



CHANDRA
SOURCE CATALOG

Progress Report

Ian Evans (CXCDs)

On behalf of the Chandra Source Catalog Project Team

Chandra Users' Committee Meeting

April 9, 2008

Summary

- Significant progress has been made since the last CUC meeting
 - About a dozen requirements/specs or updates have been delivered by the science team
 - The software team has completed the CAT 2.6 (Operational Testbed 2) and CAT 2.7 (Operational Testbed 3) releases
 - *Approximately 550 observations have been processed through the CAT 2.6 and/or CAT 2.7 systems, yielding ~25,000 sources that pass catalog quality assurance and flux significance thresholds*
 - These data are being studied extensively by the science team to identify areas where additional tweaks are needed
 - Development work for the CAT 2.8 (Production Prototype) release is well underway

Currently estimating start of production in August 2008, with first public data access in September, and formal release 1 in January 2009

- First public access includes ~1/3 of public imaging observations processed and available, and preliminary statistical characterization of catalog properties
- Catalog release includes public mission-to-date imaging observations and complete statistical characterization of catalog properties available
- This schedule reflects an approximately 3 month slip since September 2007
 - Roughly half of the schedule slip was internal, due to issues that needed to be resolved with science algorithms and results from operational tests
 - The remainder was external, because of higher priority tasks taking precedence (identified at the September CUC meeting as an increasing schedule risk)

Progress: Science Highlights

Science Highlights Since Last CUC Meeting

- Catalog science included in Cycle 10 CfP for archival proposals
- Public web site released in January 2008 to guide proposers planning to submit catalog-based archival proposals
 - <http://cxc.cfa.harvard.edu/csc/>
 - Announced via CfP, Chandra bulletin, and newsletter
 - Will be updated with detailed user documentation in time for first public data access
- Extended science and system tests with CAT 2.6 and CAT 2.7 releases have demonstrated that the system is scientifically accurate, reliable, and robust
 - Most science effort is directed at evaluating the results from the operational test system runs
- All planned science requirements and specifications required for start of catalog production are complete
 - A small number of new requirements or revisions to address issues identified from operational testing are currently in work



[Catalog Home](#)

[Project News](#)

[Documents](#)

[Schedule](#)

[Publications](#)

Introduction

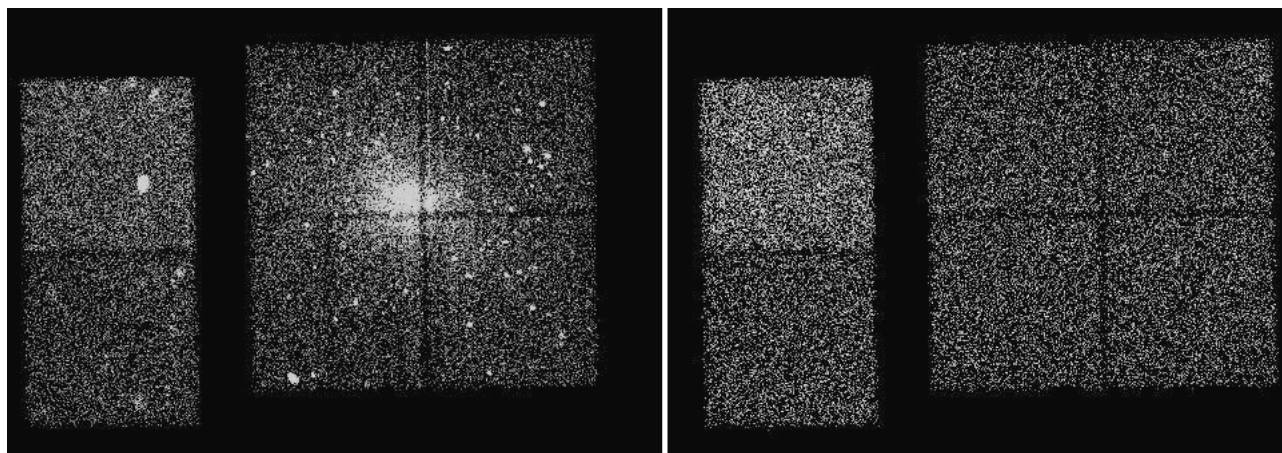
The Chandra Source Catalog will be the definitive catalog of all X-ray sources detected by the Chandra X-ray Observatory. It will provide simple access to Chandra data for individual sources or sets of sources matching user-specified search criteria. The catalog is designed to satisfy the needs of a broad-based group of scientists, including those who may be less familiar with astronomical data analysis in the X-ray regime. For each X-ray source, the catalog will list the source position and a detailed set of source properties, including commonly used quantities such as multi-band aperture fluxes, hardness ratios, temporal variability information, and source extent estimates. In addition to these traditional elements, the catalog will include file-based data products that can be manipulated interactively by the user, including images, photon event lists, light curves, and spectra for each source individually from each observation in which a source is detected.

