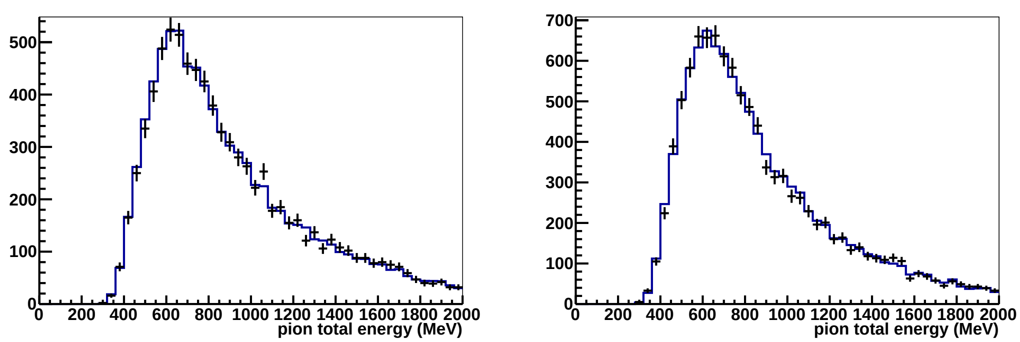

Figure 2: Energy spectrum of pi+ (left) and pi- (right) after all selections are applied to the data and MC. Events below 350 MeV and above 1950 MeV are not used in the analysis. The MC shows statistical fluctuations around the data sample that seeds the simulation.

3 Results

For every particle in the sample, the visible energy is reconstructed, followed by the calorimetrically corrected energy, examples for positive pions at bin-center energies of 800 and 1040 MeV are shown in Fig. 3, the other energy bins are shown in the supplemental big pile of plots. The MC does a good job of describing the data, both in the shape and the average. The visible energy is the simple sum of all hit energy in each plane. Because sometimes a 3D track is formed, some of the hits are in clusters which have attenuation corrections are applied, while most clusters will not, defeating the attenuation corrections raises the response by 0.5% for both data and MC.

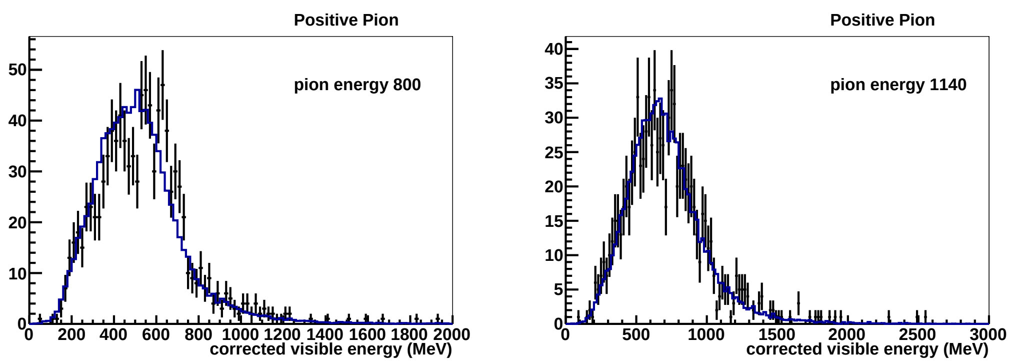

Figure 3: Visible energy distribution with calorimetric corrections applied for the bins centered at 800 MeV and 1140 MeV. This is the visible energy, including cross talk clusters, and including attenuation corrections for clusters on 3D track segments, all corrected for passive material, then scaled down 4.2% to effectively subtract off cross talk.

The summing of activity in the detector explicitly includes any and all cross-talk clusters, which is for three reasons. Once a shower starts, it becomes difficult to separate cross-talk hits from shower hits. We do not run the regular MINERvA cross-talk recognition algorithm, though also we do not need to run a kind of pattern recognition that might be sensitive to additional clusters. Finally, the PMTs included several with higher than average cross-talk according to the bench test, which were reserved as main detector spares and for use in the test beam. We have roughly tuned the level of cross talk, and estimate the fraction of energy attributed to it is 4.2±0.5% of the total energy. We apply this average correction uniformly to each pion event, rather than try to determine which clusters are cross talk or not.

For energy observed in the ECAL, we apply a calorimetric correction of 2.0773 to account for the passive material in the scintillator plane and the Pb. For the steel HCAL, the correction is 10.7271; this analysis does not have a tracker-only configuration which would have had a correction of 1.2551. The assumption here is that whatever energy deposit was observed in 1.8 cm of scintillator plane, that same dE/dx was being deposited in the passive material immediately upstream or downstream.

One other correction is applied, this one only to data. On average there is a little energy in the two adjacent 19ns buckets prior to the main event due to detector pileup, almost all of it in the ECAL portion of the detector. When this is extrapolated through the -30 to +170 ns window where we analyze the energy deposits, this comes to 2 MeV (with ECAL passive material correction, its 4 MeV). This amount is subtracted from each event. Because the correction is already less than 0.5% we will not propagate an uncertainty from it.

This amount of energy is the reconstructed response of the detector to the particle, on average about 60%, and varies from particle to particle (even with the same input energy). The development of a series of hadronic interactions produces some amount of neutral energy, neutrons, photons, and some neutrinos, which leave the detector unmeasured. Another important source of missing energy is used to unbind nucleons from the nucleus during hadron interactions, an effect most extreme for spallation type reactions. The electromagnetic components of the shower have a different response than charged hadrons. Leaving the detector without depositing energy is easier in the HCAL than the ECAL, though its easier at the very front of the detector. Fluctuations in dE/dx within a scintillator plane, especially in the HCAL, can be magnified in the total response. Finally, even with the calorimetric corrections, showering early preferentially gives lower response. These points are quantified and discussed in more detail later.

The detector response to pions is simply the corrected visible energy divided by the true total energy. This can be formed event by event, and the distribution is plotted for two selected energy bins in Fig. 4. Because it is formed event by event, and because the response ratio changes slowly with energy, this distribution is also relatively insensitive to effects in modeling the incoming pion energy spectrum.

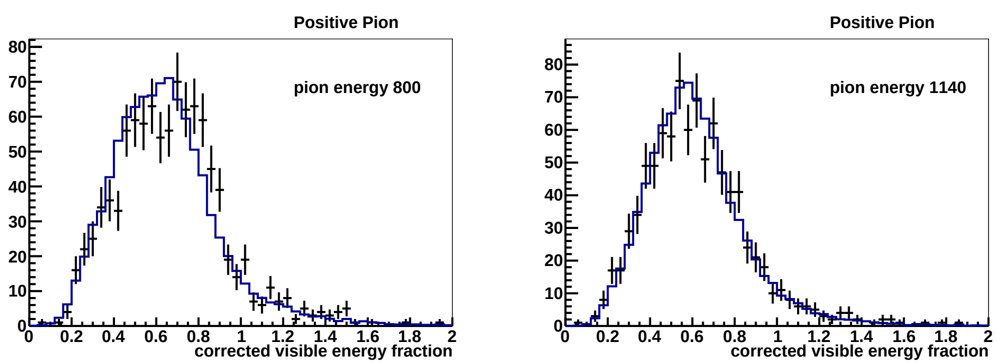

Figure 4: Ratio of visible energy to pion total energy, event by event, for selected ranges of energy.

