Activity Worksheet: My First Neutrino-Interaction Analysis

Cheryl Patrick, UCL

21 June, 2021

Many thanks to the master of CAFAna, Chris Backhouse (UCL), who helped to create the exercises for this activity.

For this activity, we'll be looking at simulation from the DUNE near detectors - specifically, the ND-LAr, which will use liquid argon as its neutrino target and detection medium, and the ND-GAr, which will use high-pressure gaseous argon.

For the group exercise, we'll divide into teams. Half of the teams will study the ND-LAr, and the other half will study the ND-GAr. We'll compare and discuss our results. If you're working on your own, you can use our example answers for one of the detectors - or run your code on both samples and compare!

Write down here which detector you will be studying:_____

Using ND-LAr simulation

The example scripts are set up to use ND-GAr simulation. To use ND-LAr simulation, you'll need to find the line `SpectrumLoader loader(NDGAR_FHC);` and change it to `SpectrumLoader loader(NDLAR_FHC);`

In this worksheet, the commands you need to type will be formatted `like this`. Anything formatted *`like this`* is something you need to substitute – with something like your login name, or the server you're using.

1 Getting started

As we said in the requirements, you'll need a Fermilab account with access to the DUNE or GENIE virtual machines. If you are in the hands-on class and don't have an account, you'll have to follow along with your team for now – you can try the exercises for yourself once you get an account. Here's how to get set up:

https://indico.fnal.gov/event/48900/page/2795-requirements.

Now you have your permissions, let's connect to the Fermilab server:

On Linux or Mac:

1. Open a terminal window.
2. Authenticate with

   ```
   kinit -f <your Fermilab username>
   ```

   and enter your Kerberos password when prompted.
3. Connect to the DUNE or GENIE server: `ssh -XY` *`server`*`.fnal.gov`
   - For DUNE users, your server will be `dunegpvm`*`xx`*`.fnal.gov` (The DUNE servers are numbered from 02 to 15, substitute any of those numbers for *`xx`*. If we all use the same one, it will be slow!)
   - For GENIE users, your server will be `geniegpvm01.fnal.gov`

On Windows:

1. Start Xming

2. Get a Kerberos ticket following the *Getting a ticket instructions* here:
   https://fermi.servicenowservices.com/kb_view.do?sysparm_article=KB0011316.
3. Connect to the server with PuTTY following the instructions in Connect to Kerberized Host using SSH Profile on the same page. Note that you can save your profile to make this easier!
   - For DUNE users, your server will be `dunegpvm`*`xx`*`.fnal.gov` (The DUNE servers are numbered from 02 to 15, substitute any of those numbers for *xx*. If we all use the same one, it will be slow!)
   - For GENIE users, your server will be `geniegpvm01.fnal.gov`
4. When your terminal opens on the Fermilab server, get another kerberos ticket on there, with the best chance possible of working: `kinit -f -p -A -r 8d -l 26h` *your username*`@FNAL.GOV`. Be sure to put `FNAL.GOV` in capitals.
5. You should now be logged onto the Fermilab servers. From now on, the instructions should be the same whether you're using a Windows, Linux or Mac computer.

Because you are all excellent students, I'm sure you've already set up your analysis area according to the instructions in https://wiki.dunescience.org/wiki/Summer_School_Software_Setup. I believe in you! So... we'll just switch to that work area now:

Do this every time you connect!

1. Go to your working area
   - On the DUNE machines: `cd /dune/app/users/${USER}`
   - On the GENIE machines: `cd /genie/app/users/${USER}`
2. Switch to the folder we're using for this school: `cd DUNESchool/`
3. Set up the analysis framework: `source lblpwgtools/CAFAna/build/Linux/CAFAnaEnv.sh`

The first time you try this, test that it has all worked:

- For DUNE users: `root -l /pnfs/dune/persistent/users/marshalc/CAF/CAFv5gas/CAF_FHC_9.root`
- For GENIE users: `root -l /genie/app/users/cpatrick/DUNESchool/CAFs/CAF_FHC_9.root`
- Enter `caf->Draw("Ev")` – it may take a few seconds or more to load, but a window should eventually pop up with a histogram - the neutrino energy spectrum from the CAF. If it does, great! Close that pop-up window, then back in the terminal, quit ROOT by typing `.q`

Congratulations! You're ready. Let's do some analysis!