Documents Delivered Since Last CUC Meeting

- Requirements and specifications released or updated
 - “*Chandra Source Catalog Requirements, Version 0.7*”, I. Evans, CXCDS
 - Refined performance requirements and resolved remaining associated TBDs
 - Transition from two-sided errors to two-sided confidence intervals for photometric and spectrometric quantities
 - Added flags to identify saturated or streak sources, and sources that required manual intervention in the quality assurance steps
 - A number of minor changes/clarifications
 - “Computation of Hardness Ratios Using BEHR”, I. Evans, CXCDS
 - “Measuring Detected Source Extent Using Mexican-Hat Optimization”, J. Houck, SDS/MIT
 - “Estimating Intrinsic Source Size”, J. Houck, SDS/MIT
 - “Adding Dither Corrections to L3 Lightcurves”, M. Nowak, SDS/MIT
 - “Removing Background Flares”, M. Nowak, SDS/MIT
 - “Operational Plans for L3 Characterization Simulations”, F. Primini, SDS
 - “Identification of Observations Containing Large Extended Sources”, A. Rots, CXCDS
 - “L3 Inter-Obi Variability Measure”, A. Rots, CXCDS
- Draft requirements pending release
 - “*Chandra Source Catalog Requirements, Version 0.8*”, I. Evans, CXCDS
 - Updated aperture photometry requirements to include some necessary quantities that are already being computed, and to improve the descriptions of other quantities
 - Added per-observation hardness ratios

Catalog Characterization

- Developed operational plan for catalog statistical characterization
 - Priority is to characterize source position, aperture photometry, limiting sensitivity, completeness, and false source rate
 - Second priority is source extent, spectral, and temporal properties
 - First characterization of high priority items available at time of first public data access
 - Detailed characterization of all items available with catalog release 1
 - Operational and manpower resources identified
- Simulation pipeline currently being developed
 - Inject simulated sources into real event lists
 - Simulate blank sky event lists using blank sky datasets from CalDB
 - Expect to be operational concurrent with CAT 2.8 release



- *Left:* M31 (ObsId 7064), ACIS ‘b’ band, block 8 *Right:* Simulated blank sky data with same exposure

Progress: Software Highlights

Software Highlights Since Last CUC Meeting

- CAT 2.6 (Operational Testbed 2) build released
 - System matches Requirements Document version 0.6, with waivers for incomplete items
 - Migration to Solaris 10
 - *Calibrate*, *Detect*, and *Source* pipeline enhancements
 - Improved background map
 - Addition of limiting sensitivity maps
 - Source properties: improved spectral fitting, additional temporal variability data, two-sided errors
 - Integration of quality assurance into pipelines
 - *Master* pipeline enhancements
 - Master source properties are populated
 - Reprocessing support
 - Automated Processing infrastructure
 - Enhancement to run system in “bulk reprocessing” mode
 - Efficiency improvements
 - Archive
 - Support for quality assurance; versioning support for reprocessing
 - New archive hardware (Solaris 10) set up and configured for CAT releases
 - Performance optimization based on CAT 2.5 test results