Because the shape (rise, peak, tail) is reasonably well reproduced, the MC agreement response can be summarized by considering the distribution simple mean energy with uncertainty on the data of RMS/sqrt(N), and plotted for each bin, as shown in Fig. 5. The top plots are obtained from the mean of the distributions like the ones shown in Fig. 3 while the right plots are obtained from the mean of the distributions like Fig. 4. The uncertainty band represents the sum of all uncertainties that apply to comparisons of the MC to data, but not those that apply equally to the data and MC and are uncertainties on the absolute vertical energy response scale.

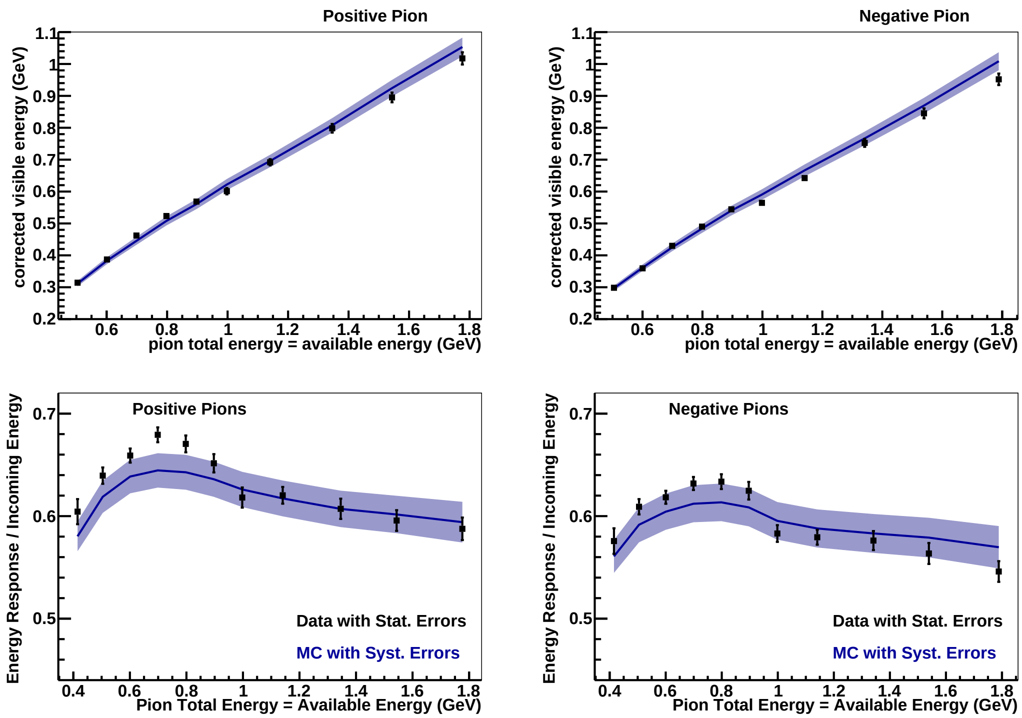

Figure 5: Calorimetric response for positive (left) and negative (right) pions. Top plots demonstrates the full response while the bottom plots show the fractional response. The errors on the data are statistical only, while the error band on the MC represents the 2.6% to 2.9% systematic uncertainty relevant to data vs. MC comparisons, growing to 3.4% or 3.6% for π^+ and π^- respectively. The absolute uncertainty is between 3.3% and 4.2% varying with energy and species, while the MC statistical uncertainty of order 0.5% for most data points.

Because these plots are taken with the same detector, and approximately the same beam, certain trends visible by eye might be more significant than the error band suggests, others require a careful analysis of the systematics. The error bands are dominated by effects that are the same between the π^+ π^- samples.

There is a trend that the response of the MC is lower for low pion energy compared to high pion energy, with the data showing a break in behavior between the data points at 900 and 1000 MeV that is not modeled in the MC. This correlates with a feature in how the energy is split between the ECAL and HCAL at different energies. At low energy, nearly half the events have zero or negligible (less than 2%)

fraction of their calorimetric energy coming from the HCAL, shown in Fig. 6. The MC describes this well up to 0.8 GeV, after which the data shows a significantly higher fraction of events in this special category, with eight of ten high energy data points statistically beyond 1σ agreement.

The underlying physics that probably best drives both the no-HCAL fraction and a reduction in the response is a decrease in the mean free path in the physical detector. A material model error would do this, but would probably not turn on at 0.9 GeV, while the reaction cross section would cause this. There is a not-quite-smooth feature in the π^- MC at this same energy (though not for π^+); a candidate explanation is that the effect the MC is already capturing should be exaggerated.

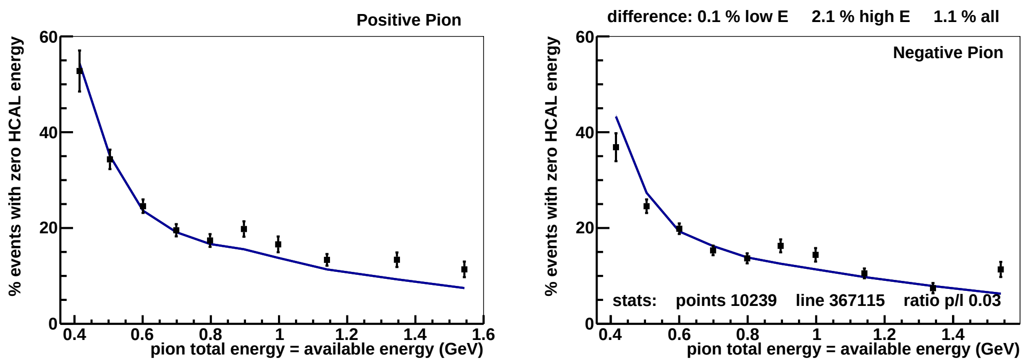

Figure 6: Percent of events with less than 2% of their energy in the HCAL for positive (left) and negative (right) pions. The change in behavior at 0.9 GeV matches the change in behavior in the response plots.

Because the shape of the corrected visible energy fraction is well reproduced, the pion energy resolution is well described by the MC. Figure 7 uses the simple RMS to quantify this in a way that agreement and trends are easy to visualize. It is expressed as percent resolution (RMS/Mean), so typical values are 0.25 RMS / 0.60 mean = 42%. The errors here are from the ROOT histogram function GetRMSError(), which seem to underestimate the true error, but clearly the trend is accurately modeled. At higher energies it would be typical to express this as %/sqrt(E) which reflects the statistical fluctuations of a multi-stage hadronic shower, but at these energies we are instead looking at shower fluctuations dominated by which fates occurs for just one or two interactions. A reasonable way to summarize this in a manuscript describing a neutrino analysis is “below 2 GeV, the calorimetric resolution for singple pions is typically 40% and the MC does a good job of describing the non-Gaussian shape”.

