CXC Science Data Systems (SDS) Priorities, Scope, Status & Plans

Martin Elvis
SDS Group Lead

SDS Priorities

- Prime Directive:

  To enable the best science to be extracted from Chandra data
- These Goals follow:
  - ***Support core software for Chandra users***
    - I.e. support updated calibration and changes in instrument performance
    - Support reprocessing with latest calibration/algorithms
  - ***Support Chandra-unique capabilities***
    - Full exploitation of Chandra's spatial & spectral resolution
  - ***Support full spectrum of users***, newbies to sophisticates
    - Bring advanced analysis methods to all users via systematic review of capabilities, followed by incremental improvement with advanced tools, scripts, visualization, guides
  - ***Support multi-observatory analysis*** to correlate Chandra results with multi-wavelength data
    - Coordinate with other large astrophysics data analysis efforts
    - Use Data Model within VO to extend CXC multi-mission capability

SDS Responsibilities in support of Chandra Goals

- Pipelines. *Recent issues: co-ordinates, CC-mode times, gain(t), gapmap*
- Reprocessing. *Repro 3 coming up in Spring 2004*
- Instruments. *changes with time, calibration: ACIS, HRC, LETGS, HETGS*
- Level 3 processing. *source catalog, properties, limits…*
- V&V. *new system -release early 2004*
- Documentation, Web site
- Chandra/CIAO workshops, X-ray schools (with GSFC, COSPAR), Astro-statistics workshops
- Web services. *ChaRT, GUIDE*
- Helpdesk backstop
- CIAO *development, testing, maintenance, upgrades*

SDS Resource Distribution

- SDS is *16 FTE scientists (7 @ MIT), 4 FTE data aides*
  - Pipelines + Reprocessing + Instruments
  - Level 3 processing
  - V&V
  - Documentation, Web site
  - Workshops, X-ray astronomy schools
  - Web services
  - Spectroscopy database*
  - Helpdesk backstop
  - CIAO development, testing *~7 FTE scientists (4 @ MIT). I.e. ~1/2*
  - *Research*: *30%- 40% of all scientist FTE*
- Outlook to FY05: *13-14 FTE Scientists, 4 FTE data aides*
  - Transferring V&V to Data Systems Division Operations group

* On Jan 15 N. Brickhouse takes up Associate Directorship of SSP Division at CfA

Current CIAO release: 3.0

- Principle: *CIAO must serve full spectrum of users*
- Key CIAO 3.0 feature: extensibility
  - Also tool upgrades (`dmextract, dmgroup/dmgroupreg`), bug fixes
- Obtained by opening CIAO internals to a macro language
    - Note: not available readily in XSPEC/FTOOLS
- Gain for advanced users:
  - automation of your own complex tasks
  - Fairly easy importing of C, Fortran code as ‘modules’
- Gain for regular users:
  - More rapid response from SDS than CIAO release schedule via ‘scripts’ page containing useful macros
- Demos of this capability will follow this presentation
- Bottom line: *Extract Better science from Chandra*

Beyond CIAO 3

- Principles: *Support core software for Chandra users, Chandra-unique capabilities, Support full spectrum of users, support multi-observatory analysis*
- **Reprocessing 3:** updated calibration, algorithms for L2 products
- **Level 3** pipelines: Chandra source catalog(s)
- **CIAO 3.1**: "ACIS bakeout" support
- **CIAO 4**: graphics, speed, new functionality through scripts
- **CIAO 5**: ultimate Chandra performance, VO ready

