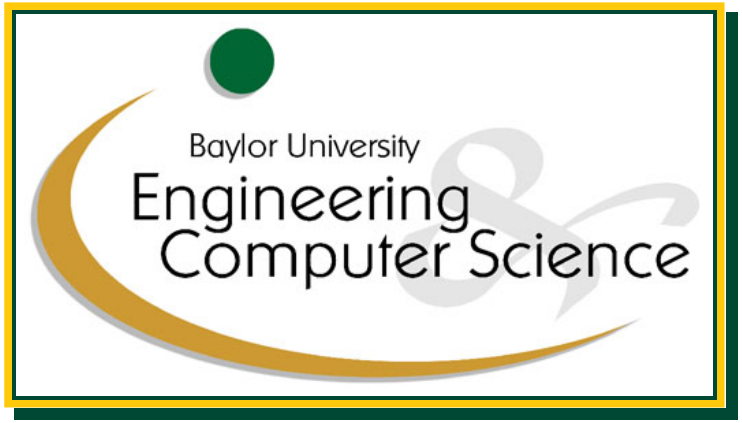
pCT Data Storage at Baylor

Folder/File Naming and Organizational Organizational Organizational Scheme

Blake Edward Schultze

February 13, 2016



Description of User/Shared Directories on Kodiak/Tardis Cluster Nodes

/home/<*user name*> : This is the user's private home directory where the user account's settings/history files (`.bash_profile`, `.bash_history`, etc.) are stored. Each user only has access to their directory but they can access it from each of the Kodiak/Tardis nodes. These directories have a limited storage capacity and their contents are not backed up, meaning its contents will not be recoverable in the event of failure. Hence, these should not be used to store user data/code, particularly if it is large and/or important.

/data/<*user name*> : This is the global home directory of each user where the input/output data for code/program execution and all permanent/important personal data should be stored. As a subdirectory of **/data**, the contents of these directories are backed up periodically, but like the local home directories, each user only has access to their personal directory.

/ion/... : This directory is dedicated to proton and ion therapy research and this is where users can upload data/code and selectively share portions of it with the rest of the collaboration by moving the desired files to the staging area (**/staging/<*user name*>**)and an administrator will then move them to the appropriate shared data/code directory. The data in this directory is located on a network storage device and this drive is mounted on both the Kodiak and Tardis clusters, making it accessible from all of the master/compute nodes. To prevent a permanent loss of data in the event of drive failure, the contents of this directory are periodically backed up to tape drive.

/ion/incoming/<*user name*> : This is the directory where users should upload data/code to the Baylor server prior to preparing and organizing it for sharing with the collaboration, with each user only having access to their personal directory in **/incoming**. When a user is ready to share data/code, it should be moved to their directory in the staging area and an administrator will move it to the appropriate shared directory.

/ion/staging/<*user name*> : This is the directory where users move data/code they wish to share with the collaboration, at which point an administrator will move the data from the staging area to the appropriate shared directory. Since the administrators are not familiar with pCT or its organizational scheme, it is important for users to organize the data/code prior to adding it to the staging area so an administrator can easily determine and move the data to the appropriate shared directory. In particular, users should create the entire target hierarchy of shared subdirectories below **/ion** inside the staging area. The administrator can then follow these hierarchies downward until encountering a subdirectory in the staging area hierarchy that does not currently exist in the target shared directory, at which point they can simply move this subdirectory and its contents from the staging area to the target shared directory. The subdirectories of the target shared directory that already exist must still be created and included in the staging area hierarchy so the directory hierarchies in the staging area and target shared directory are identical below **/ion** and **/ion/incoming/<*user name*>** , respectively, providing the administrator with the target data path and required organizational framework and prevents the need to create/name/organize subdirectories.

/ion/pCT_data/... : This directory is where the raw, preprocessed, projection, and reconstruction data/images are moved to make them available to the other pCT users. Each type of data is stored in separate subdirectories and soft links to this data are created and organized in a directory hierarchy

indicating their input/output data dependencies. The directory/file naming and organizational scheme for each type of data and the soft links are outlined in the next section of this document. Data/images should only be moved to this shared directory after having been verified as valid/accurate and having been organized appropriately.