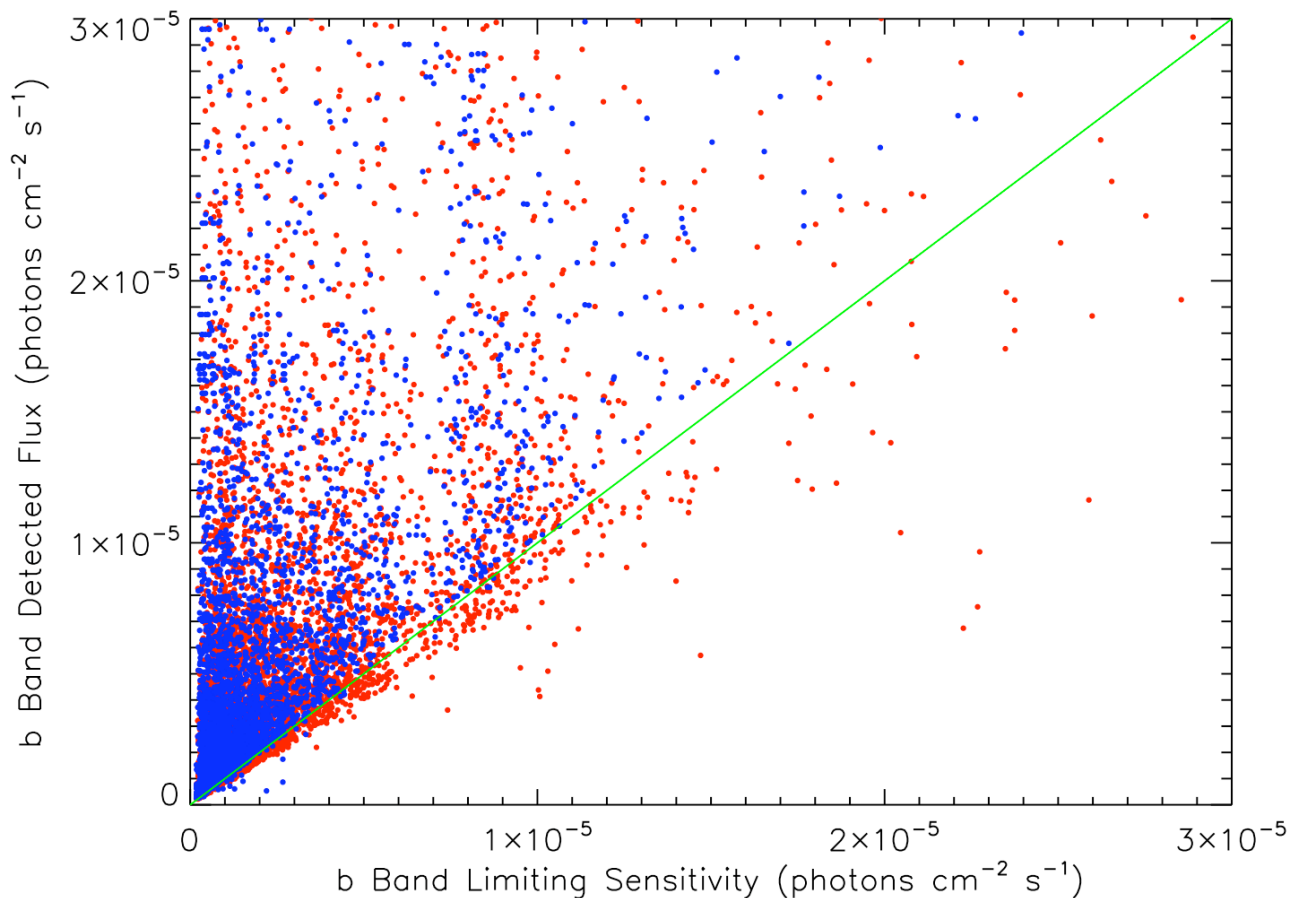
Progress: Software Highlights (cont.)

Software Highlights Since Last CUC Meeting (continued)

- CAT 2.7 (Operational Testbed 3) build released
 - System matches Requirements Document version 0.7, with waivers for incomplete items
 - Migration to Sybase 15
 - *Calibrate*, *Detect*, and *Source* pipeline enhancements
 - Improved background flare removal
 - Source properties: improved spectral fitting, wavelet source extent, intrinsic source size, two-sided confidence limits, correction for dithering in temporal variability measures
 - Integration of manual quality assurance feedback into pipelines
 - *Master* pipeline enhancements
 - Improved source matching for new barely overlapping sources
 - Two sided-confidence limits
 - Integration of quality assurance into pipeline
 - Automated Processing infrastructure
 - Enhancement to run system in “SAP” mode
 - Support for manual quality assurance threads
 - Archive
 - Test database and archive servers, and all catalog clients upgraded to Sybase 15
 - Support for SAP and manual quality assurance threads