2 Introducing CAFs

The files we'll be using here are known as **CAFs**: Common Analysis Files. Our CAFs are a DUNE-specific set of ROOT ntuples - a tree-structured format containing information about every simulated neutrino-interaction event (other experiments also have their own versions). ROOT is a CERN-produced piece of software designed to make it easy to work with this ntuple format and to use it to do physics analysis - producing histograms and other plots. You can find information about the CAF format here and general tutorials about working with ROOT ntuples here. For this tutorial, we'll be using some CAFs that have already been created for you, but you can use this page to find out how to make your own.

1. Let's look at one of these files right now.
   - DUNE users: `root -l /pnfs/dune/persistent/users/marshalc/CAF/CAFv5gas/CAF_FHC_9.root`
   - GENIE users: `root -l /genie/app/users/cpatrick/DUNESchool/CAFs/NDGAr/CAF_FHC_90.root`

   Ignore any `Warning in <TClass::Init>: no dictionary for class` warnings.

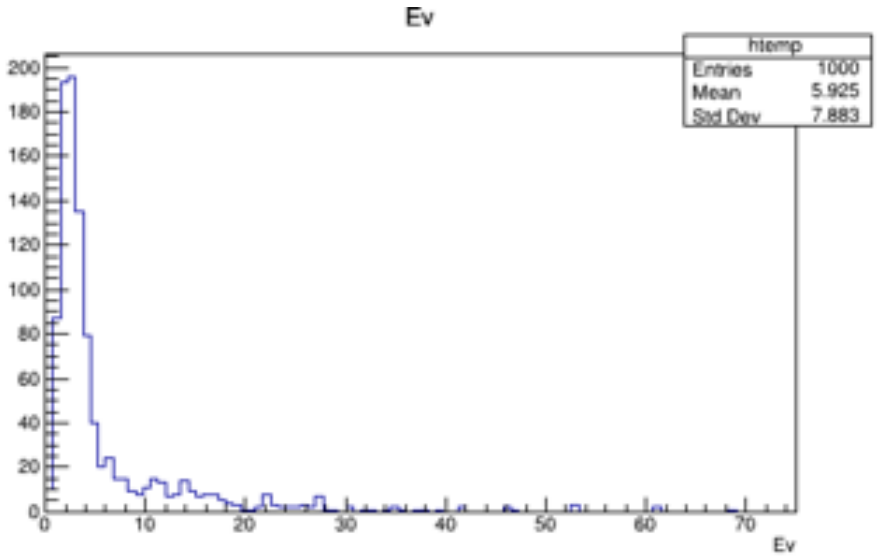

Figure 1: *True neutrino energy from a DUNE ND-GAr simulation CAF in the TBrowser window*

2. Once ROOT opens, look at the CAF: `TBrowser b`. A ROOT object browser window should open. In the left-hand pane, you'll see your file listed underneath ROOT files. Double click on its name to open it. It might be slow. Once it opens, double-click on `caf;1` to look at the CAF contents.

3. CAFs are in the form of a 'tree' called `caf` whose 'branches' correspond to information about a simulated (or real, when the detector has been built and the beam turned on!) neutrino-scattering event. For example, the branch `Ev` stores the energy of the simulated neutrino, in GeV (giga-electron-volts). When we click on this branch in our ROOT object browser, we'll see a histogram of all the neutrino energies, for every scattering event in the file. Go ahead and click it! It should look like figure 1. That tells you (roughly, as we only have 1000 events in this file) the spectrum of the DUNE neutrino beam. As you can see, most of the neutrinos in the beam are in the few-GeV energy range.