/ion/pCT_code/... : This directory contains the source code of the programs involved in the pCT reconstruction process, from data acquisition to image reconstruction and analysis of reconstructed images. Each of the programs is maintained as a GitHub repository to provide version control and simplify collaborative program development efforts. Although these may have originated as a personal repository, these are currently being added to a collaboration wide GitHub account so users can locate and access all available pCT programs from a central location. The group of developers of each program will inevitably change as people join the collaboration and others move on to other projects, so transferring repository ownership from individual users to a persistent collaboration account not only provides more efficient access and collaboration, but permanent ownership and centralized storage will provide seamless transitions as program developers either join or leave the collaboration.

Each program repository has a master branch, which corresponds to the current release version, and each of the program's developers will typically have their own branch which they can use to develop and test new ideas/features. Once the implementations of such development mature and have been thoroughly tested, they can propose that these be merged into the master so they are included in subsequent releases. The group of developers of a program should decide amongst themselves what the process will be for approving merges with the master branch and when to release a new version of the program, which may include the results of several separate merges.

Although users can access each program's source code from GitHub, each repository and its constituent branches are cloned to **/ion/pCT_code/...** and establishing a link between the local and remote (GitHub) repositories provides a mechanism to check for and automatically apply available updates by entering `git pull --rebase`. Hence, users can ensure they always use the most recent version of a program by performing this update check prior to using the release version of a program. **Note that this should only be done when using the release version of a program; if a developer branch is "updated" to its most recent version, any additions/changes they have made since their last commit will be overwritten**. Developers will usually commit changes only after achieving some relatively significant goal so reverting to the previous commit can result in a substantial loss of work. If a developer wants to make their intermediate developments available for use, they should consider creating another branch they can use as a release version of their code and restrict access to the branch exclusively used for development.

/ion/pCT_data/pCT_Documentation/... : Documentation relevant to pCT is stored in this directory, such as descriptions of the data format, coordinate system, and phantoms and pCT related publications (including student theses/dissertations). This is a GitHub managed local repository allowing everyone to "push" contributions to the repository and "pull" updates/additions from others into their own local clone ensuring everyone has access to the latest information.

KEY:

Dark Green : directories whose names do not change

Brown : directories whose names vary depending on object name or date

***Italic/Royal Blue* : individual files**

***Italic/Dark Blue* : multiple files**

/data/ion/pCT_data/organized_data : directory containing all of the pCT data (raw, processed, projection, and reconstruction), primarily as soft links to the actual data itself (with each type of data stored/organized in separate directories), organized in a hierarchy of subdirectories indicating data dependencies.

(1) /**Phantom** : directory containing all of the experimental/simulated data and reconstructed images associated with the phantom/object named "*Phantom*".

- **(a)** /**Reference_Images** : directory containing reference images (xCT, RSP, etc) relevant to analysis/comparison of the data/images for this object and data type.
- **(b)** /**Experimental** : directory containing data and images generated from an experimental scan of the object.
  - **(i)** /**YY-MM-DD** : directory containing data and reconstructed images corresponding to all experimental scans of the object performed on this date.
    - **(1)** /**XXXX[_AAA]** : directory containing data and reconstructed images corresponding to the experimental scan of the object for this particular run # of this date, where the run # is of the form "*XXXX*" with an optional descriptor tag "*_AAA*" added specifying additional pertinent information about the scan, such as inferior "*_Inf*" or superior "*_Sup*" positioning as in the case of a head phantom.
      - **(a)** /**Input** : directory containing raw data generated by scan of object from each gantry angle and transmitted by event builder.
        - **(i)** ***raw_xxx.bin*** : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", $\cdots$}.
      - **(b)** /**Output** : directory containing calibration and post processed data generated from analysis of raw data and used as input to image reconstruction.
        - **(i)** /**YY-MM-DD** : directory containing the post processed "*projection_xxx.bin*" data generated on this date and the reconstructions using this data.
          - **(1)** ***calib.txt*** : text file containing calibration curve coefficients for WEPL calibration.