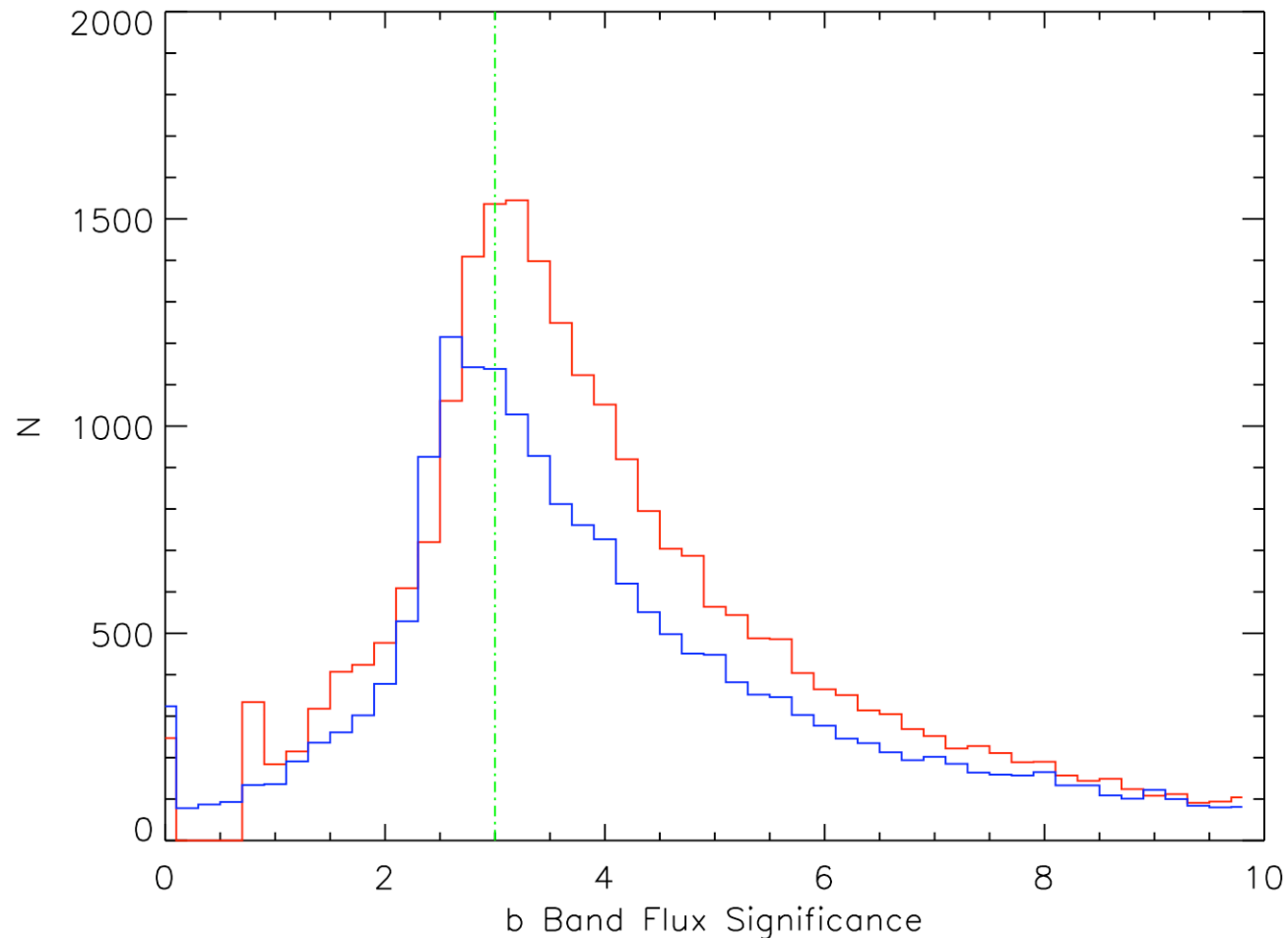
Progress

- CAT 2.6 operational test used for science and performance testing
 - CAT 2.6 processed 500 random ACIS observations in 13.1 days wall-time, generating ~31,000 raw detections and ~22,000 sources that passed quality assurance and catalog flux significance criteria (flux estimates $> 3\times$ the 1σ uncertainties)
 - $< 1\%$ error rate for observations, $\ll 1\%$ error rate for individual sources
 - Run characterized by science team identified issues corrected in CAT 2.7
 - Times are based on test hardware; production hardware expected to be $\sim 4\times$ faster
 - Evaluation of the results was used to verify algorithms and implementations, and establish expected performance
 - Rapid iterations based on an evaluation of an initial run of a “baseline” set of 32 well-characterized observations resolved several issues and allowed fine tuning of key parameters prior to the two-week run
- CAT 2.7 running and providing feedback to science and software teams
 - CAT 2.7 processed 243 random ACIS observations in 7.3 days wall-time
 - Processed approximately 50 additional observations, including the baseline set of 32 observations and 17 designed to extensively exercise source matching
 - Evaluation of the results is currently in progress