Finally, because the systematic uncertainties on the response are highly correlated, forming the ratio π^+/π^- is interesting. The two ratios are separately consistent with flat, and the data ratio is 1.3% $\pm$ .5% more than the MC ratio which is at 1.048, with π^+ giving intrinsically higher response. Many systematics (described in the next section) are the same for the two samples. There is no evidence for

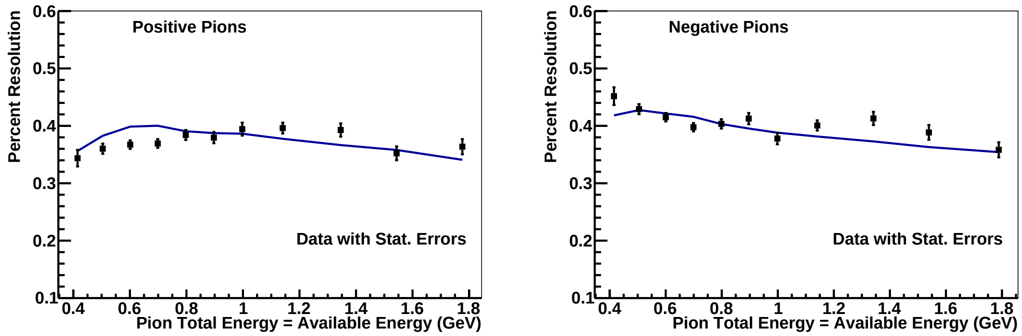

Figure 7: Resolution of the calorimetric response for positive (left) and negative (right) pions.

a change in behavior at 0.9 GeV in the ratio.

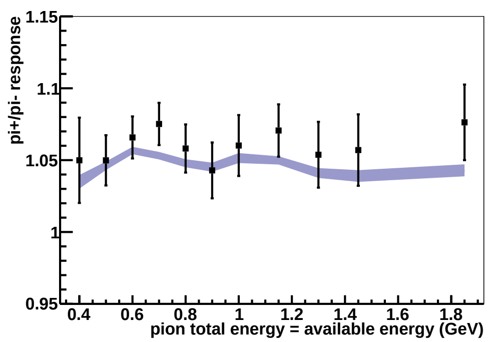

Figure 8: Ratio π^+/π^- of the response for data and MC. The MC predicts 1.048, the data yields 1.062. It looks like the statistical error bars are too big, and I need to look again what miscalculation might be happening.

As a prelude to a closer look at the systematic error band and conclusions that follow, these results can be summarized by offering these two scenarios. First, the MC models the data well, the error bars touch at most energies. An alternative interpretation is that there is a 1.3% energy scale shift between the the π^+ and π^- response that the MC does not model (on top of the 5% shift it does model), combined with an effect where the MC has too high a mean free path in the detector starting around 0.9 GeV, and an overall energy response uncertainty approximately flat with energy.

4 Systematic uncertainties

This is a systematics dominated measurement. Data statistical error on single points are between 1.2% and 1.8%, but much of the power of the analysis comes from considering the whole data set, which implies statistical uncertainties of around 0.5%. Enormous effort has gone into calibrating, constraining, and cross checking systematic effects. In the final stage of the analysis, many candidate systematics are scripted into plots for two documents, one each for positive and negative pions, and included as supplementary material in docdb:9474. The results are summarized in a table below with a discussion of the most important or interesting ones. The strategy is to apply a change to the data or MC (or both) corresponding to the uncertainty in some feature of the selection, simulation, or reconstruction, remake the energy response distributions, and remake the summary plot of the means of those distributions. Some effects are data only, like dependence with beam intensity or differences between early or late in the run. Others are MC only, such as reweighting or rerunning to produce a change in material or stepsize in the simulation. If its done in both data and MC such as a selection, then the MC predicts a shift and the discrepancy is whether the data shifts differently than predicted.

4.1 beamline momentum uncertainty

The bias uncertainty in the reconstructed beamline momentum changes as a function of momentum (Energy x 1% / GeV) reflecting the increasing importance of the alignment uncertainties in the wire chambers at higher energy. Added to this are uncertainties three 0.5% uncertainties in the magnet alignment, the longitude size of the magnet, and other magnet uncertainties. Taken together, the uncertainty is 1% at 0.5 GeV and 2.0% at 1.8 GeV. The resolution, the unbiased uncertainty on individual event momentum assignments is well modeled.

In the calorimetry data, we see this interesting effect at 0.9 GeV, but conclude that sources of bias in the beamline could not be large enough or of the right type to cause that effect. In particular, the slope in Fig. 6 is so shallow at 0.9 Gev that shifting the blue line to the right by a few percent to mimic a beamline bias would produce only mildly better agreement.

4.2 existence of an event slice between 200ns and 800ns and afterpulsing

By default we eliminate events when we find another slice within (-400,200) ns of the slice with the pion. This certainly reduces the chance that activity attributed to the pion is not split between two slices nor messed up by another event from the beam in an nearby, early 19ns bucket. But there is other, more subtle physics at work in both the data and the MC, and to evaluate the robustness of our result, we rerun while rejecting all events that produce a slice within 800 ns.

The change in response is shown in Fig. 9. The format of this figure is typical of all such studies in this note and in the plot dump documents. The blue is the baseline result and the points show the result with the modified cut, MC is on the left and data is on the right.

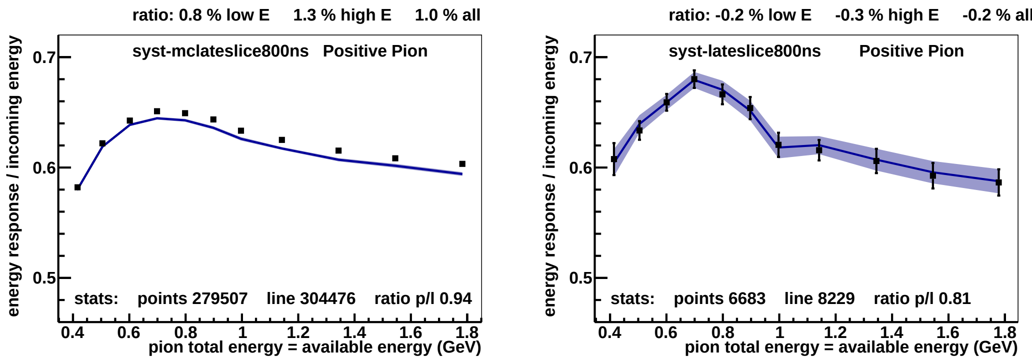

Figure 9: Effect of modifying the pileup cut to reject events with another slice within 800 ns after the selected event. Left is the effect in MC, right is the effect in data. In both cases, blue is the default with statistical uncertainties, while the points show the effect of the additional cut. The values printed at the top show the percent shift for the high energy five data points, and for the next five data points, and the total, but not including the 400 MeV data point. The effect illustrated here is discussed.