4. You might have more questions – how many of these 1000 'neutrinos' are actually antineutrinos? Are they of the muon- or electron-neutrino flavors? We can't tell from this simple browser plot. Instead, we'll be using DUNE's analysis software, CAFAna, to analyze the CAFs. You'll be running and modifying some template code to learn about the sort of information available in the CAFs, and how we work with it. We'll also use this to think about the neutrino-interaction physics we've learned up until now. Let's get to it!

5. Before you go – the branch `LepE` shows the energy spectrum of outgoing leptons (muons, electrons, or neutrinos) from the interactions. Find the branch. How does the spectrum compare to the incoming neutrino spectrum?

3 Investigating the neutrino beam flux

Neutrino experiments like DUNE use beams with a broad neutrino-energy spectrum. We'll be looking at CAFs generated in the *forward-horn-current* mode which selects a beam rich in muon-neutrinos μ_ν. However, there will be some contamination from electron-neutrinos, and from both muon- and electron-antineutrinos.

The list of CAFAna cuts is linked from our wiki page. What combination of cuts would you use to select the following?

ν_μ: _____ $\bar{\nu}_\mu$: _____

ν_e: _____ $\bar{\nu}_e$: _____

The script `Example1.C` uses the CAFs to draw the ν_μ flux - the energy spectrum of the incoming muon neutrinos. Let's try running it.

1. First, copy the example scripts to your own computer (you should be in your DUNESchool directory – if you aren't, switch to it now. Then, on the DUNE machines, get the scripts from GitHub:

```
git clone https://github.com/cherylepatrick/DUNESchoolInteractions.git
cd DUNESchoolInteractions/
```

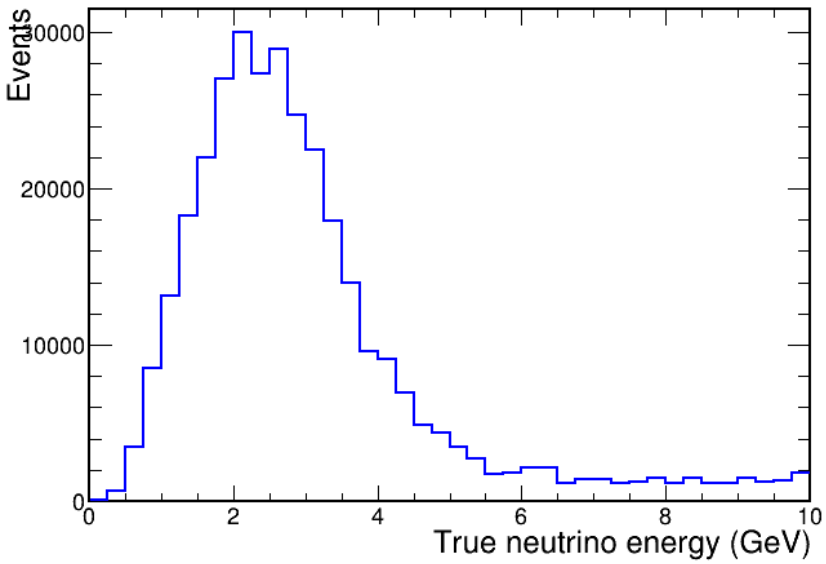

Figure 2: *True neutrino energy for ν_μ charged current interactions.*

If you're on the **GENIE** server, you'll need a modified version. Instead of getting the scripts from GitHub, do this:

```
mkdir DUNESchoolInteractions
cd DUNESchoolInteractions/
cp /genie/app/users/cpatrick/DUNESchool/DUNESchoolInteractions/*.C .
```

You'll notice solutions for each exercise - no peeking until you've tried them!