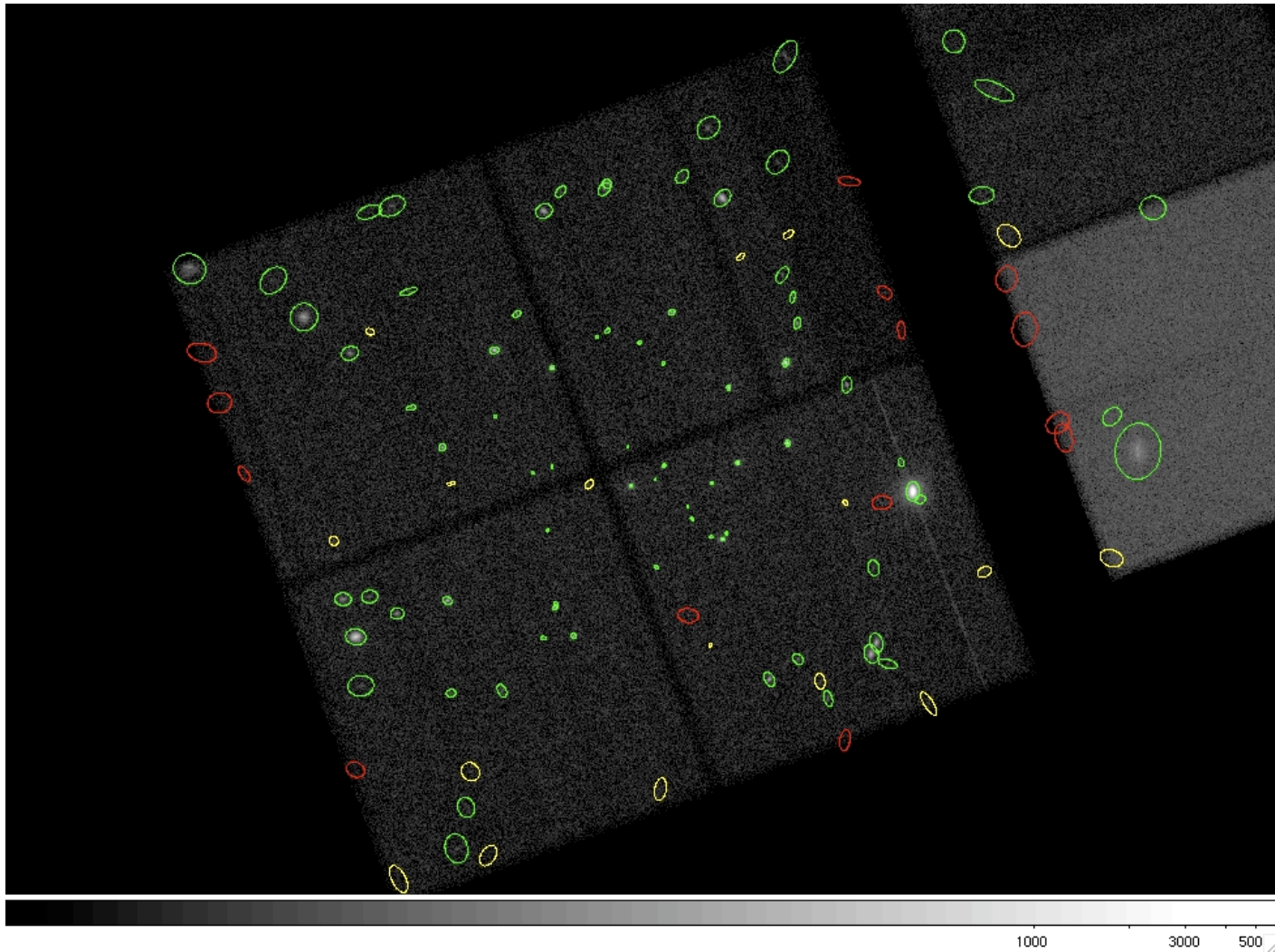


- ACIS broad band fluxes for good quality, uncrowded sources plotted vs. limiting sensitivity at the locations at which the sources were detected