Starting with the MC (left plot), there is a shift to higher response. The MC contains events which cause energy to show up early and late, which sometimes is caught by the chronobuncher and declared its own event slice. When Geant4 liberates a soft neutron, it may bounce around in the detector for a while before kicking a proton with enough energy to produce a few photoelectrons which can be formed into a one or few hit slice. Secondly, a Michel electron from inelastic pion production has a good probability to appear within 800 ns. Events that have these characteristics, especially those that produce numerous neutrons, preferentially have lower energy response, so eliminating them yields a higher overall response.

In the data, we presume this also happens, but the observed response in the data with the new selection is consistent with statistical fluctuations. However, the MC does not model afterpulsing, nor do we have a good check that the chronobuncher responds identically to small data and MC events. And of course, the MC's idea of neutron and inelastic pion production might not be perfectly accurate. And there is additional pileup events in this larger window, though we suppose the data events cut in this way are an unbiased subset of all interactions. In summary, that the data response is 1.3% lower overall says that the MC is not modeling one or more of the first four properties. Not shown, this discrepancy grows to 1.65% at high energy.

A team of two UM Duluth undergraduates have scanned a selection of events in data and MC, looking for smoke or fire, and has not found substantial evidence of misbehavior or serious afterpulsing.

[In fact, this might not properly be a systematic error, in the sense that we are testing Geant4]

4.3 event selection

This is a switch to the old sliding mass cut used in the RogerBlough era. This is applied on top of the current cuts and trims another 20% of the events in data, eliminating regions with higher electron, positron, and kaon contamination. Electrons and kaons passing the rest of the event selection will have 10% to 20% higher response than pions; electrons intrinsically give higher response for the same energy while mis-id kaons have about three times more available calorimetric energy for a given measured particle momentum.

Indeed this cut lowers the average response by 0.7% for π^+ and 1.5% for π^-. We take this as a systematic on both the data/MC and absolute response. If we could demonstrate definitively that the electron, kaon mechanism was the answer, then arguably the tighter cut represents better accuracy and we should adopt it. However, at this time we can not rule out alternate instrumental contributions to this shift, and we take the larger, conservative uncertainty. The shift in response is uniform in energy.

A second approach to this is use the tignt beamline selection instead of the default loose selection. This tightens several tolerances for the reconstruction of the track in the beamline WC. This change has about half the effect as the one above, and is picking up on the same kind of challenging triggers and backgrounds. We simply take the larger uncertainty above.

4.4 material model

Changing the Pb density by -2σ has 0.5% change on the calorimetric response, and caused a slight but statistically significant 0.3% decrease of events with negligible HCAL energy. The latter is in the expected direction because it is equivalent to increasing the mean free path (in planes) before the pion interacts.

Lowering the polystyrene density by 2σ lowers the calorimetric response by 2.6% and lowers the number of events with negligible HCAL energy by 0.5%.

There are two effects related to the material model. One is the active/total material fraction, which the fundamental uncertainty in the absolute energy scale, and is described in the next paragraph. Most of the 2.6% of the polystyrene result is due to this effect, but it is calibrated out. There is a milder effect in a comparison of MC to data if the material budget allows pions to propagate deeper into the detector. Imagine the simpler scenario, where the passive Pb material is lowered significantly in the MC. Then less energy is lost in the Pb, and more will appear as energy one plane deeper than before, and overall the shifted MC response could be higher. This statement is independent of uncertainties in absolute energy scale calibration. According to this test it is about 0.5%.

The per-plane response, through the MEU tuning, is done in such a way that the data/MC relative response is not affected by the (active scintillator)/(total material) but the absolute response for both is affected. After passive material corrections are applied, the uncertainty in the active/total fraction cancels out, and the uncertainty in the total material remains. For the 20E20H detector, this is the quadrature sum of the 1.4% from the scintillator and 1.2% from the Pb or Fe, and is one of the largest contributions to the absolute uncertainty.

The Pb and Fe absorber is very simlar to the Full MINERvA, but we model our own material accurately. The model for the material in the scintillator plane is intended to be the same, but the test beam geometry has been updated relative to the Titan/Resurrection, especially it has about half the content of heavier atoms Si, Al, Ti, and Cl.

4.5 cross talk correction

As described earlier, the MEU scale is built using a muon sample which excludes crosstalk hits. The reconstructed energy in each plane along a muon track is measured *without* using cross talk hits, and so any energy scale should also not include them. However, when confronted by a hadronic or electromagnetic shower, it is not always possible to separate out cross talk from regular energy deposits on an individual basis. We correct for crosstalk by reducing the response for both data and MC by 4.2$\pm$0.5%, so this latter number is included in the uncertainty. This covers both the imperfections in the cross talk tuning and that the size of the correction is averaged over all planes, not weighted for energy in a particular plane. It applies equally to both to data/MC and to the absolute scale.

4.6 Birks' parameter and non-linearity

Rerunning the analysis with the MC Birks' parameter shifted down $\pm$15% has a $\mp$1.3% effect on the response. Of that, 0.3% is already accounted for in the way the MEU tuning is done, so only the remaining 1.0% is taken as an uncertainty.

The standard non-linearity uncertainty (half of Howard's line) causes a shift of 0.7% in the response at low energy, increasing to 1.0% at high energy. We found murky evidence for a non-linearity effect at this level, but with our data, we have not been able to rule it out either.

The Resurrection MC uses a Birks' parameter that is at +2σ relative to this value, and by itself it would produce a response that is 2% lower for these showering pions. The testbeam MC also uses a finer step size for the MCHit Aggregation, which allows the Birks' parameter to apply with its full magnitude, switching to the Resurrection MCHit aggregator raises the response 0.5% for the testbeam simulation.

4.7 Geant4 step size

The testbeam analysis uses a set of framework defaults which allow Geant4 to choose the optimal step size, subject only to a maximum step of 0.5 mm for electrons, positrons, and gammas. As of framework release v10r7p4 and for the Eroica processing this is expected to be the default for all of MINERvA. In addition, for test beam processing we turn off the MC hit aggregator.

To evaluate whether the step size of the simulation has an effect, we have rerun the simulation with a maximum allowed step of 0.05 mm applied essentially to all charged particles including charged nuclear fragments and nuclei. As of this writing, we have completed this test for a mono-energetic pion sample with momentum of 600 MeV/c. The smaller step size redues the MC response by 1%. A similar test with 500 MeV/c electrons shows no effect, suggesting that the step size applied to neutrons, protons, and pions dominates the effect.