(2) ***TVcalib.txt*** : text file containing calibration curve coefficients for tv corrected WEPL calibration.

(3) ***settings.cfg*** : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and initial settings to use for preprocessing/reconstruction.

(4) ***projection_xxx.bin*** : binary files containing tracker coordinates and WEPL data associated with gantry position "*xxx*" = {"*001*","*002*","*003*",$\cdots$} converted from raw data using tracker alignment, track reconstruction, and WEPL calibration routines and used as input to image reconstruction.

(5) **/Reconstruction** : directory containing preprocessed data and reconstructed images generated using the "*projection_xxx.bin*" data along with reference images relevant to the object.

- **(a)** **/YY-MM-DD** : directory containing the preprocessed data generated on this date and the reconstructed images generated from this data.
  - **(i)** ***settings_log.cfg*** : copy of ***settings.cfg*** with any changes made to parameters/options applied at execution, if any.
  - **(ii)** ***execution_times.txt*** : execution times for various portions of preprocessing and/or reconstruction and total program execution time.
  - **(iii)** ***bin_counts.txt*** : linearized bin # for each proton history, where linearized bin # = t_bin + angle_bin * T_BINS + v_bin * T_BINS * ANGULAR_BINS.
  - **(iv)** ***mean_rel_ut_angle.txt*** : mean relative ut angle ($\angle ut_{out} - \angle ut_{in}$) by linearized bin #.
  - **(v)** ***mean_rel_uv_angle.txt*** : mean relative uv angle ($\angle uv_{out} - \angle uv_{in}$) by linearized bin #.
  - **(vi)** ***mean_WEPL.txt*** : mean WEPL value by linearized bin #.
  - **(vii)** ***stddev_rel_ut_angle.txt*** : standard deviation of the relative ut angle ($\angle ut_{out} - \angle ut_{in}$) by linearized bin #.
  - **(viii)** ***stddev_rel_uv_angle.txt*** : standard deviation of the relative uv angle ($\angle uv_{out} - \angle uv_{in}$) by linearized bin #.
  - **(ix)** ***stddev_WEPL.txt*** : standard deviation of the WEPL value by linearized bin #.
  - **(x)** ***sinogram.txt*** : mean WEPL after statistical cuts with the t_{bin} and angular bin θ_{bin} plane for each vertical bin v_{bin} stacked on each other.
  - **(xi)** ***hull.txt*** : text file specifying hull in 1s/0s with the xy plane for each slice stacked on each other.
  - **(xii)** ***FBP.txt*** : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.
  - **(xiii)** ***x_0.txt*** : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
  - **(xiv)** ***sin_table.bin*** : file containing the tabulated values of sine function

(xv) ***cos_table.bin*** : file containing the tabulated values of cosine function

(xvi) ***coefficient.bin*** : file containing the tabulated scattering coefficient values for Σ_1/Σ_2 for $u_2 - u_1/u_1$ values

(xvii) ***poly_1_2.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{1, 2, 3, 4, 5, 6\}$

(xviii) ***poly_2_3.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{2, 3, 4, 5, 6, 7\}$

(xix) ***poly_3_4.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{3, 4, 5, 6, 7, 8\}$

(xx) ***poly_2_6.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{2, 6, 12, 20, 30, 42\}$

(xxi) ***poly_3_12.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{3, 12, 30, 60, 105, 168\}$

(xxii) ***MLP.bin*** : binary file with MLP path data for each history entering hull.

(xxiii) ***WEPL.bin*** : binary file specifying WEPL value for each history entering hull.

(xxiv) ***histories.bin*** : binary file specifying entry/exit coordinates/angles, bin number, gantry angle, and hull entry x/y/z voxel # for each history entering hull.

(xxv) /**Images** : directory containing reconstructed images generated using this preprocessed data.

(1) /**YY-MM-DD** : directory containing the reconstructed images generated on this date using the preprocessed data above.

(a) ***x_k.dcm*** : DICOM images of x after k iterations.

(b) ***x_k.txt*** : text images of x after k iterations.

(c) ***x_k.png*** : PNG images of x after k iterations.

(c) /**Simulated** : directory containing data and images generated from a simulated scan of the object.

