JET GENERATION IN CLIC

MATT ZHANG

Generating Samples

- As shown last time, we have a process for taking HepMC generator samples (made in Pythia), and using the CLIC environment to extract calorimeter hit information (after simulation and reconstruction).

- For each event, we end up with info regarding:

  - ECAL [25 x 25 x 25]

  - HCAL [5 x 5 x 60]

  - Object ID and properties

Generating Samples

- Using the ee➔jj HepMC file Maurizio provided last time, we first convert to HEPEvt format.

  - HepMC and HEPEvt data available here (/afs/<u>cern.ch/user/m/</u> <u>mazhang/Projects/JetCalo/Data</u>)

- After this, we feed the HEPEvt sample into Mokka for simulation, and feed the resulting output into Marlin for reconstruction. This gives us a ROOT file with information on detector response.

  - Available here ➔ /afs/<u>cern.ch/user/m/mazhang/Projects/</u> <u>JetCalo/Output</u>

- Continued…

Generating Samples

- After this, we take two Python scripts and convert the ROOT files into H5 format. These scripts "zoom in" on regions of interest, selecting 25x25x25 ECAL and 5x5x60 HCAL regions to save in the H5 file. These sizes can be changed.

- These scripts were originally written by Kaustuv Datta, Jayesh Mahaptra, and Nikolaus Howe. Maurizio has extended the original singular script by splitting it in two. This is to get around an issue with Python dependencies in the ILCSOFT environment.

  - The ILCSOFT version of Python does not have all necessary dependencies for creating H5 files, so first we convert from ROOT to a txt format.

  - Then we switch back to the lxplus default Python and convert from txt to H5.

 - Scripts here ➙ /afs/cern.ch/user/m/mazhang/Projects/JetCalo/ConversionScripts

Examining H5 Files

- Looking at the output H5 samples, I noticed that the ECAL and HCAL arrays were mostly empty. The generator reconstruction ROOT files were good, so seemingly something was going wrong in the script.

- The Python script works by finding the weighted center for each event, then taking an ECAL and HCAL window around there.

