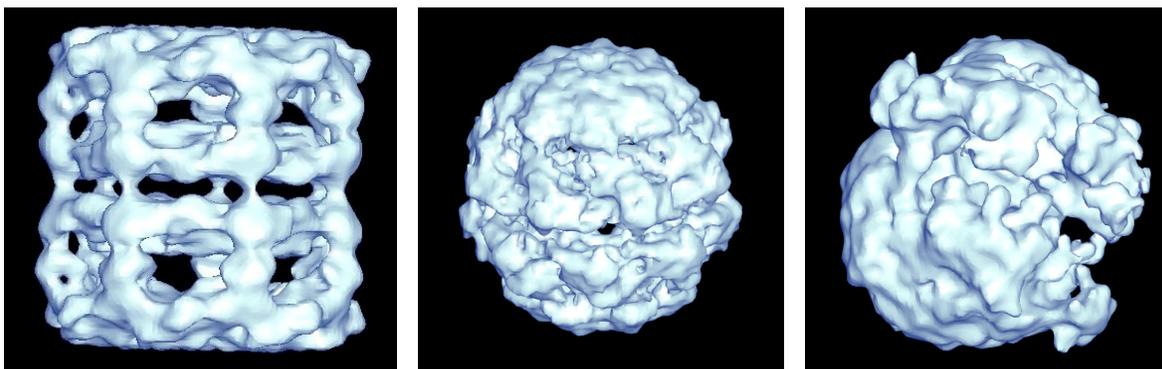


EMAN2 Reconstruction Tutorial

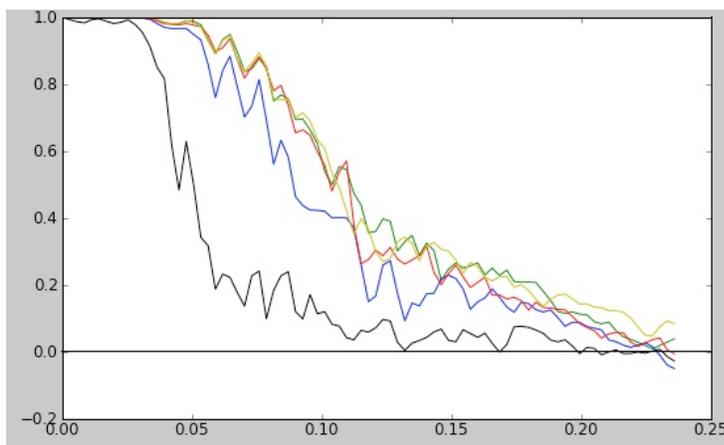
Note that this tutorial was updated on 3/15/2011 to reflect any changes in the interface, etc. It should not be used with versions of EMAN2 prior to 2.01.

Evaluating your refinement

- Hopefully you managed to start a refinement before finishing the first day, and you will now have a complete preliminary refinement to look at. Select **3D Refinement**. This will bring up a list of all of the completed refinements you've run.
- Double click on the single item you should see in this list. This will open a 3-D display showing you your preliminary refined structure. middle-click on it to get a control panel, and play with the isosurface threshold. It should look something like:

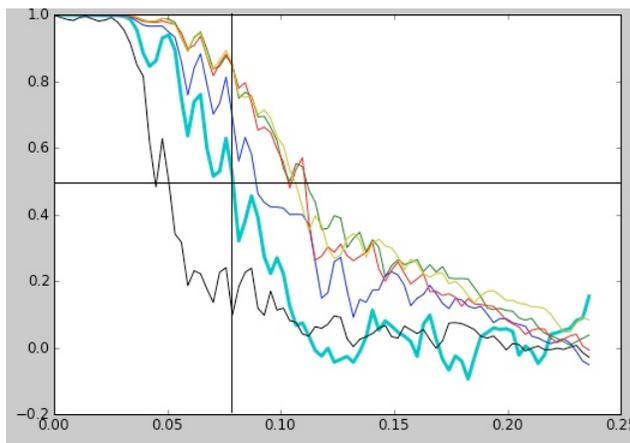


- Now, if it doesn't look exactly like this, or as good as these, don't panic. You may still be able to get a very nice final reconstruction with some additional refinement. You probably shouldn't start high resolution refinement until you have a reasonable low resolution structure though.
- Next, close this window and select the **Resolution** item. Again, double click on the single item you see on this list. This will give you a plot containing a number of FSC curves. Before you get too excited, these curves are NOT used for measuring the resolution of your structure. These are NOT FSC curves computed between even and odd halves of your data. Rather they are computed by comparing the model from one refinement iteration to the next, to monitor convergence of the iterative process. Eventually these curves should stop changing from one iteration to the next. If we had *true* convergence, then the map would be completely unchanged from one iteration to the next, and the curves would be 1.0 at all resolutions. Due to noise, this never happens, and you will see something like this:



If you open the control panel you can see which color represents which comparison. The black line compares the initial model to the 1st iteration, the blue line the 1st iteration to the 2nd, ... You can see that the process had already begun to converge after 3 iterations. This curve can NOT tell you resolution.

- Computing resolution requires running an even/odd test, where 2 halves of the data are used to produce (somewhat) independent 3-D maps, then these maps are compared by FSC. This curve CAN be used to measure resolution. Select the **Run e2eotest** item.
- ➔ It is a bit pointless to compute a resolution curve on this roughly refined result, and the computation does take a few minutes to complete. If you would like to start a better refinement running, you may skip this step for now. If it seems unlikely that you will have time to produce a proper refined model, you may want to learn how to do this.
- The resulting dialog will have a subset of the tabs that the e2refine dialog contained. Select the **Directory and iteration**, then set everything else in this window to the same values you used for e2refine. Press **OK**
- When this finishes running, opening the Resolution plot again, will now show an additional thick line:



The thick line at some cutoff value (0.5 shown) can be used to assess the resolution of your structure. Note that resolution is distinct from resolvability (perceived level of detail). A map at 10 Å resolution low-pass filtered to 20 Å resolvability would still assess as 10 Å resolution. Most maps must thus be filtered so the level of visual detail corresponds to the assessed resolution. We do not wish to do this at this preliminary stage, but on our final reconstruction, we need to take this step.

- Finally, we should look at the distribution of particle orientations for our reconstruction. Select the **Eulers** item. This will open the Euler angle explorer. Drag with the left mouse button (not clicking on one of the cylinders) to rotate the view. You may also want to zoom out a bit. Each cylinder represents a single class-average. The height of the cylinder tells how many particles were in that particular average. The viewer shows you one asymmetric unit of a single iteration of a refinement at a time. Middle-click to bring up the control panel and select which iteration of which refinement you want to view. If you are refining ribosomes, you will see an entire hemisphere. For GroEL or mm-cpn, you will see a narrow triangular spherical section extending from pole to equator.
 - If working on GroEL or mm-cpn, try selecting **Mirror** and **All syms**. This will show you the orientations for the entire unit sphere (turn them off when you're done, as it will slow things down a lot).
 - If working on ribosome, it can be difficult to visualize all sides of a hemisphere at one time. If you press **f** on the keyboard it will toggle 'flat' mode, where the hemisphere is flattened on to the plane, making it easier to see the distribution.
 - Left-click on a cylinder to bring up an image display. The image shown will be the projection of the map in that orientation. Pressing up arrow, will show the corresponding class-average with matched 1-d structure factor. The 3rd image is the original unmodified class-average. If you see significant differences between image 1 and 2 this represents a discrepancy between the projection and class-average in that orientation, and should be kept in mind.