We do not use the MC hit aggregator for testbeam simulation. It combines hits in the same scintillator volume together, and passes the set (both dE and dx) to the Birks calculation and the simulation of the photo-electron yield. At the end of a proton, taking these hits together causes some dE to be spread out over too large a dx, reducing the amount of suppression from Birks' effect, thereby increasing the overall response by 0.5%, the testbeam default is lower. We should consider whether we can change the default code to perform the Birks' suppression or save the dE/dx information and still get the right PE yield while processing and saving fewer MCHits.

The MCHit aggregation effect is not a systematic because the testbeam uses the higher fidelity simulation. The step size is an interesting question. Clearly it is also reasonable to take in quadrature with the rest of the uncertainties. For a correction, we should probbly consider a correlation with Birks' parameter and how its one-sided, however its just one of many 1% uncertainties. In a way, it is a known component of the discrepancy between reality and Geant4, so it is a component of what we are trying to measure. In this discussion we comment on it separately, rather than wrap it into the MC uncertainty.

One final comment. It is likely that the Resurrection processing uses a Geant4 steering that contains a bug in the options file that cuts neutrons with energy less than 10 MeV. Under this circumstance, we're not sure what happens to the energy associated with these neutrons, except that certainly they are not available to deposit energy in the detector as they bounce off nuclei in time and for several microseconds after their production. Compared to the testbeam default processing without this bug, the effect decreases the resposne between 1% and 2%.

4.8 energy scale

Because of the way the MEU tuning is done, the uncertainty on the energy scale is quite small, 0.6%. In some sense, the measurement presented in this note is actually a pion/muon measurement with the absolute uncertainty separated. The largest contribution to the energy scale uncertainty in fact comes from the observation of a likely discrepancy in the muon response between the earlier 20E20H data and the later 20T20E calibrations. In that case we took half the discrepancy, stating that the 1σ uncertainty for each sample sits at the halfway mark between them, and then kept it symmetric for simplicity. The MEU factor extracted for the 20T20E data is 0.9% higher, the gains are higher by about the same amount, but the extracted light yield is not. A piece of corroborating information, the dE/dx for stopping protons as they travel through the first planes of the detector also suggest a 1% energy scale shift; supposing its a calibration effect not a physics effect, the 20E20H data has a lower response relative to the 20T20E data.

This observation applies to the pion calorimetry situation in an indirect way, because we see mild evidence for an energy scale shift causing the π^+ to have 1.3% higher response relative to the earlier π^- data. We don't know the cause of the 20T20E vs. 20E20H energy scale error, but one hypothesis is that it has to do with delta ray production and brems in the different absorbers, another has to do with retuning the HV and changing the Gains methodology between the two configurations. Neither of these would cause a shift within the 20E20H calorimetry results. Another possibility is some time dependent aging phenomena which is (for some unknown reason) not captured with our nightly calibrations or is unexpectedly something in the beamline; this kind of hypothesis could cause a trend across the entire 2010 testbeam dataset.

To check, we divided π^+ and π^- into first half and second half, and repeated the calorimetry measurement on each of the half-statistics samples. There are fluctuations at the edge of the statistical uncertainties, but they do not form an identifiable time dependent trend at all energies that would explain a 1.3% offset between the samples. With our uncertainties, we only rule out a sample dependent instrumental effect at something like 2σ, leaving a good but not certain chance that Geant4 is failing to capture another 1% difference between pion polarities, above and beyond the 5% difference it does capture.

4.9 sanity checks

Cluster timing: The timing distribution of clusters is not perfectly reproduced. The analysis integrates over all clusters in the slice from 30 ns before the event begins. This includes energy on pixels where the discriminator did not fire, much of which is cross talk but can also include soft energy deposits from real physics processes. As a check, the analysis can be rerun including only the main timing peak from -30 ns to +50 ns, but then turn off the data pileup subtraction, but don't turn off the cross-talk subtraction. In this case, the MC response decreases 1% more than the data. This test is picking up imperfections in the timing distribution which are not necessarily imperfections in the energy response, plus imperfections in the amount of cross talk, plus imperfections in the beam pileup background, and possible actual discrepancies between the Geant4 model and reality. This discrepancy seems adequately covered already.

Low energy clusters: Another check is to eliminate all low energy clusters dE < 0.5 MeV and redo calorimetry. This removes most cross talk clusters, but will also remove an unknown number of real physics clusters in data and MC. The response in the data goes down 4.2% but in the MC it goes down 5.5%. The overall magnitude of this is consistent with the $4.2\pm0.5\%$ we are eliminating and attributing to cross talk, but the difference is exaggerated in the MC. Because hits outside the timing peak are preferentially late in the timing distribution and preferentially small, this is very likely measuring the same effect as the timing distribution.

Divide sample in halves: The data sample was divided in half according to several parameters: early/late in run, high/low temperature, vertical high/low, and beam intensity high/low. With these variations, the total data, low energy sample, and the lower statistics high energy sample were inspected. The results were consistent with the statistical uncertainy for half-samples. Two excursions of 3% were observed, both in the high energy sample, also consistent with statistical uncertainties. This validation is not reveal anything that should be treated as an uncertainty.

4.10 use with Titan/Resurrection and Eroica processing

When applying these energy scale uncertainties to analyses of MINERvA neutrino data, we should realize that several systematics are already included in the response energy scale uncertainty, and should not be double counted. The error here includes the Birks' effect (except for Resurrection), non-linearity, and material model. This is not to say they should not be checked, if any of these affect something other than calorimetric response (such as tracking, PID, or dE/dx based energy reconstruction), then these testbeam results do not inform that corner of your neutrino analysis.

In certain kinds of neutrino analyses, uncertainties in the hadronic energy scale are dominated by uncertainties in the amount of intranuclear rescattering final state interactions (FSI). This might not be true if this FSI effect is what is explicitly being measured, or if the analysis explicitly states the observable is post-FSI. An interesting, clean example is the situation of coherent pion production. In that case, the signal process pions' energy scale uncertainty is exactly what is constrained and validated by the analysis presented here.

However, these testbeam simulation being compared to testbeam data is different than the one used in the Titan/Resurrection processing, and the discrepancy between the two simulations should be taken in quadrature, if they are not corrected. There are mild differences in the v10r6pX geometry that are within the uncertainty on the mass model, most important are probably the effects of the Cl,Ti,Al,Si elements with their higher Z, and most significant differences in the neutron-cut bug and Birks' parameter. These effects can be isolated as follows:

- To go from testbeam to Resurrection Birks' parameter decreases response 2%
- The effect of the MCHit aggregator step size that increases the response by 0.5 %
- The neutron cut in Resurrection decreases response (energy dependent) by 1 to 2%
- The neutron cut in Resurrection also eliminates out-of-slice late activity from slow neutrons.