CAT 2.6 data are shown in red; CAT 2.7 data are shown in blue. If the limiting sensitivity maps are correct, all sources would fall at or above the green line. For CAT 2.6, ~10% of sources fall below the line, with a bound on the error in the limiting sensitivity of ~20%. Background map corrections applied in CAT 2.7 significantly reduce the error in the limiting sensitivity: <1% of the sources fall below the green line using the CAT 2.7 algorithms.



- Histogram of number of detected sources versus ACIS broad band flux significance**
 CAT 2.6 data are shown in red; CAT 2.7 data are shown in blue. The normalizations of the histograms differ because the number of observations processed in the CAT 2.6 and CAT 2.7 test runs differ. The 3.0σ catalog flux significance threshold is shown in green. There is a slight deficit of sources in the $3.0\text{--}3.5 \sigma$ range in the CAT 2.6 data due to the parameter settings used in the source detection step. This has been addressed in the CAT 2.7 run.



- Catalog quality assurance criteria very effectively exclude false sources
 In this example (ρ Oph A; ObsId 637), only sources in green will be included in the catalog; source regions in red violate CAT 2.6 quality assurance criteria and source regions in yellow violate the catalog flux significance criterion.

Progress: Operational Testing (cont.)

Issues or Limitations Identified from CAT 2.6/2.7 Operational Tests

- Will be addressed in CAT 2.8 except where otherwise noted
 - False source detections may not be eliminated by quality assurance in some cases
 - Streak maps for ACIS subarrays with 128 rows are often not effective because of insufficient source-free rows; these will be excluded from release 1
 - Highly piled-up sources are detected as a group of sources on the edge of the crater
 - ACIS BI-FI chip boundaries can generate false sources if the background is bright
 - ACIS chip S4 can generate false sources because of background non-uniformity after destreaking is applied
 - Inspection of 493 observations from the CAT 2.6 operational test indicates that ~1% clearly include visually questionable detections after applying filters
 - » The contribution of these detections to the total false source rate is ~ 0.2%
 - » The fraction of false source detections from observations that do not clearly include visually questionable detections has not yet been established
 - Master source properties may not be computed correctly if the corresponding per-observation source properties are indefinite
 - Source position error ellipse is approximated by a circular error at the per-observation level
 - Source extent estimates may fail in crowded regions; deconvolved source extent will be approximated by a RSS estimator for release 1
 - Upper limits will not be considered when computing inter-observation properties if a source is not detected in an observation for release 1
 - Quality assurance parameters are not optimized

Planned Software Releases

- DS 7.6.11.8 (DS Archive Upgrade; May 2008)
 - Upgrade DS operational system to Sybase 15
 - Merge archive servers and migrate to new hardware running Solaris 10 and Sybase 15
- CAT 2.8 (Production Prototype; May 2008)
 - System matches Requirements Document version 0.8
 - Pipelines complete
 - Archive support for master quality assurance
 - Integration of manual merge review thread (GUI, integration with pipelines/AP/archive)
 - Catalog User GUI phase I
 - Extended pre-production test on operational hardware with operational archive configuration
- CAT 3.0 (Production Release; August 2008)
 - System matches Requirements Document version 1.0
 - Operations interfaces complete
 - Miscellaneous cleanups
 - Catalog User GUI phase II
 - Tune system
 - Catalog production

Chandra Source Catalog: Query Builder

Save Query Load Query Clear Query Submit Query

Property Selection:

Master Source Properties:

- dec
- err_ellipse_ang
- err_ellipse_r0
- err_ellipse_r1
- extent_flag
- flux_aper90_b
- flux_aper90_h
- flux_aper90_hilim_b
- flux_aper90_hilim_h
- flux_aper90_hilim_m
- flux_aper90_hilim_s
- flux_aper90_hilim_u

Observation Source Properties:

- instrument
- kp_prob_b
- kp_prob_h
- kp_prob_m
- kp_prob_s
- kp_prob_u
- kp_prob_w
- ks_prob_b
- ks_prob_h
- ks_prob_m
- ks_prob_s
- ks_prob_u