High Resolution Refinement

- We aren't likely to have enough time at the workshop to make much of an attempt at this, but if you are going through this tutorial on your own at home, the next step is to try to improve the resolution of our reconstruction. This refinement will take quite a bit of time to run, particularly if you are working on ribosome.
 - Once again, select **Run - e2refine**, and select the **Phase flipped-hp** particles to use.
 - **Particles:** Instead of using the **small** set, this time you should select the **all** set of particles. I'd suggest trying ~5 iterations.
 - **Model:** We will need to add our preliminary refined model as a reference and select it. Press **Browse To Add**, and add your **refine_01/bdb/threed_filt_XX** structure.
- ➡ If your preliminary structure has the wrong handedness (we have not determined handedness), you can flip the handedness of the reference here before starting the full refinement. Note that it won't make any difference resolution-wise, but will save you from having to do it later. Here is an example command:
- ```
e2proc2d.py bdb:refine_01#threed_filt_04 flipped.hdf --process=xform.flip:axis=z
```
- ➡ If you do this, of course, you will want to add **flipped.hdf** to the Model panel and use that as an initial model.
- **Project 3D** - We'll need to reduce our angular step (**value:**) to get a decent resolution structure:  
GroEL - 1.607, mmcpn - 2.0, ribosome - 2.5
  - **Simmx** - All 3 comparators should be set to **frc** with **zeromask=1:snrweight=1**
    - shrink=1
    - otherwise same as before
  - **Class averaging**
    - averaging iterations = 1
    - class separation = 5
    - keep = 1.5, sigma-based
    - Averager = ctf.auto with set proj SF
    - All 3 comparators **frc** with **zeromask=1:snrweight=1**
  - **Make 3d**
    - pad = 192
    - keep = 1.5, sigma based
    - set sf checked
    - post processor = filter.lowpass.gauss
    - params cutoff\_freq=.08 (this corresponds to ~12 Å, for mmcpn/GroEL, you can try .125)
  - **OK**, then wait a reaaaaly long time (or use a cluster or multi-core workstation)

## Additional topics we may cover

- Parallel computing
- 2D analysis, dynamics, pseudotime movies
- e2refinemulti - multi-map simultaneous refinement
- Variance maps - e2refinevariance.py

## Particle Picking

We skipped particle picking in day 1 of the tutorial, however, it's still an important thing to understand, as is micrograph evaluation. While we may not formally cover this material depending on available time, but you can go through these steps, time permitting.

- **cd to workshop/eman2/boxing.** This directory contains 5 raw scanned micrographs from the GroEL data set. They are not all high quality images. Part of this task will be to assess image quality.
  - run **e2workflow.py** to create a project in this directory
  - There are 2 possible forms for bringing raw data (micrographs or CCD frames) into the project. You can click directly on the **Raw Data** task, or you can use the **Filter Raw Data** subtask. The first option will leave the original data where it is, and simply reference it. There may be some limitations to this approach, and if you ever move the original data, you will no longer be able to access it from the workflow. If you use the **Filter Raw Data** task instead, the data will be (optionally) filtered, and imported (copied) into the project. After this, if you wish, the original files can be deleted. If you keep the original files as well, this approach has the side effect of taking 2x as much disk space for the raw data. Select one of the following:
    - Select the **Raw Data** task. You will see a form appear with an empty table. This form is a status display showing you all of the frames which have already been imported. You can add raw data to this form by hitting the 'Browse To Add' button or by right clicking and choosing 'Add'. Note that the right click menu also has 'Remove' and 'Save As' options.
    - Select the **Filter Raw Data** subtask. Click on the 'Browse To Add' button. You should see the sample micrographs in the raw\_data directory. Select them all by first clicking on image 1160.mrc, holding shift and then clicking on 1792.mrc (or click in the upper left corner of the table). This should highlight all images. Then hit OK. The browser dialog should close and the selected images should appear in the table on the first form. Make sure that the 'Edge norm' and 'Thumbnails' and 'Associate with project' options are selected. If the particles in your images appear dark on a light background, select the 'invert' option. In addition, if the images are CCD frames you may wish to use the x-ray pixel filter as well. When options are selected, hit 'OK'.
  - ➔ Do **not** use the x-ray pixel filter option with negative stain or phase-plate data (if you are lucky enough to have a phase-plate). The filter will likely not work properly.
  - ➔ The Edge norm option will adjust the images so the mean value around the edge of the frame is zero and the standard deviation of the pixel values is 1. This normalization helps regularize the data for later processing and minimizes problems with brightness and contrast, though additional per-particle normalization may occur later. It is particularly critical for CCD data.
  - ➔ The invert option should be used as necessary to make your particles white on a dark background. Depending on acquisition method and whether your data is in ice or stain, you may or may not need this. For the workshop demo data, you do not need to invert.
  - ➔ The 'Associate with project' options adds the filtered images to the list of frames in the project - you can see list of the frames in the project by going back to the **Raw Data** task.
  - ➔ The 'Inplace processing' option writes the filtered images back to the input files, which saves on disk space, but has the disadvantages discussed above with simply adding images to the project without importing.
- 
- Go to the **Interactive Boxing – e2boxer** task and hit **OK**.
    - ➔ The displays on the workshop PC's are completely inadequate to this task. To box particles you really need to have a display with at LEAST 1600x1200 or 1080p resolution. Dual display setups are even better. Some laptops have adequate displays for this, but most do not.
  - You should see the e2boxer interface form appear which displays a table. The left-most column lists the image names of the micrographs in the project. The other column, most probably blank at this point, is titled Stored Boxes.
    - ➔ Note that you can double click on any of the image names in the left column and the associated image will appear in an interface for viewing (Try it).