(i) /**G_YY-MM-DD** : directory containing data and reconstructed images corresponding to all GEANT4 simulated scans of the object generated on this date.

(1) /**XXXX[_AAA]** : directory containing data and reconstructed images corresponding to the experimental scan of the object for this particular run # of this date, where the run # is of the form "*XXXX*" with an optional descriptor tag "*_AAA*" added specifying additional pertinent information about the scan, such as inferior "*_Inf*" or superior "*_Sup*" positioning as in the case of a head phantom.

(a) /**Input** : directory containing raw data files generated by simulated scan of object for each gantry angle.

(i) ***raw_xxx.bin*** : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", $\cdots$}.

(b) /**Output** : directory containing calibration and post processed data generated from analysis of raw data and used as input to image reconstruction.

(i) /**YY-MM-DD** : directory containing the post processed "*projection_xxx.bin*" data generated on this date and the reconstructions using this data.

(1) ***calib.txt*** : text file containing calibration curve coefficients for WEPL calibration.

(2) ***TVcalib.txt*** : text file containing calibration curve coefficients for tv corrected WEPL calibration.

(3) ***projection_xxx.bin*** : binary files containing tracker coordinates and WEPL data associated with gantry position "*xxx*" = {"*001*","*002*","*003*",$\cdots$} converted from raw data using WEPL calibration routine and used as input to image reconstruction.

(4) /**Reconstruction** : directory containing preprocessed data and reconstructed images generated using the "*projection_xxx.bin*" data along with reference images relevant to the object.

(a) /**YY-MM-DD** : directory containing the preprocessed data generated on this date and the reconstructed images generated from this data.

(i) ***hull.txt*** : text file specifying hull in 1s/0s.

(ii) ***FBP.txt*** : text file specifying filtered back projection image.

(iii) ***x_0.txt*** : text file specifying voxel values of initial iterate.

(iv) ***MLP.bin*** : binary file with MLP path data for each history entering hull.

(v) ***WEPL.bin*** : binary file specifying WEPL value for each history entering hull.

(vi) ***histories.bin*** : binary file specifying entry/exit coordinates/angles, bin number, gantry angle, and hull entry x/y/z voxel # for each history entering hull.

(vii) /**Images** : directory containing reconstructed images generated using this preprocessed data.

(1) /**YY-MM-DD** : directory containing the reconstructed images generated on this date using the preprocessed data above.

(a) ***x_k.dcm*** : DICOM images of x after k iterations.

(b) ***x_k.txt*** : text images of x after k iterations.

(c) ***x_k.png*** : PNG images of x after k iterations.

(ii) /**T_YY-MM-DD** : directory containing data and reconstructed images corresponding to all TOPAS simulated scans of the object generated on this date.

(1) /**XXXX[_AAA]** : directory containing data and reconstructed images corresponding to the experimental scan of the object for this particular run # of this date, where the run # is of the form "*XXXX*" with an optional descriptor tag "*_AAA*" added specifying additional pertinent information about the scan, such as inferior "*_Inf*" or superior "*_Sup*" positioning as in the case of a head phantom.

(a) /**Input** : directory containing raw data files generated by simulated scan of object for each gantry angle.

(i) ***raw_xxx.bin*** : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position "*xxx*" = {"*001*","*002*","*003*",$\cdots$}.

(b) /**Output** : directory containing calibration and post processed data generated from analysis of raw data and used as input to image reconstruction.

(i) /**YY-MM-DD** : directory containing the post processed "*projection_xxx.bin*" data generated on this date and the reconstructions using this data.

(1) ***calib.txt*** : text file containing calibration curve coefficients for WEPL calibration.

(2) ***TVcalib.txt*** : text file containing calibration curve coefficients for tv corrected WEPL calibration.

(3) ***projection_xxx.bin*** : binary files containing tracker coordinates and WEPL data associated with gantry position "*xxx*" = {"*001*","*002*","*003*",$\cdots$} converted from raw data using WEPL calibration routine and used as input to image reconstruction.

(4) /**Reconstruction** : directory containing preprocessed data and reconstructed images generated using the "*projection_xxx.bin*" data along with reference images relevant to the object.