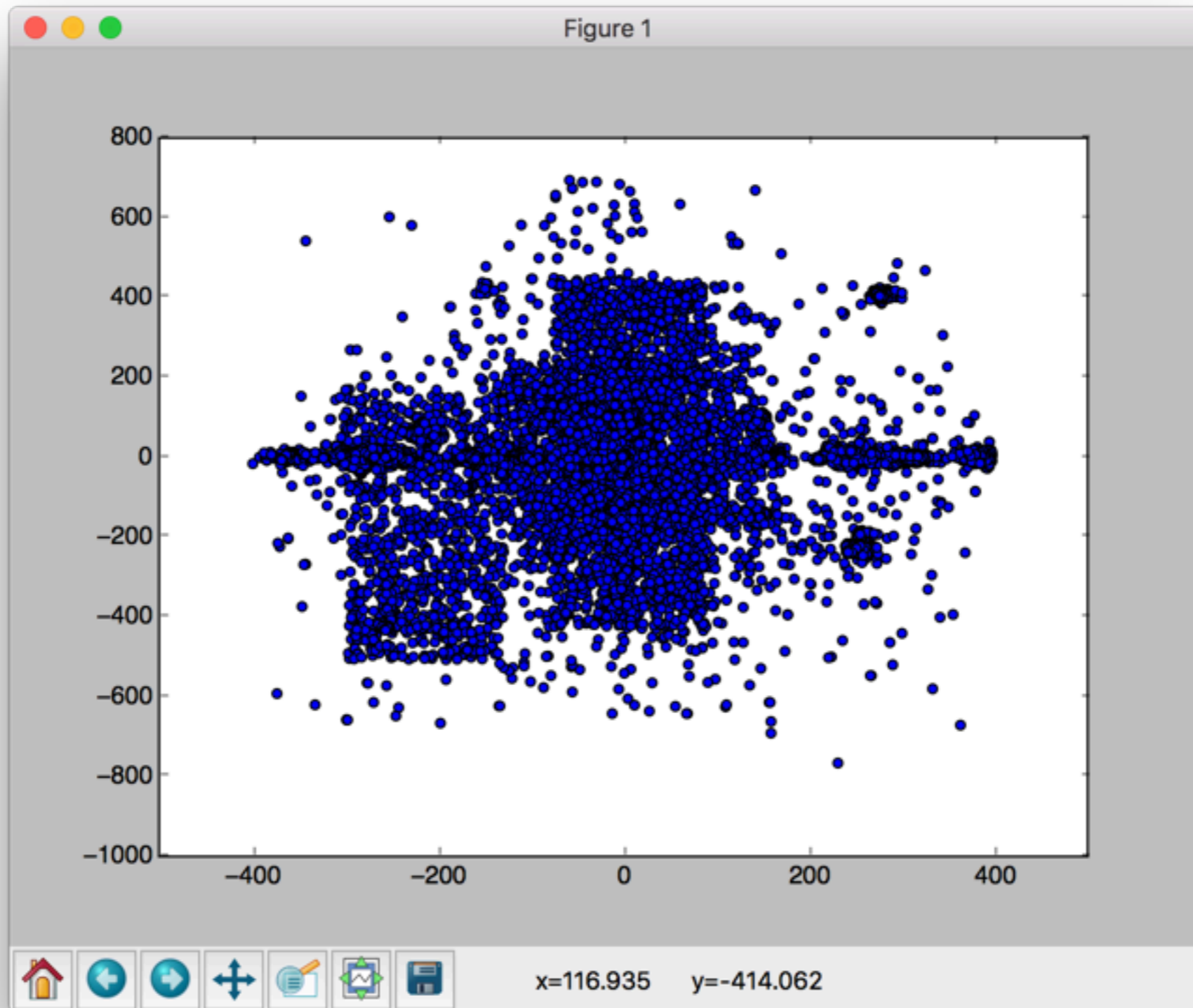- I plotted the x and y distances away from the weighted centers for all hits in all events.
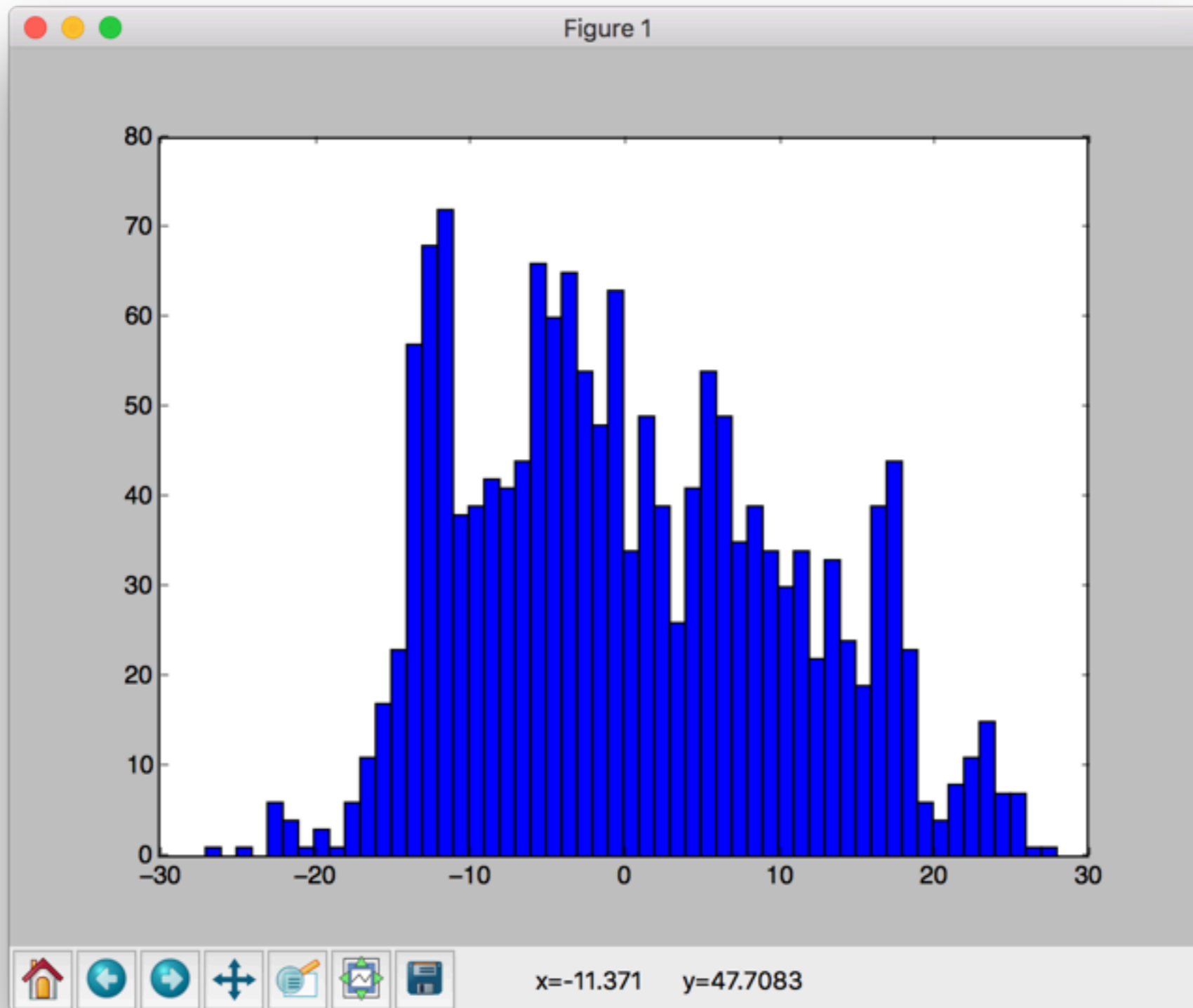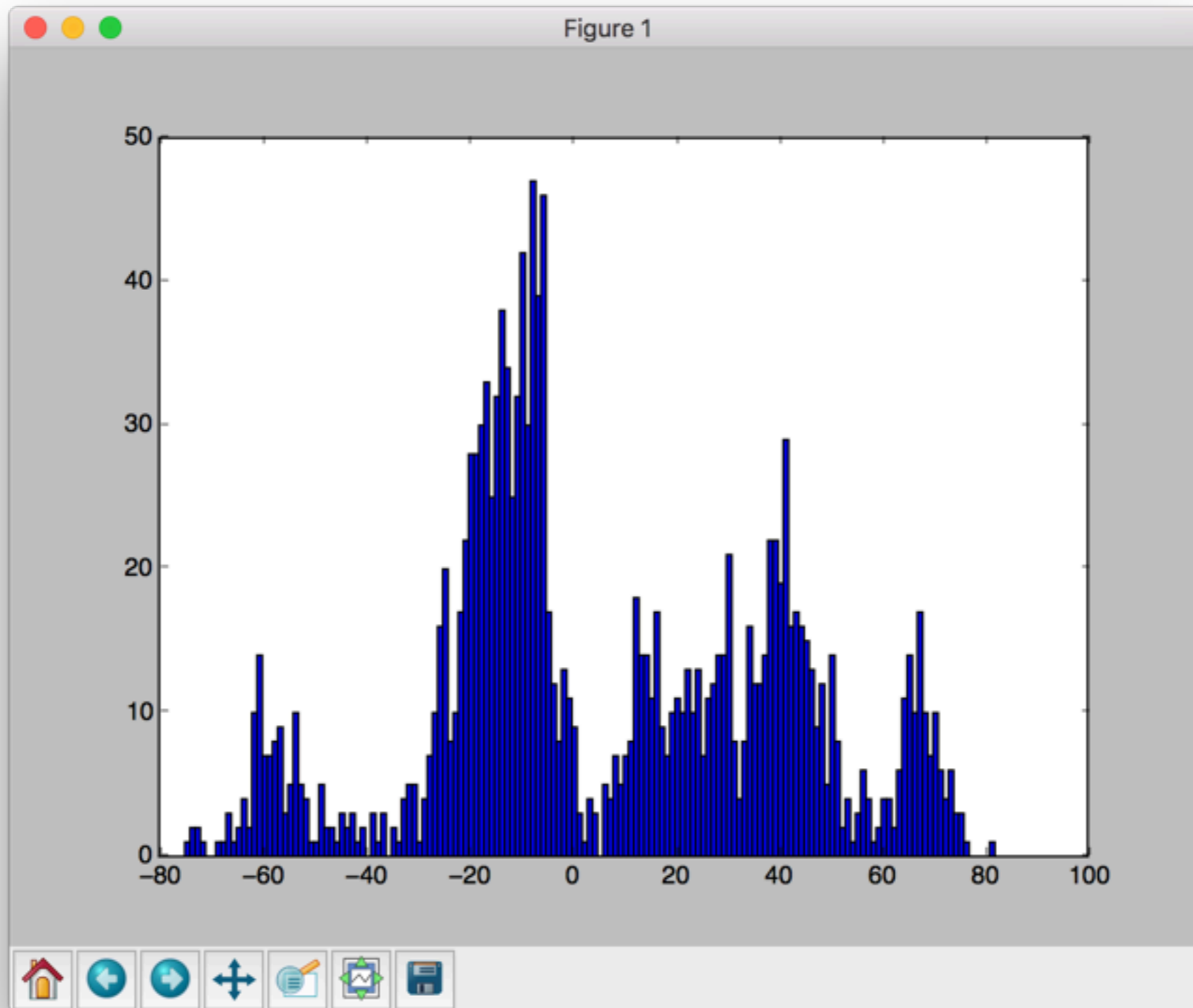
ECAL, dX
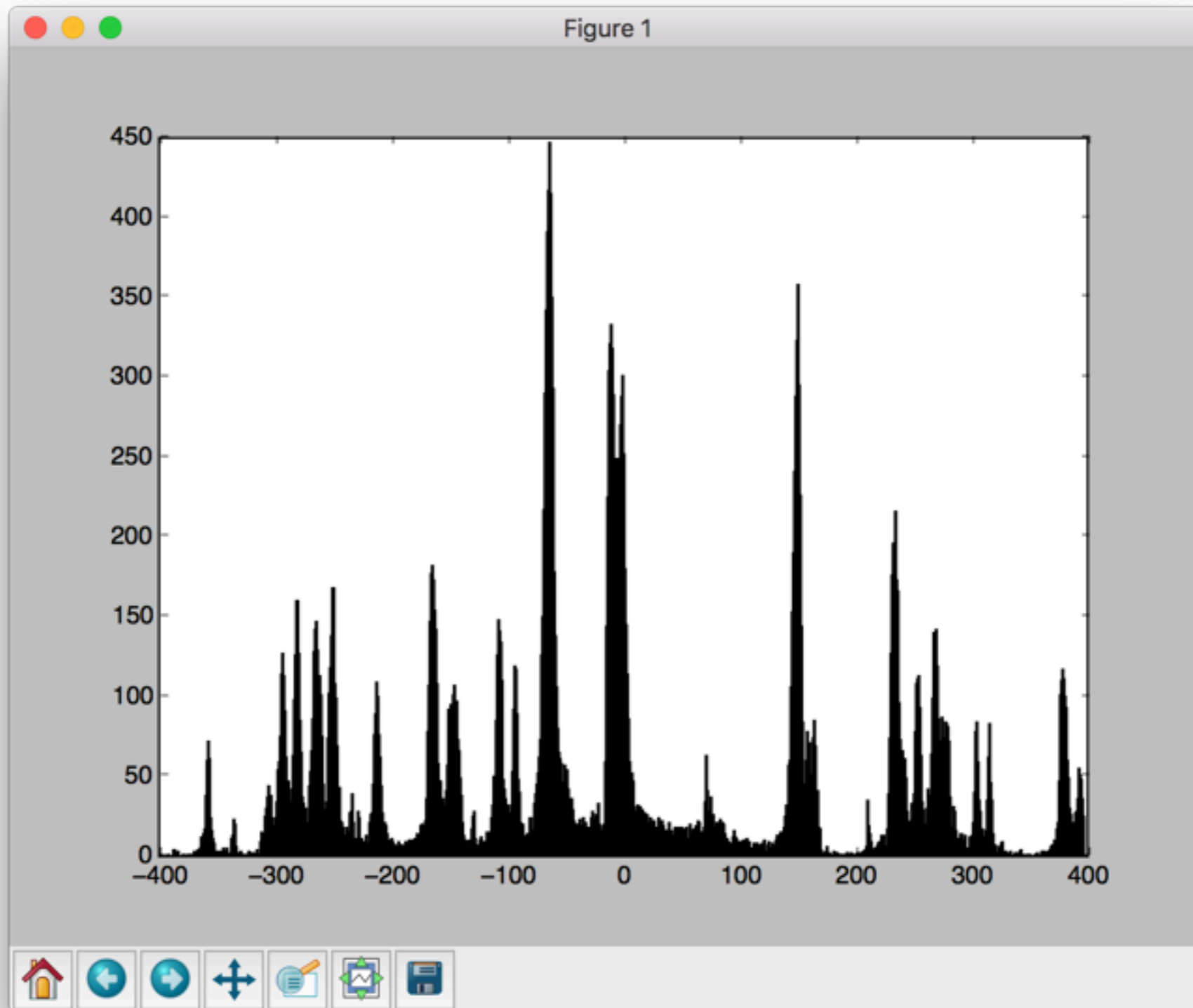