- If you have a decent workstation with a reasonable amount of RAM, you could open all of the micrographs at once, but on a laptop, this probably isn't wise. Choose 2 or 3 images (or just 1 if you have problems with more, implying your machine is really not adequate for this tutorial), enter a boxsize of 256, then hit **OK**.
- Wait a moment while e2boxer loads.

## e2boxer – the GUI

- You should see at least 3 windows appear. You will need to arrange the windows yourself. On a laptop with a low-resolution display, this can be a challenge. The windows are:
  - The main controller - looks like a regular window; it has an assortment of buttons and text entry boxes
  - the main image display window - shows the 2D micrograph currently selected
  - the particle display window - (with nothing in it) that will eventually show the boxed particles.
  - micrograph thumbnails - This will only appear if you selected 2 or more micrographs in the previous step. Clicking on one of these will select the current image to be boxed.

## Getting started with interactive autoboxing

- The particle picker (e2boxer.py when started from the command line) has 3 modes of operation: 'Manual', 'Erase', and 'Swarm'. The program will start in Manual mode, in which particles are picked by hand with no option for automatic selection. Erase mode permits you to mark large regions of the image which should be excluded from automatic picking. Finally, Swarm mode is the interactive automatic particle picker. In Swarm mode you select a few particles manually, then the remainder of the image is automatically selected based on parameters determined from your manually selected images. The more particles you select manually, in general, the more accurate the automatic picking should become.
  - ➔ Note that particles picked in Manual mode have no impact on the Swarm picker at all, so if the automatic picker is doing a good job, but it missed a few particles you'd like to add, rather than adding them to the reference set in Swarm mode, you may elect to switch to Manual mode instead, to pick up the missing particles.
  - ➔ Also note that the Swarm picker works well on some projects, and less well on others. While we do intend to add other automatic picking options in future, at the moment Swarm is the only option in EMAN2. While you can pick particles with a non EMAN2 tool, then import the boxed particles, if you do, you will lose the information locating where each particle was on the original micrograph.
- You likely want to start with the **Swarm** picking mode. Select this mode, and start by manually selecting 2-3 particles. You should see a bunch of other boxes appear automatically. The more manual references you pick, the better the automatic selection should become. Some particles may cause the picker to become too liberal. In this case, generally picking a few more references manually will help. **The Particle Diameter** parameter is also quite important. This is not the same as box-size. If you are having troubles getting good Swarm results, try changing this parameter. You can also change the method to 'More Selective' to get fewer particles.
  - ➔ You can delete 'bad particles' by holding down shift and clicking on them. This includes particles that were automatically selected. If you delete one of the references you selected, it will update the autopicking results.
  - ➔ Bad particles or boxes that contain just noise can be damaging to your reconstruction, as they permit noise/model bias to become stronger. It is much better to miss a few good particles if it permits you to exclude obvious 'bad' particles.
  - ➔ One caveat to the above, if the good particles that get excluded are all in one orientation, that would be a bad thing
  - ➔ Note that this isn't your only opportunity to eliminate bad particles. Try not to be overly liberal, but realize that you will have another chance to clean up any false positives after CTF Wiener filtration.