(a) /**YY-MM-DD** : directory containing the preprocessed data generated on this date and the reconstructed images generated from this data.

(i) ***hull.txt*** : text file specifying hull in 1s/0s.

(ii) ***FBP.txt*** : text file specifying filtered back projection image.

(iii) ***x_0.txt*** : text file specifying voxel values of initial iterate.

(iv) ***MLP.bin*** : binary file with MLP path data for each history entering hull.

(v) ***WEPL.bin*** : binary file specifying WEPL value for each history entering hull.

(vi) ***histories.bin*** : binary file specifying entry/exit coordinates/angles, bin number, gantry angle, and hull entry x/y/z voxel # for each history entering hull.

(vii) /**Images** : directory containing reconstructed images generated using this preprocessed data.

(1) /**YY-MM-DD** : directory containing the reconstructed images generated on this date using the preprocessed data above.

(a) ***x_k.dcm*** : DICOM images of x after k iterations.

(b) ***x_k.txt*** : text images of x after k iterations.

(c) ***x_k.png*** : PNG images of x after k iterations.

/data/ion/pCT_data/raw_data : directory containing the raw experimental data organized by scan date

(1) **/YY-MM-DD** : Folder containing all raw experimental data acquired from the scan beginning on "*YY-MM-DD*"

(a) ***<object>_XXXX[_AAA]_xxx.dat*** : raw experimental data for the object named "*<object>*", from run # "*XXXX[_AAA]*", where "*XXXX*" is a 4 digit # with leading zeros and "*_AAA*" is an optional descriptor tag, and "*xxx*" is the gantry angle at which the data was acquired.

/data/ion/pCT_data/preprocessed_data : directory containing the preprocessed experimental data organized by scan and processed dates

(1) **/YY-MM-DD** : Folder containing all processed experimental data corresponding to the raw experimental data acquired on "*YY-MM-DD*"

(a) **/YY-MM-DD** : Folder containing all processed experimental data generated on "*YY-MM-DD*" from the raw data

(i) ***<object>_XXXX[_AAA]_xxx.dat.root.reco.root.bin*** : processed experimental data with tracker coordinates, recovery of missing hits when possible, and calibrated WEPL measurements for the object named "*<object>*", from run # "*XXXX[_AAA]*", where "*XXXX*" is a 4 digit # with leading zeros and "*_AAA*" is an optional descriptor tag, and "*xxx*" is the gantry angle at which the data was acquired.

File List

1. ***<object>_XXXX[_AAA]_xxx.dat*** : raw experimental data for the object named "*<object>*", from run # "*XXXX[_AAA]*", where "*XXXX*" is a 4 digit # with leading zeros and "*_AAA*" is an optional descriptor tag, and "*xxx*" is the gantry angle at which the data was acquired.
2. ***<object>_XXXX[_AAA]_xxx.dat.root.reco.root.bin*** : processed experimental data with tracker coordinates, recovery of missing hits when possible, and calibrated WEPL measurements for the object named "*<object>*", from run # "*XXXX[_AAA]*", where "*XXXX*" is a 4 digit # with leading zeros and "*_AAA*" is an optional descriptor tag, and "*xxx*" is the gantry angle at which the data was acquired.
3. ***projection_xxx.bin*** : binary files containing tracker coordinates and WEPL data associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", $\cdots$} converted from raw data using WEPL calibration routine and used as input to image reconstruction.
4. ***raw_xxx.bin*** : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", $\cdots$}.
5. ***calib.txt*** : text file containing calibration curve coefficients for WEPL calibration.
6. ***TVcalib.txt*** : text file containing calibration curve coefficients for tv corrected WEPL calibration.