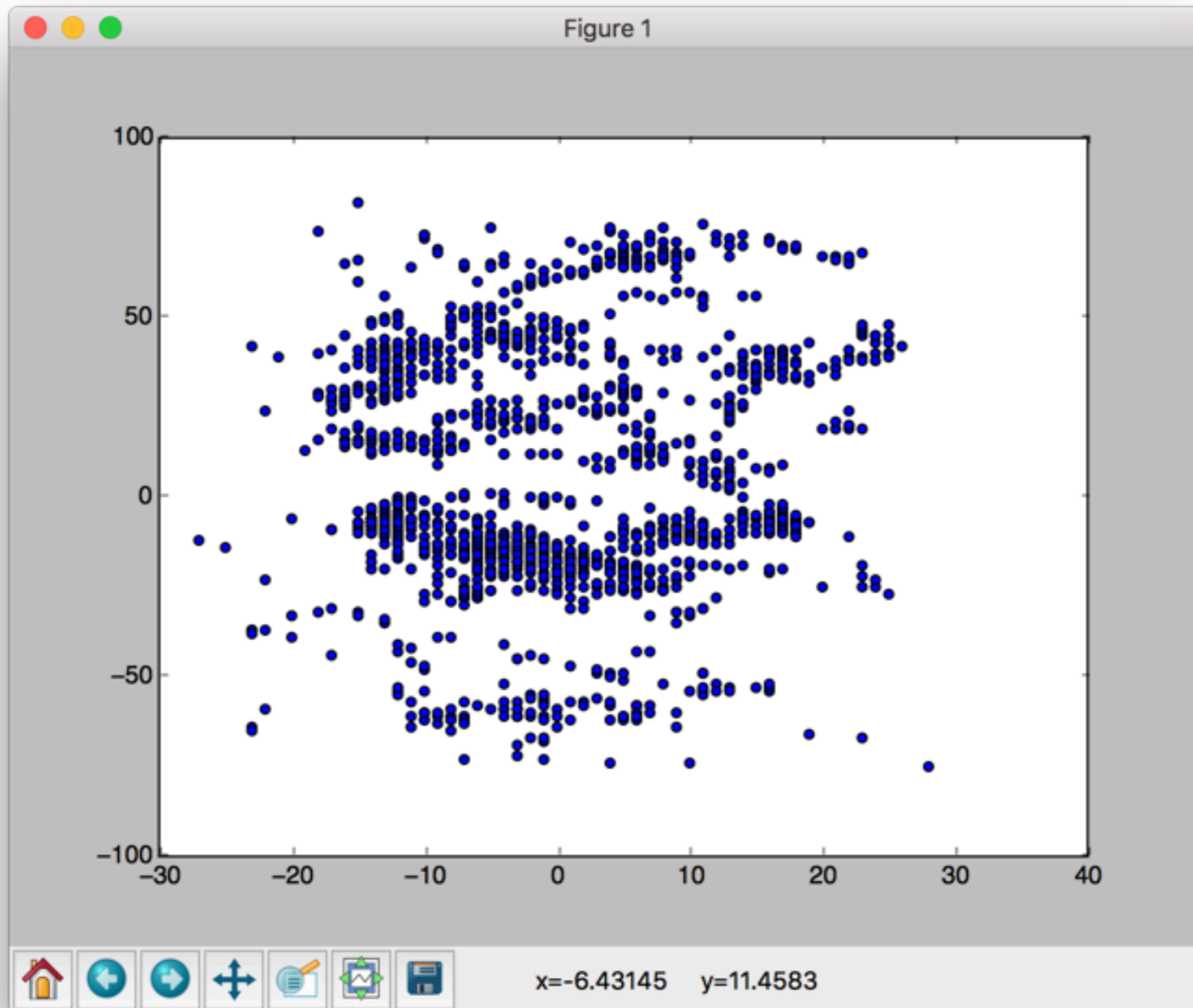# ECAL, dY
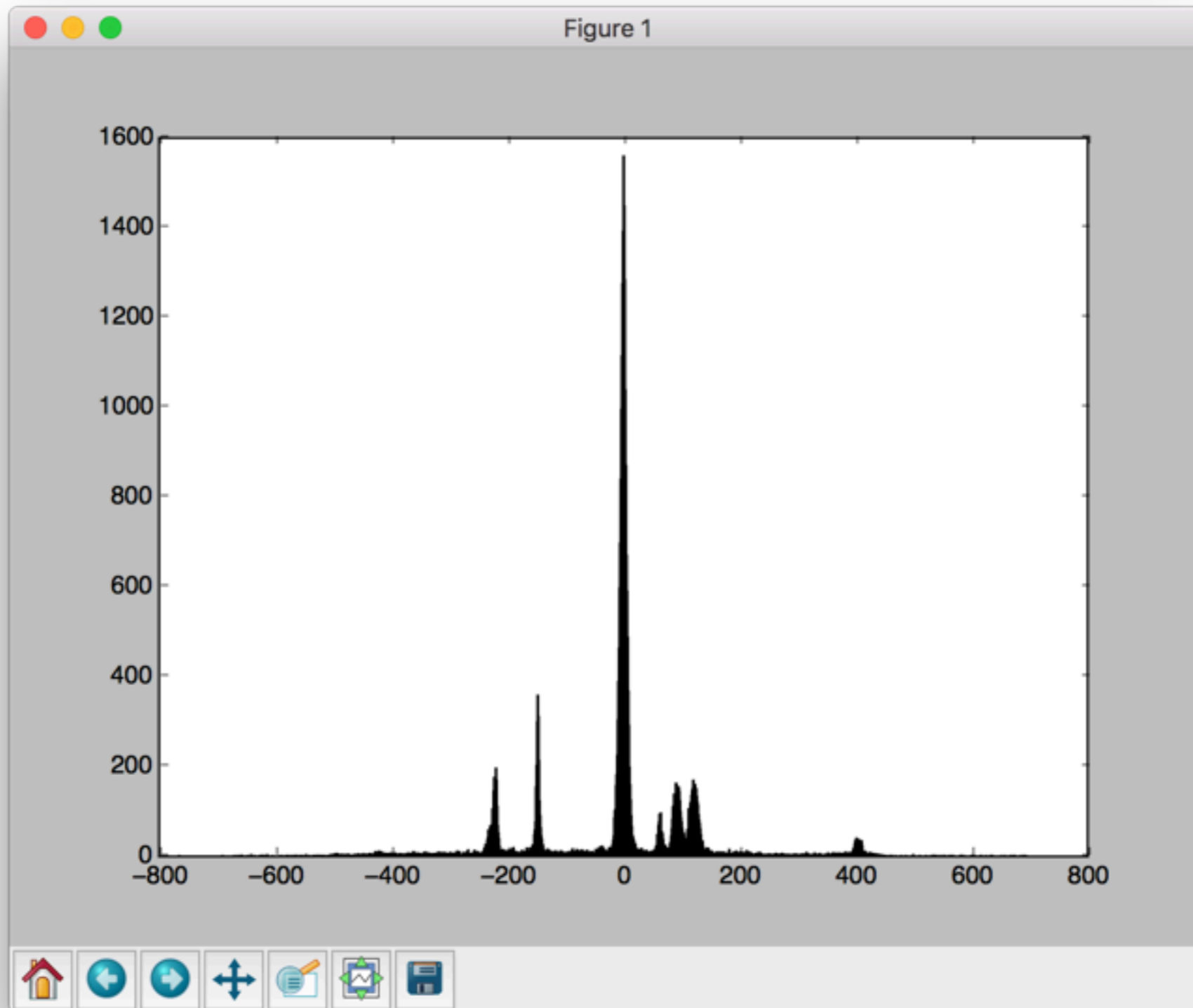
ECAL, dX vs. dY

HCAL, dX

HCAL, dY

HCAL, dX vs. dY

Thoughts

- Clearly, the ECAL and HCAL windows are not capturing relevant bunches of hits.

- This is probably due to the calculation of a single center for each event - there is no clustering beforehand.

- Before, when we had single particle guns, this did not give us problems, so probably nobody noticed. But now we have ee→jj.

- Two branch attack:

  - Find other software for reconstruction. For example, Marlin interfaces with FastJet, so we can feed in the ROOT files and get reconstructed jets to work with (for traditional object selection and for BDT studies).

  - Figure out a clustering method, since we need the ECAL and HCAL slices for neural net training. We can either implement a traditional clustering algorithm, or run an unsupervised machine learning algorithm.

Next Steps

- Generate some reconstructed jets using FastJet.

- Code up a cluster-finding algorithm.