- ➔ When picking, it can sometimes help for purposes of more accurately centering and identifying particles, to use a box size smaller than the final size you plan to use for processing. This is absolutely fine. You can use whatever box size you like during the picking process. When you get to 'generate output' later, you will be given a chance to select the final box size to be used for processing.

## Boxing multiple images interactively

- ➔ Note that this section applies only if you have more than one image loaded into the e2boxer interface.
- Choose the next image in the image thumbnails window. This should load the new image into the main display window, and if you used swarm, you should also observe that autoboxing has occurred, based on whatever autoboxing was done on the last selected image.

## Finishing e2boxer

- Once you are done boxing the images you may be tempted to press the 'generate output' button in the e2boxer control panel, but don't. When you have finished boxing each set of micrographs, simply click Done in the main controller and return to the e2workflow interface. We will deal with 'generate output' later.
- If you have run e2boxer correctly you should be able to click on the Interactive boxing task and observe that the Stored Boxes column now has entries in it. These entries should correspond to the total number of boxes currently stored in the database for the given image.
- Generate boxed particle output by going to the **Generate Output (e2boxer)** task in the **Particles** section of the workflow. You should see a form appear that lists the micrographs you have in the project as well as the number of boxes you have stored for each image in the database.
  - Select all of the images
  - enter a boxsize of 256.
  - make sure the normalize.edgemean option is chosen
  - make sure the output image format is "bdb"
  - Hit Ok. Once again this will spawn processes on your operating system that you can monitor in the Running tasks window.
- ➔ It is critical for accurate CTF correction that the box size be substantially larger than the particle, ideally almost 2x. That is, if a box that just barely contains your particle has a size of 128, you should use a final box size in the 192-256 range. This is larger than typically used in EMAN1, but there is a good reason for it, as you will see later. See: <http://blake.bcm.edu/emanwiki/EMAN2/BoxSize> for information on 'good' sizes to use to optimize processing speed. Since we are just practicing here, it doesn't matter much, but if this were your own data, picking an appropriate box size at this stage is critical !
- ➔ Assuming you chose the 'bdb' format your boxed particle output you'll note that the filenames that appear in this list aren't simply "1160\_ptcls", but much longer and complicated looking 'bdb:particles#1160\_ptcls'. The specifications for bdb database access are fairly straightforward:
  - 'bdb:dbname' which refers to the database (think of it as an image file) in the local directory named 'dbname'. 'dbname' can contain individual images, numbered sets of images, named images and other named metadata
  - 'bdb:/path/to#dbname' allows you to specify a database residing in a different directory. Note that '#' is used for the final separation between the path and the database name
  - files in the EMAN2DB directory should NEVER be renamed, individually deleted, moved, etc. It can cause the system to become 'confused' and produce a range of seemingly unrelated errors.

- If you really insist on deleting files in the BDB directory, make sure you aren't running any EMAN2 programs, run 'e2bdb.py -c' and only then erase the files. Note that the easiest way to remove files from an EMAN2 database is using the EMAN2 browser.
- For more info on the database please see: <http://blake.bcm.edu/emanwiki/Eman2DataStorage>
- You should now proceed to CTF correction as we did earlier in the workshop and autofit the data you've selected. Evaluate the 1-D and 2-D power spectra you get for the particles from each image. Are some better than others ?