Conclusion: out of the box Resurrection single particle response is 2% to 4% lower than testbeam MC for the same 20E20H test case.

source	relative	absolute	comment
beamline	1.0	1.0	grows with p
next slice	1.4	1.4	grows to 1.65%
pion selection	0.7	0.7	is 1.5% for π^-
energy scale	0.6	1.1	
plane material	<0.3	1.4	
absorber material	0.3	1.2	
cross talk model	0.5	0.5	
Birks' parameter	1.0	1.0	
PMT nonlinearity	0.9	0.9	
Step size	1.0	1.0	lowers MC
All others	0.7	0.7	
Total π^+	2.6	3.3	high E is 3.4 , 4.0
Total π^-	2.9	3.6	high E is 3.6 , 4.2

Table 1: The most important systematic uncertainties. The comment field identifies which ones grow with energy, or are significantly different for negative pions.

4.11 summary

When combined in quadrature, the systematic uncertainty on direct comparisons between data and MC are between 2.4% and 3.5% depending on species and energy, and the absolute uncertainty ranges from 3.2% to 4.1%, and is summarized in Tab. 1, including variations for positive and negative pions, and for low and high energy pions. The discussion above could accidentally be interpreted to mean that we have much investigation to do on our calibrations, which isn't totally fair. The other way to see it is that we proposed an experiment with systematics in the 2% to 3% range, and we have indeed succeded in beating down the systematics to this range.

Before proceeding with a discussion of the physics of pion interactions, these results have a few good candidate interpretations for how they should be applied to MINERvA neutrino analyses. [As of this draft, maybe I'm carrying some extra choice, but after a round of discussion on this draft, we will presumably eliminate one, or keep it in the text to show what our choices were..]

Flat energy scale systematic from discrepancy Because the data and test beam are mostly in agreement, but show as much as 4% disagreement at lower energies, we should use a 4% uncertainty on our simulation's modeling of the single particle response in the Eroica processing era. For Resurrection, we should take an additional 2% to 4% uncertainty in quadrature for the uncorrected Birks' and neutron effects. This is the most conservative choice.

Flat energy scale systematic from agreement If the MC and data agreed within errors, we would naturally take the 2.6 to 3.6% systematic uncertainty to be our constraint on the MC model. By some metrics, that all but two data points error bars (out of twenty-two) touch the systematic error band, we do indeed have agreement that would give an adequate χ^2 goodness of fit. Again, for Resurrection, an additional 2% to 4% uncertainty should be included. This is the most agressive choice.

Energy dependent systematic Since the naive reading of the discrepancy is localized to the low energy region, we could apply a 4% uncertainty for pions below 0.9 GeV, and the smaller 3.6% uncertainty above that. In this case, we might consider inflating it again to capture our current lack of data above 2.0 GeV. This is the choice with the most complicated implementation.

Nuanced interpretation We in fact see evidence the MC is not modeling something at 0.9 GeV and above, even though that is where the data and MC response have the best agreement. The latter might be coincidence, we may have an energy-scale systematic effect at the level of the uncertainty we show (or even a Geant4 bias) that applies to all data points which would bring the low energy data into agreement, and let the discrepant high energy data fall out of agreement. This is a good set of talking points, but is not a prescription for using this result in neutrino analysis.

5 Hadron physics and discussion

Overall, there is good agreement between the data and the MC, better than we were expecting. Good job Geant4, have a biscuit. Had there been a major discrepancy between the simulation and the data, we would have investigated what aspects of the Geant4 hadron simulation were in need of adjustment. In the present situation, there is one mild discrepancy, and we instead quantify what deviations from the Geant4 model we could have been sensitive to.

The first subsection contains a quick introduction to the basic source of fluctuations and the flow of energy to other particles in a hadronic shower. Following that is a description of the Geant4 prediction for containment. These introductory topics give more detail about the basic structure of the result and agreement presented in this technical note. Immediately after that, we discuss the pion mean free path effects (reaction cross section), how it could lead to the mild discrepancies we observe in the test beam data. Finally, there is short discussion on how our analysis is not sensitive to interaction fates, and what kind of analysis would be sensitive.

[Disclaimer**. The energy flow estimates are based on a 2012 study. Something has changed and now a bug prevents me from repeating them reliably in 2014. And even without the bug, there is a technical sensitivity to the Geant4 step size and how Geant4 saves information to our TG4Trajectory structure. These older estimates are useful enough to be a guide to the physics, but probably not robust enough to be cut and pasted into a NIM article without being redone and checked.]

5.1 calorimetric fluctuations and energy flow

Fluctuations are intrinsic to the calorimetric reconstruction of hadrons. An excellent reference for general and surprisingly specific discussion of calorimetry is Richard Wigmans' "Calorimetry: Energy Measurement in Particle Physics" (Oxford Science Publications, International Series of Monographs on Physics 107, Clarendon Press, Oxford University Press, 2000). Here we capture the parts relevant to the MINERvA case using the Geant4 simulation, which also cover a lower energy situation than most of the discussion in Wigmans' book.

At the interaction level, fluctuations are dominated by energy that goes missing when unbinding nuclei or to neutral particles that may leave the detector unseen. In contrast, a particle like a proton or pion that stops at the end of its range can expect to have its energy reconstructed with good resolution, about 10% in the MINERvA tracker.

The energy flow for these pions is illustrated in Table 5.1. This was constructed from the Geant4 truth information by stepping through truth trajectories and assigning each trajectory's ΔE to a specific particle or parent particle. To be clear, this is not the same as adding up all the MCHit truth objects from the active scintillator. Energy lost by an EM shower was attributed to its parent pion or gamma ray by tracing the original parentage of each particle in the shower. The energy attributed by Geant4 to

destination	mean
unbinding	17.9%
neutrino	4.4
neutron	6.5
primary	22.5
new pion	8.1
proton	24.5
pizero	8.9
muon	3.4
electron	1.3

Table 2: Approximate energy flow percent by species (from a 2012 study, see disclaimer), according to the Geant4 MC truth for the 20E20H configuration. The first three are sources of missing energy that will usually not be visible in the test beam detector The next set are forms of charged particles, including "primary" ionization by the original pion before a major interaction occurs, and separately ionization by secondary pions which can produce energy deposits.

unbinding nuclei is the difference between the outgoing and the incoming particles' energies at every inelastic interaction. The specific simulation is the 20E20H one used for analysis here, and the truth requirement that the pion did not interact upstream of the detector.

Energy flow to missing energy accounts for much of the 35% to 40% difference between our energy estimator and the true available energy. Another major portion comes from the electromagnetic component of the hadron shower. This component does not produce much missing energy, but in the ECAL the energy deposit in the Pb is much higher per g/cm^2 than it is in the CH scintillator plane. A similar, but milder effect occurs in the Fe HCAL. When the reconstruction applies a passive material correction based only on g/cm^2, this leads to an underestimate of the shower energy which is 75% to 80% of what it should be. Because the showering EM fraction of these pion events is typically 10% of the total, this should account for another 4 to 5% of the bias. Another component is lack of containment for charged particles, discussed in the next subsection, which is estimated to be another 5% or so. Lastly, a final contribution comes from particles that do deposit energy late, outside the time slice being analyzed. This portion is due primarily to slow neutrons and Michel electrons and is X%.

