ADASS XIII
ASP Conference Series, Vol. XXX, 2004
F. Ochsenbein, M. Allen and D. Egret eds.

Chandra Level 3 Processing: Creating an Automated Source Catalog for Chandra Data

Roger Hain, Adam Dobrzycki, Ian Evans, Janet DePonte Evans, Peter Freeman, Kenny Glotfelty, Margarita Karovska, Michael McCollough, Francis Primini, Arnold Rots

Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138

Abstract. Efforts are underway at the Chandra X-ray Center to develop a new level of data processing capability which will result in the creation of a source catalog spanning the full set of Chandra observations. Level 3 processing will include detailed source properties derived from all available Chandra data. The resulting catalog will provide easy access to Chandra data for an expanded number of astronomers, particularly those less familiar with analysis in the X-ray regime. It will allow easy searching of the archive for specific sources or for statistical properties of different classes of targets. In addition, consistent sets of data products along with all relevant calibrations will be available for detailed analyses.

Work has begun on developing a preliminary data processing pipeline, combining existing processing tools and identifying functionality which needs to be developed. Some effort has been made to identify a quick source detection algorithm which can work in the presence of the low-count background and source photons typical of Chandra X-ray data, and also be robust enough to find multiple or extended sources. Eventual goals of Level 3 processing include refining source detection and properties by simultaneously fitting multiple observations, and cross-matching identified sources with other catalogs. In this paper we present the current design, challenges, and discuss the various analysis tradeoffs.

1. Goals

There are three main goals that drive the development of the Chandra Level 3 pipeline project. The first is the desire to create a catalog of all detected sources from all Chandra observations. This will be done by automating source detection to extract all sources in the field of view. It specifically envisions making use of all available data from combined observations to maximize the ability to detect and quantify all possible sources.

In addition to identifying sources and their locations, the second goal is to provide detailed source properties for all detected sources. This will mean archiving a uniform set of properties for all catalog objects. All relevant calibration data will also be made available in the catalog.

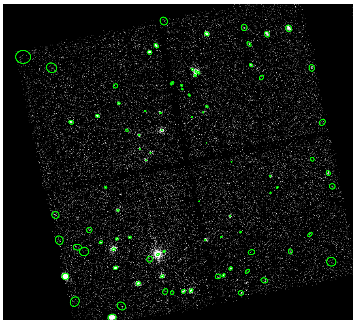

Figure 1. Sample source regions identified by the Level 3 pipeline

Having detected sources and cataloged detailed information, the third goal is to provide easy access to all Chandra data for a broad based group of astronomers. The catalog and interface will enable easy searches for sources and properties, and also enable easy searches for statistical properties over a class of sources. In addition to providing information for pure scientific research, the data should also aid in preparing future proposals or observations.

2. Prototype Pipeline

The prototype Level 3 pipeline is currently being developed. In its initial steps, it processes individual observation IDs (obsids) to automatically extract source data and populate the archive. Event data from an obsid is preprocessed by removing bad pixels and creating a simplified exposure map. This information is used by the source extraction software to identify source regions (Figure 1). Source detection is performed on the broadband image, and also on the hard, medium, and soft energy ranges independently. Successively larger regions centered at the mirror focal point are examined at higher blocking factors. These larger blocking factors are matched to the increase in size of Chandra's point spread function (PSF).

Sources from the event list examined at different blocking factors and energy levels are then merged to create a complete source list. Each source region is identified, along with an appropriate background region. Further analysis is performed on a per source basis, with parallel paths for each source through the pipeline. To assist in the forward fitting step, in which sources are identified as either a single point source, two point sources, or a more complex source, the Chandra PSF is calculated at multiple energy levels (Figure 2). Exposure maps of both the broadband background region and of the individual energy regions are also calculated.

The forward fitting step identifies the type of source (single, double, complex) and provides source property information. Flux, hardness ratio, and other quantities are calculated. Additional information, such as lightcurves and spectral information, are extracted. Data are then fed into the archive. The archive is examined for existing data on the source, and if found, portions of the Level 3 pipeline will be run again with merged data from multiple obsids. The archive will maintain the most complete information from all obsids that have been in-