7. ***settings_log.cfg*** : copy of ***settings.cfg*** with any changes made to parameters/options applied at execution, if any.
8. ***execution_times.txt*** : Listing of the preprocessing/reconstruction options/parameters used for a particular reconstruction along with the execution times for various individual tasks and portions of preprocessing and/or reconstruction and total program execution time organized as a series of "key=value" entries in a text file in the directory with the corresponding results.
9. ***execution_times.csv*** : The same information listed in the ***execution_times.txt***files for each execution are also stored in this global comma separated value (.csv) file with the execution information organized into separate columns and each row corresponding to a different reconstruction, thereby maintaining a single file where the execution information for all previous reconstructions can be found in the same place. Upon execution of a new reconstruction, a new row entry is added for storage of the corresponding execution information.
10. ***sin_table.bin*** : file containing the tabulated values of sine function
11. ***cos_table.bin*** : file containing the tabulated values of cosine function
12. ***coefficient.bin*** : file containing the tabulated scattering coefficient values for Σ_1/Σ_2 for $u_2 - u_1/u_1$ values
13. ***poly_1_2.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{1, 2, 3, 4, 5, 6\}$
14. ***poly_2_3.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{2, 3, 4, 5, 6, 7\}$
15. ***poly_3_4.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{3, 4, 5, 6, 7, 8\}$
16. ***poly_2_6.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{2, 6, 12, 20, 30, 42\}$
17. ***poly_3_12.bin*** : file containing the tabulated MLP polynomial values with coefficients $\{3, 12, 30, 60, 105, 168\}$
18. ***bin_counts.txt*** : linearized bin # for each proton history, where linearized bin # = t_bin + angle_bin * T_BINS + v_bin * T_BINS * ANGULAR_BINS.
19. ***mean_rel_ut_angle.txt*** : mean relative ut angle ($\angle ut_{out} - \angle ut_{in}$) by linearized bin #.
20. ***mean_rel_uv_angle.txt*** : mean relative uv angle ($\angle uv_{out} - \angle uv_{in}$) by linearized bin #.
21. ***mean_WEPL.txt*** : mean WEPL value by linearized bin #.
22. ***stddev_rel_ut_angle.txt*** : standard deviation of the relative ut angle ($\angle ut_{out} - \angle ut_{in}$) by linearized bin #.
23. ***stddev_rel_uv_angle.txt*** : standard deviation of the relative uv angle ($\angle uv_{out} - \angle uv_{in}$) by linearized bin #.
24. ***stddev_WEPL.txt*** : standard deviation of the WEPL value by linearized bin #.
25. ***sinogram.txt*** : mean WEPL after statistical cuts with the t_{bin} and angular bin θ_{bin} plane for each vertical bin v_{bin} stacked on each other.

26. ***MSC_counts.txt*** : text file specifying Modified Space/Silhouette Carving generated image of integer values indicating the # of times each voxel was identified as lying outside the object with the xy plane of each slice stacked on each other.

27. ***MSC_hull.txt*** : text file specifying Modified Space/Silhouette Carving generated hull in 1s/0s with the xy plane of each slice stacked on each other.

28. ***hull.txt*** : text file specifying hull in 1s/0s with the xy plane of each slice stacked on each other.

29. ***hull_avg_filtered.txt*** : text file specifying hull after applying an average filter and applying a threshold to the result so that all voxels with value less than some threshold (typically 0.01-0.1) are set to 0, thereby restoring image to 1s/0s with the xy plane of each slice stacked on each other.

30. ***FBP.txt*** : text file specifying filtered back projection image with the xy plane of each slice stacked on each other.

31. ***FBP_hull.txt*** : text file specifying object hull generated from the filtered back projection image by setting all voxels with RSP less than a user definable threshold (typically 0.6) with the xy plane of each slice stacked on each other.

32. ***FBP_median_filtered.txt*** : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.

33. ***x_0.txt*** : text file specifying voxel values of initial iterate with the xy plane of each slice stacked on each other.

34. ***x_k.dcm*** : DICOM images of x after k iterations.

35. ***x_k.txt*** : text images of x after k iterations.

36. ***x_k.png*** : PNG images of x after k iterations.

37. ***MLP.bin*** : binary file with MLP path data for each history entering hull.

38. ***WEPL.bin*** : binary file specifying WEPL value for each history entering hull.

39. ***histories.bin*** : binary file specifying entry/exit coordinates/angles, bin number, gantry angle, and hull entry x/y/z voxel # for each history entering hull.