Compared to most of Wigmans' discussion, the MINERvA situation being tested with these test-beam data is lower energy and is more prone to fluctuations due to a single or few interactions. Also, as we step down to the $E_\pi = 350$ MeV pion energy threshold in our data, we pass below the threshold for multi-pion production. Also, we end up well begun up toward the pion absorption Δ peak at $T_\pi = 160$

MeV = $E_\pi = 300$ MeV = M_Δ-M_N (see e.g. Ashery et al., Phys.Rev. C23, p. 2173, 1981). At the lowest energies for pions, probably an increasing number of pions that survive to the end of their range without interacting, while at the highest energies the decay in flight fate is becoming vanishingly small.

5.2 containment, energy escaping the detector

Energy attributed to neutrinos obviously escapes, unbinding energy is obviously unobservable, and much of the neutron energy listed above escapes or is captured in later timeslices. Of the energy that is carried out of the active detector area, half goes out the side, 40% goes out the front, and 10% goes out the back. The amount of invisible energy escaping the 20E20H configuration increases with energy, doubling between 750 MeV and 2000 MeV.

Another source of fluctuations arises because even particles that would contribute calorimetrically escape the testbeam detector. Another study using Geant4 truth information indicates that for 750 MeV pions, only 3% of such energy escapes on average, including a steady leak of gamma ray photons and electrons (uncontained EM showers). Half of these photons (by energy, not by number) are soft < 20 MeV, the other half substantial. Occasional fluctuations of pions and protons that kink to the side or backward and escape happen 6% of the time, but Geant4 predicts that on average they carry away 40% and 15% of the total energy (with large RMS) when they do.

In the full MINERvA, the neutrino and unbinding energy is still missing energy. Some larger fraction of the escaping neutron energy will become measured energy deposits in the larger MINERvA detector, and some fraction of that latter case will be clustered with the event, and included calorimetrically.

5.3 mean free path

The most prominent feature in the response is a sharp trend to lower response at 0.9 GeV pion energy. This 0.04 shift across few data points for both pion polarities is mostly not mirrored in the MC which shows a smooth, decreasing trend. There is a hint in the π^- result that the MC is trying to model some feature at that energy. This feature in the data correlates with another feature: the probability that a pion leaves reconstructed energy only in the ECAL. One explanation of this discrepancy is generically the mean free path in the detector is too long in the MC, or specifically the pion reaction cross section in Geant4 is too low starting at 0.9 GeV pion energy.

The correct way to study this is with the Geant4 cross section modification implemented by Juan Pablo Velasquez and Trung Le and run a special purpose MC. As of this writing, there is something about the code or options files that is not testbeam safe. Such a situation constantly happens for testbeam analysis, and we've not yet resolved this one.

In place of the correct study, here are the results of a more hacky way to obtain the same thing. In this case, the modified MC is changed by looking at how deep in the detector the MC truth first interaction occurred, and randomly removing events (like a reweight, but instead a random picker) with a probability that increases with depth in the detector. In this way, the modified distribution approximates a reduction in the mean free path. This is similar in spirit to the kind of studies Brandon Eberly did for the pion analysis. Rik's implementation has an additional energy transfer < 10 MeV requirement to be considered the first interaction. The magnitude of the change here roughly corresponds to doubling the reaction cross section. The two plots in Fig 10 show the results.

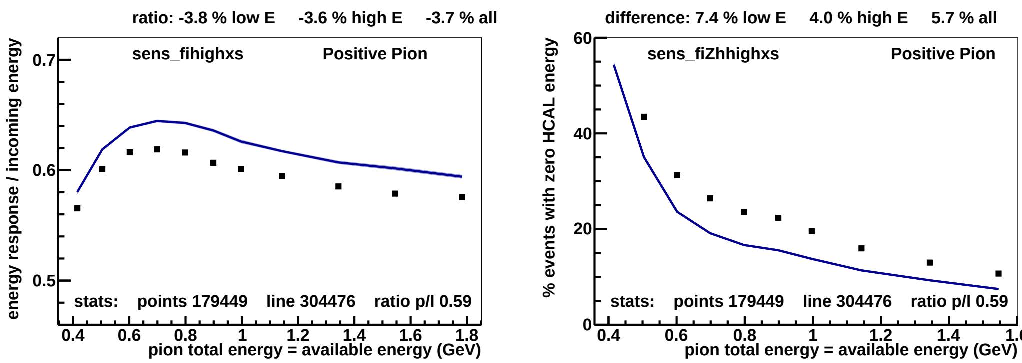

Figure 10: The change in response (left) and percent of events with zero HCAL energy (right) when reducing the pion mean free path using the picker/reject method.

Implemented this way for pions with energy above 0.9 GeV, this effect would lower the MC prediction by a percent or two and raise the zero HCAL component by a percent or two. Both these directions, but something with lesser magnitude than illustrated here would be adequate to bring the shape into significantly better agreement across the spectrum, leaving only an energy scale offset within the known uncertainties.

5.4 interaction fates

The main features of the Geant4 pion simulation are the probability to have an inelastic scatter, and the relative probability that the interaction produced one of a few particular fates. Following an interaction, there will be some number of hadrons in the final state. For this analysis, the Geant4 truth information is used to divide the pion sample into the following categories:

- two or more pions in the final state, "inelastic pion production" 15% on average, with threshold
- exactly one pion of different charge, "single or double charge exchange" 11%
- exactly one pion of the same charge, energy transfer > 10 MeV, "hard kink", 37%
- zero pions, any number of nucleons out, "absorption" 30 to 40%, more at low energy
- interaction upstream or nothing, 5%

- decay, 3%

The relative fraction of these, as well as the mean free path, affects the MC energy scale. For this analysis, this categorization of the MC is restricted to the first interaction in the detector, and in the case of a hard kink is restricted to an energy transfer of at least 10 MeV. Softer scatters than this are considered not the first interaction. Of course, the resulting fragments may continue to interact downstream, but that produces a classification decision tree that is uncomfortably complicated.

Of these categories, the absorption fate is the most dramatic. At 1.5 GeV, the energy response is 0.52 rather than the whole sample average of 0.60, confirming the expectation that this fate produces more missing energy than other fates. On the other hand, it has a similar response than the sample average at the very lowest energy, partly because this fate is increasingly more important at the 300 MeV Δ peak.

The multi-pion production fate has a threshold, and turns on slowly in the sample, and is fully on by pion energy of 1 GeV. At lower energies, this fate produces somewhat less visible energy on average, about 0.52, rising until eventually it produces 0.62 instead of the sample average of 0.60 fraction. The other fates are relatively mild mannered, producing a little more energy than the sample average, counteracting the large trend to below average response from the two above.