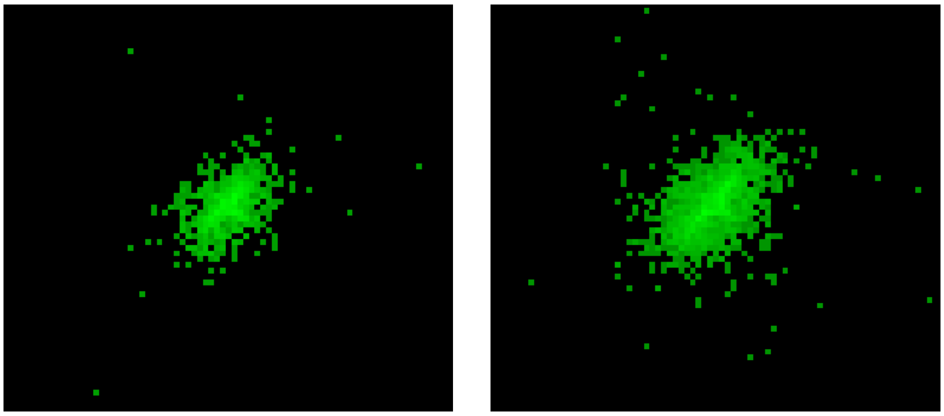

Figure 2. Sample source PSFs for 2 energy bands: soft (left) and hard

corporated into the Level 3 pipeline. It will include not just source property information, but also ancillary data, such as the regions from which source and background information was extracted, exposure map data, etc.

3. Challenges

This section highlights four of the most significant challenges that will need to be met in order to create the Level 3 pipeline.

Low Count Data. Chandra X-ray data typically consist of low-count source data when compared to background data, often making it difficult for source detection algorithms to distinguish sources, reject false sources, and separate two nearby sources. Since the Chandra PSF grows significantly away from the focal point of the mirrors, resolution between multiple sources far from the focal point becomes increasingly difficult. Shown on the left in Figure 3 are some test studies performed to determine the ability of one candidate source extraction routine, SExtractor (Bertin), to resolve sources in the presence of varying backgrounds and distances from the focal point. The graphs show that increasing distance from the focal point and decreasing source events increases resolving difficulty for a pair of sources 4 arcseconds apart.

Regions of Interest. Determining the source and background regions for cleanly separated sources is not difficult. But when multiple sources are near each other and source and background photons overlap, identifying regions of interest (ROIs) becomes harder. For example, at the right in Figure 3 two sources are near enough that the source and background photons (nested circular regions) overlap in a small area. Should the overlapping photons be assigned to both sources, divided equally, or ignored? Various characterizations of the source and background region are currently being evaluated in a forward fitting algorithm.

Processing Complexity. Efficiently handling the large number of per source files and supporting data created by the Level 3 pipeline is necessary to minimize processing times. The current pipeline produces a very large number of files. Focal plane images at several blocking factors and energies are produced.

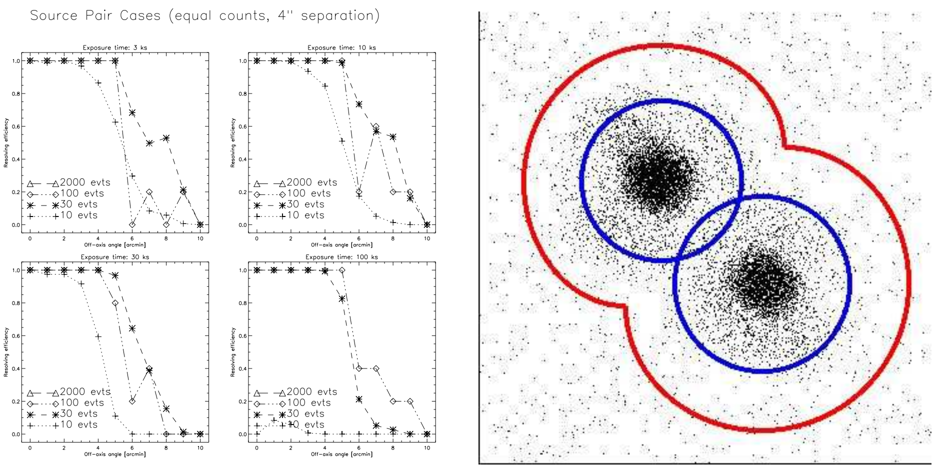

Figure 3. Challenges: Source detection results (left), overlapping ROIs

Per source exposure maps and PSF files are created at each energy band. Spectral and temporal information, per source viewable image files and additional smoothed images are created. Then, if prior Level 3 processing has already created archive entries for a source, all processing is repeated combining data from current and prior observations of this source for a new archive entry.

Archive. Challenges for the Level 3 archive come from the need to provide uniform data for all sources, despite the different character of the sources and the different needs of potential users. Issues also arise from combining multiple observations of the same source. Some archived quantities are not easy to merge into a combined catalog entry. Spectral information, for example, may vary among different observations of the same source. Which method is best to represent these types of data? How should factors such as exposure length vs. off-axis angle be weighed against each other?

4. Conclusion

Work is underway on the Chandra Level 3 pipeline to produce a source catalog with detailed information for all Chandra sources. An initial prototype pipeline is being developed, highlighting numerous challenges that need to be addressed. Refinements will be made to the pipeline as research into these issues matures.

Acknowledgments. This project is supported by the Chandra X-ray Center under NASA contract NAS8-39073.

References

Bertin, E. SExtractor User's Guide, http://terapix.iap.fr/cplt/oldSite/soft/sextractor