Search Criteria:

Property Search:

- o.theta < 15
- o.theta > 12
- o.edge_flag = 0
- o.flux_significance_b > 10
- o.pileup_warning < 0.2
- o.instrument = ACIS

Position Search:

Result Set:

Remove Column Remove All Columns Move Column Up Move Column Down

Sort By: m.name ascend


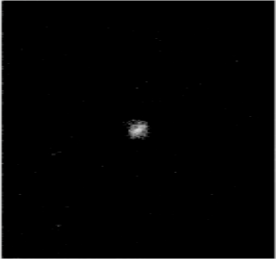
Max Rows Displayed: 10

Chandra Source Catalog: Source Viewer

Return To Query Return To Properties

name	ra	dec	theta	pileup_warning	flux_significance_b	flux_aper90_b	flux_aper90_hilim_b	flux_aper90_lo
CXO J002332.4-011445	5.68500309	-1.24596953	12.137410...	0.008353773...	11.4114	8.55084E-13	9.30883E-13	7.80042E-13
CXO J122349.6+072658	185.95697...	7.44958639	12.540067...	0.002861974...	34.5778	3.88339E-13	4.00104E-13	3.76692E-13
CXO J220107.3-315411	330.28076...	-31.903270...	12.775592...	3.926315327...	10.759	3.06864E-14	3.48392E-14	2.64916E-14

tricolor regimg (broad) regimg (hard) regimg (medium) regimg (soft) regimg (ultrasoft) fow psf (broad)

tricolor regimg (broad)

Schedule Risk Changes Since Last CUC Meeting

- Processing and/or archive hardware performance is inadequate (Risk ↓↓)
 - New Beowulf hardware delivered week of Mar 31 has sufficient margin based on operational and pre-production tests
 - Archive hardware upgrades (in-house, to be installed) are expected to have sufficient margin based on CAT 2.6 and CAT 2.7 test results and scaling to larger database table sizes (for many thousand observations)
- Science algorithm development takes longer than planned (Risk ↓)
 - Development of key algorithms is now completed
 - Updates based on CAT 2.6 and 2.7 tests are not expected to delay the schedule
 - Certain science capabilities will be deferred until after the start of processing
- Science algorithms or software are inadequate (Risk ↓)
 - Each software release is followed by a test and evaluation period to identify issues early
 - CAT 2.6 and 2.7 tests give high confidence that algorithms and software are adequate; some minor tweaks and corrections remain to be completed in CAT 2.8
- Software implementation takes longer than planned (Risk ↓)
 - Specifications exist for all planned tasks, and these tasks are either implemented or good time estimates exist for the remaining effort
 - New tasks or revisions based on feedback from CAT 2.6 and 2.7 tests are not expected to require significant effort (but note that reviews are not yet completed)

Schedule Risks (continued)

- Establishing catalog operations takes longer than planned (Risk \leftrightarrow)
 - Working with operations group to establish procedures and operations requirements for observation pre-filtering and batch creation
 - Planning to document system and train the operations group using pre-production test to refine procedures
 - Software and science teams will support operations during transition phase
- Other tasks compete for resources (Risk \leftrightarrow)
 - Same science and software resources support live *Chandra* mission
 - Since the last CUC meeting, there have been 5 DS operational system releases plus 1 pending DS operational system release, as well as 2 CIAO releases to support the Cycle 10 CfP
 - » 4 releases in response in operational changes, 1 to support cycle 10 CfP
 - Significant changes required to archive/database software and infrastructure
 - New hardware required archive and database software to migrate to Solaris 10
 - Sybase discontinued support for Sybase 12, hardware required archive and database software to migrate to Sybase 15
 - Planned transition to minimize schedule impact, but significant impact nevertheless

Schedule

- Schedule risks have decreased or remained the same since the last CUC meeting, as tasks are completed
- Most concerned about schedule risk associated with competition for available resources
 - Constantly working to minimize impacts by reassigning resources and shuffling schedules
 - Reshuffled CAT 2.7 release contents and added CAT 2.8 release to avoid delaying large scale operational tests
- Continue to estimate that all catalog schedule components will complete at about the same time

THERE ARE STILL NO LONG POLES

Bottom line best estimate is start of production in August 2008, with first public data access in September, and formal release 1 in January 2009