Running CAFAna scripts

Once you've done the setup listed earlier, running a CAFAna script is simple.
To run a script, type `cafe` *`scriptname`*`.C` at the command line. That will work for all the scripts in this tutorial.
When you are finished, quit the CAFAna environment with `.q`
By default, CAFAna will print any plots you make to the screen. You can speed it up by running it without graphics (any images you save to disk will still be created): `cafe -bq` *`scriptname`*`.C`. In this case, CAFAna will automatically quit when it finishes your script.
In future, when you're developing and testing a script, you might want to speed it up by running over only a fraction of your CAFs. To do this, we use the `--stride` option. For example, to test with only 1/5 of your CAFs, you'd type `cafe --stride 5` *`scriptname`*`.C`

2. Run the script: `cafe Example1.C` (Ignore any warnings.)
3. You should see a plot like figure 2, which should be saved in a file called *Example1.png*.
4. Let's take a look at that *Example1.C* script. You can use the `less` command to scroll through it, or open it in a text editor like gedit, nano, or emacs. There's some advice about them here.
5. Look for the lines that:
   - Load the data from the CAFs
   - Define the variable to be plotted
   - Define the binning
   - Define the event selection (cuts)
   - Set the axis title
   - Choose the color of the line on the plot
6. Use the code to answer:

- How many bins are there?
- How many protons on target are we modeling?
- Where are the input files?
- Optional bonus: Using the `ls` command on the folder of inputs, find how many files you've been looking at. Change the code to run with only one of these input files. How does the output plot change?

7. Quit CAFAna with `.q`
8. Now edit *Exercise1.C* (there's advice about text editors here). It's a copy of the script you were just looking at, with some extra comments to help you. Your task is investigate the beam contamination. Make `Spectrum` objects for the contributions from $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$, and draw them on the same plot. I've put in some code to help you get started, and marked the bits of code you need to edit/add, or should pay particular attention to, with asterisks. Use your program, the plot you draw, and your knowledge of neutrino interactions to answer:
   - Which contamination do we have most of: $\bar{\nu}_\mu$, ν_e or $\bar{\nu}_e$?
   - Do all the components have the same energy spectrum? Why might that be? (This is hard to answer unless you know how a neutrino beam is produced. Ask your team-mates if you don't know...)
   - These show the relative spectra of *interacting* (anti-)neutrinos. Do you think the spectra of *all* produced (anti-)neutrinos will be the same? Why (or why not)?
   - I've set the plot to a log scale. Try turning it off to see why I turned it on.
   - Optional bonus: change the argument to the SpectrumLoader to point to reverse horn current (antineutrino-rich) CAFs. Switch it over and re-run the script (don't forget to change the output png file name so you can compare). Which beam suffers from more contamination - the FHC (neutrino) or the RHC (antineutrino) beam?
   - How might we tell a neutrino from an antineutrino interaction? Will this be easier in the ND-LAr or the ND-GAr?
   - Optional bonus: We've plotted these as lines without error bars, and scaled to a realistic beam exposure - but have we really used enough simulation to understand the beam? Try changing the plotting style from "HIST" to "E" to add error bars. What does this tell you? How much simulation might we need for a real analysis?

4 Quasi-elastic-like scattering

In week 1, we looked at the different types of neutrino scattering. We're going to think about quasi-elastic scattering of muon-neutrinos. (So make sure to use a neutrino-rich FHC sample!) What particles would you expect to see in the final state? (Hint: draw the Feynman diagram. You should have 1 lepton in your final state, plus one or more hadron(s).)

If you don't know, look it up in the slides from the "Introduction to interactions" talk.

1. Use the CAFs to draw the true neutrino energy spectrum of quasi-elastic events. To do this, you need to make a cut (also known as an event selection) to select only ν_μ events where the true interaction type is quasi-elastic. Put that cut into the script *Exercise2.C*. Look for the asterisks to see where you need to edit.
2. In real life, we don't know the true interaction type, so we have to rely on the particles we can detect to try and identify a type of interaction. We'll use the information from the diagram we just drew. Make a second histogram on the same axes, copying the one in the script. This time, instead of the cut on true quasi-elastic events, we'll make a cut (or cuts) selecting events where the *true final-state particles match the signature for quasi-elastic muon-neutrino scattering*. You'll find that in the example, defined as `kHasQEFinalState`. Read through the code and check you understand it. Then plot a second line on the same plot to show the spectrum you get using this cut.
3. Compare the two spectra. Are they the same?
4. You might notice that there are fewer events when you look for the final state signature than when you look for true CCQE events. The reason for this is that when a proton makes its way out of the argon nucleus, it will often re-scatter, transferring its energy to other nucleons. For this reason, experiments will often use a slightly broader definition for QE-like events, which we call CC0π (one muon, at least one proton, no pions). Try plotting the neutrino energy spectrum with the CC0π cut applied. How does it compare to the two spectra you already plotted?
5. Some questions:

- Can you remember the name for the process that knocks additional nucleons out of the nucleus? We're looking at simulation for ^{40}Ar. Do you think we'd see more nucleons in the final state, or fewer, if we were studying ^{12}C?
- It's unsurprising that your CC0π cut allows in more events than your $1\mu^-1p$ cut. But it doesn't match your true CCQE signal either. What could be contributing to your CC0π sample, apart from true CCQE events?

6. We don't need to guess what's contributing: we can find out from the CAFs. The script *Example2a.C* shows you how to select only the CC0π events that are true QE, and puts them into a *stacked* histogram - a ROOT object that lets you stack multiple histograms on top of each other. Edit the script to also select the CC0π events that were generated by resonant interactions, deep inelastic scattering, and scattering from correlated pairs joined by meson-exchange currents — then add those Spectrum objects to the stack.

7. Questions:

   - Which of these interaction modes would produce a CC0π final state, even without final-state interactions?
   - How would you make a CC0π final state from a resonant interactioin followed by FSI?
   - Are particular modes making more CC0π events at certain energies? Why might that be?
   - All these events have nucleons in the final state. Can you think of any properties of those nucleons that a future study might use to distinguish between the different interaction modes?

8. If these concepts are unfamiliar to you, or you're not sure of the answers, check out the slides from the *Introduction to neutrino interaction physics* talk to refresh yourself!

5 'True' vs 'Reconstructed' quantities

So far, we've been looking at 'true' information. By that we don't mean anything about the actual true physics — we mean the quantities simulated by our Monte Carlo interaction generator software, GENIE. We've been looking at distributions of 'true' neutrino energies to study how the GENIE predicts that they will interact.

In a real detector, of course, things aren't so easy. We certainly don't know something like the neutrino energy! We have to *reconstruct* it from the final-state particle information. Let's think about these quasi-elastic events again. To keep it simple, we'll focus on that sample with the classic final-state signature - 1 μ^- and 1 proton.

1. Think about conservation of energy. Write down an equation for the energy of the neutrino. You'll want to consider:

   - Final-state particle energies
   - Binding energy of the knocked-out nucleon (28 MeV)
   - The mass of any particles created

2. Starting from *Exercise3.C*, code a new `Variable` to correspond to the neutrino energy, calculated according to this formula. As an input, use the true final-state muon and proton energies. For these calculations, you'll need to use the variables in the CAF, through the `StandardRecord` object. You can find what variables are available at https://wiki.dunescience.org/wiki/CAF_ntuple_format .You'll see that some particle masses have been defined in the script for you. Plot this Spectrum.

3. Sadly, of course, a real detector wouldn't be able to measure the true proton and muon energies - we'd go through a complex procedure of detecting energy deposits in the detector, assigning them to particle tracks, and using what we know about their physics to *reconstruct* their energies. As we don't yet have a real detector, the CAF-maker software does a mock version of this. Plot another Spectrum on the same axes as your previous one, but this time, instead of the true proton and muon energies, use the reconstructed energies. What difference does it make to the plot?

4. The CAF maker will actually attempt to reconstruct the neutrino energy for you! Add a third Spectrum, plotting the neutrino energy from the CAF.