Despite these effects, the overall calorimetric response from each of these fates is really not so different, just a few to ten percent. Changing the probability of any one fate relative to the others, even by very large factors, produces results that remain within our error bars. Calorimetrically, we are not sensitive to the mix of interaction fates. Followup analysis, possibly with additional pion data, may be able to look into this more by tagging fates topologically instead of calorimetrically.

6 electron/pion response ratio

The tertiary beamline does not have very many electrons in it, and they are all at low energy. Will Bergan, undergraduate at William and Mary, has formed a clean sample of 100 electron events with energies from 300 to 500 MeV in the 20E20H detector. His results are described in another technical note [cite], and summarized briefly here. Four examples from the data comparing typical electron events (left) and pion events (right) are shown in Fig. 11.

[This is a rough telling of the story still.]

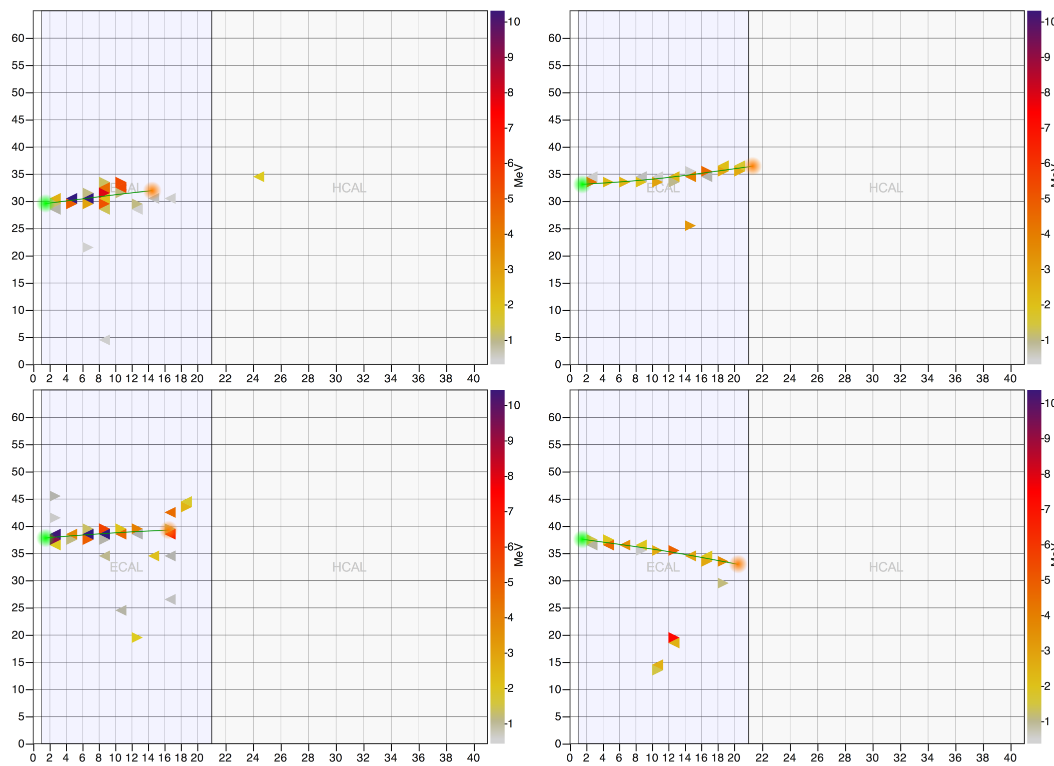

Figure 11: Typical electron (left) and pion (right) events. Note especially how the electrons start to shower early, and have reached shower max by plane 6 (the third X plane).

In his analysis, he compared the electron response in data and MC and finds the response in the MC to be systematically low, about 5%, possibly getting worse with energy. At the same time, the pion response shown here and in Bergan's independent analysis, shows the MC does a good job describing the data below 500 MeV.

The electron response (regardless of a data MC discrepancy) is higher than for pions, between 0.75 in the MC and 0.80 in the data, and an e/pi ratio of around 1.25 in the data, which increases at the lower energies at least because the pion response is decreasing toward the Δ peak. This leads to a double ratio (MC/data)(e/pi) of 0.90 at 500 MeV, increasing to 0.98, again illustrating the energy dependent discrepancy.

Show Will's 2D plots for electrons and pions.

7 Conclusions

The conclusions from the end of section four are repeated here. [Right now, cut and paste]

Flat energy scale systematic from discrepancy Because the data and test beam are mostly in agreement, but show as much as 4% disagreement at lower energies, we should use a 4% uncertainty on our simulation's modeling of the single particle response in the Eroica processing era. For Resurrection, we should take an additional 2% to 4% uncertainty in quadrature for the uncorrected Birks' and neutron effects. This is the most conservative choice.

Flat energy scale systematic from agreement If the MC and data agreed within errors, we would naturally take the 2.6 to 3.6% systematic uncertainty to be our constraint on the MC model. By some metrics, that all but two data points error bars (out of twenty-two) touch the systematic error band, we do indeed have agreement that would give an adequate χ^2 goodness of fit. Again, for resurrection, an additional 2% to 4% uncertainty should be included. This is the most agressive choice.

Energy dependent systematic Since the naive reading of the discrepancy is localized to the low energy region, we could apply a 4% uncertainty for pions below 0.9 GeV, and the smaller 3.6% uncertainty above that. In this case, we might consider inflating it again to capture our current lack of data above 2.0 GeV. This is the choice with the most complicated implementation.

Nuanced interpretation We in fact see evidence the MC is not modeling something at 0.9 GeV and above, even though that is where the data and MC response have the best agreement. The latter might be coincidence, we may have an energy-scale systematic effect at the level of the uncertainty we show (or even a Geant4 bias) that applies to all data points which would bring the low energy data into agreement, and let the discrepant high energy data fall out of agreement. This is a good set of talking points, but is not a prescription for using this result in neutrino analysis.

When applying these energy scale uncertainties to analyses of MINERvA neutrino data, we should realize that several systematics are already included in the response energy scale uncertainty, and do not need to be double counted. The error here includes the Birks' effect (except for Resurrection), non-linearity, and material model. This is not to say they should not also be considered, if any of these affect something other than calorimetric response (such as tracking, PID, or dE/dx based energy reconstruction), then they do need to be checked.

In most cases, the single particle response is just one element of the hadronic system response from a neutrino interaction, which also includes uncertainties and fluctuations coming from the Genie interaction model, its final state rescattering model, and its production of neutron final states and unbinding energy. However, the special case, coherent pion production, where the signal process is not affected by FSI, and the results presented here should carry over explicitly.

In conclusion, we have made a calorimetric response of our detector for pions with energy between 350 and 2000 MeV and systematic uncertainties from 2.6 to 3.6% (for data vs. MC) and 3.4 to 4.2% (absolute). We find the MC does a good job of describing the data. Our systematics are controlled enough that there is a modest discrepancy that starts in for pions with 0.9 GeV.