5. In the *Introduction to neutrino interaction physics* talk, we discussed it's possible to reconstruct the neutrino energy from just the muon kinematics. Let's add a fourth Spectrum, with the neutrino energy calculated with the quasi-elastic reconstruction formula (use the reconstructed muon and proton energies).

$$E_\nu^{QE} = \frac{m_p^2 - (m_n - E_b)^2 - m_\mu^2 + 2(m_n - E_b)E_\mu}{2(m_n - E_b - E_\mu + p_\mu \cos\theta_\mu)} \quad (1)$$

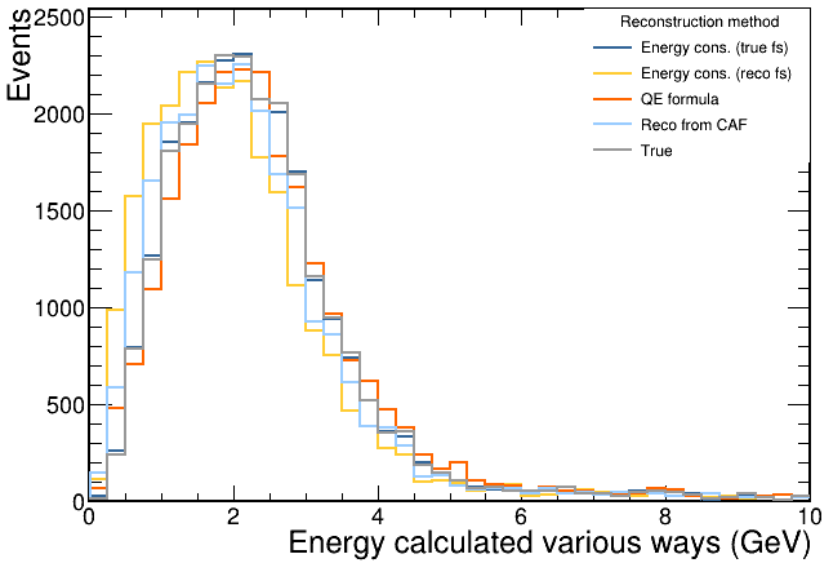

Figure 3: *Reconstructed energy with a larger sample of simulation (ND-GAr)*

6. Finally, plot the true neutrino energy Spectrum.
7. Some things to think about:
   - Do any of your manual reconstruction techniques match CAFAna's reconstruction?
   - Which reconstruction technique does the best job of reproducing the true neutrino energy? Why? Is it a valid technique you could use in a real detector? If not - which technique would you choose?
   - We've selected events with a true 1-muon, 1-proton state. Could we do that with a real detector? What might we do instead? (You can try coding it if you like!)
   - The QE formula was derived using momentum conservation. You also tried a reconstruction based on energy conservation. Do they agree? If not, does that mean that either the law of energy or of momentum conservation is being broken? Or is there some other physics going on here?
   - Do you think that you have used enough simulation to fully interpret your results? You can try re-running with more files, if you like, or look at figure 3 to see the spectra I got using 10 times the amount of simulated data.

6 Optional bonus tasks

Are you having fun yet? If you want to investigate the CAFs more, try some of these ideas. (We haven't tried these ourselves, so there are no solutions available, but feel free to chat about them in the drop-in sessions.)

1. Do these techniques assume that we have a particular physics process going on? Try splitting your different reconstructed distributions up into the QE, MEC, RES and DIS components (as we did before). Do some reconstructions work better than others for different modes? (I would recommend abandoning the one where you use the true muon and proton energies at this point, as it's artificial.)
2. As we found out earlier, the 1 muon-1 proton final state misses a lot of the QE sample. We'd rather use the CC0π sample. How might we modify our energy-reconstruction algorithms for use with that sample? Try the energy-conservation and QE-formula reconstruction on that sample, and compare to the true and reconstructed neutrino energy from the CAFs. Which method works better? (Again, ignore the method where you use the true muon and proton energies.)
3. For the brave (or at least, for ROOT experts): try making a 2-dimensional plot of true vs reconstructed energy for each reconstruction method (smearing matrix), or a 1-dimensional histogram of the difference between the true and the reconstructed energy (the energy resolution). Which reconstruction method performs best? To make a 2-D Spectrum, you'll need to try out some different methods from CAFAna. You can find the CAFAna code in two github repositories:

- https://github.com/cafana (Common to various experiments' versions of CAFAna)
- https://github.com/DUNE/lblpwgtools (DUNE-specific)

You might want to start by looking at the options for a Spectrum.