CXC 2-Year Software Plan: FY04, FY05

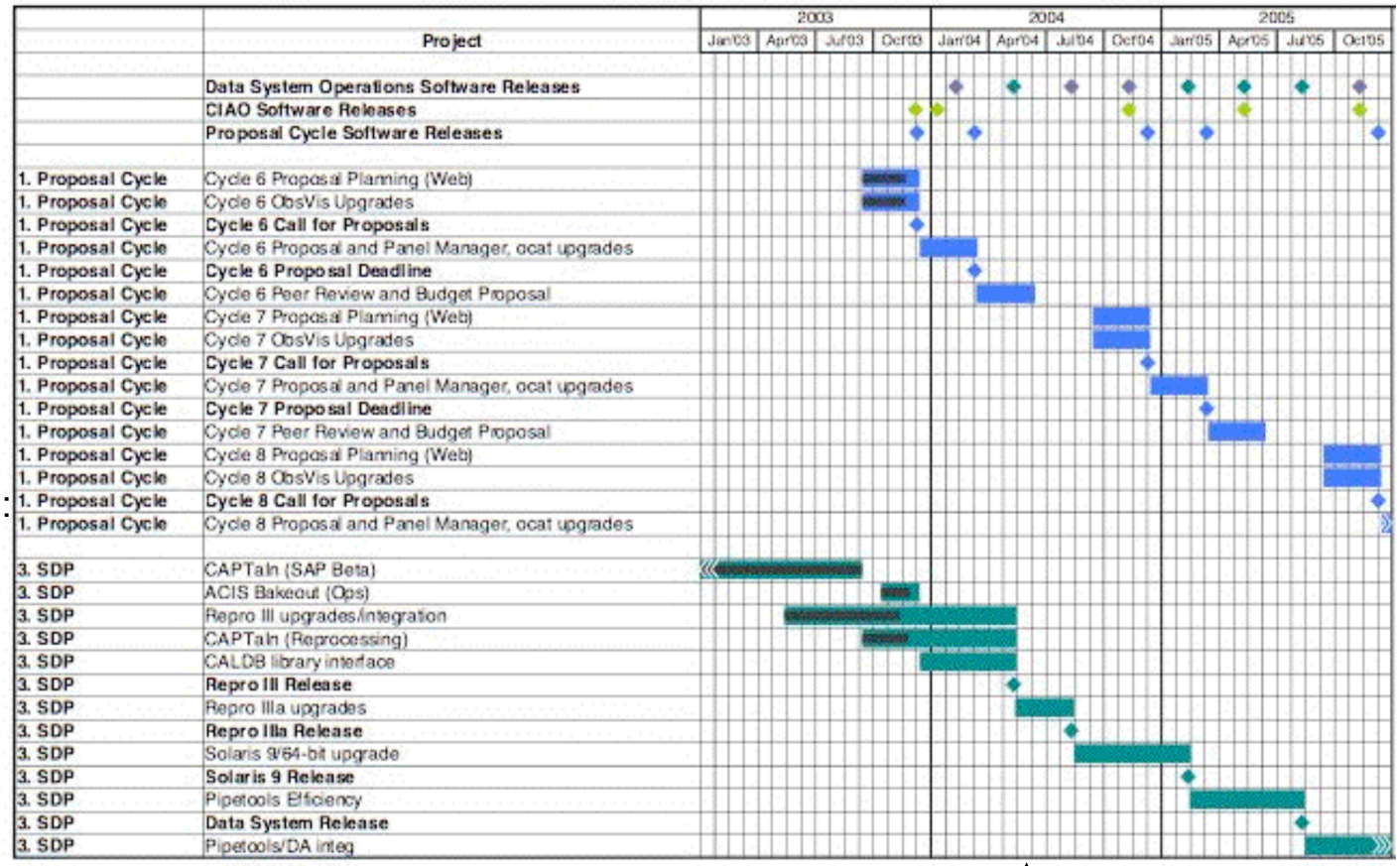
- 2-year plan presented to MSFC November 2003
- Approved and now reported against monthly
- Software Project Areas: ***SDS areas in bold***
  - Proposal cycle - *new capabilities (e.g. HST, XMM, RXTE coordination), adapt to S/C changes*
  - MTA (S/C monitoring & trends)
  - Archive - *incorporate L3. Chaser/Webchaser updates*
  - **V&V** system - *detailed automated checks*
  - **Pipeline processing** - *control and repro.3*
  - **Level 3** - *in stages (OBI based point sources, merged, extended)*
  - **CIAO 3.1** - April 2004
  - **CIAO 4** - Fall 2004
  - **CIAO 5** - Fall 2005

CXC Software Schedule 2003-2005: 1

1. Proposal cycle

2. Mission Planning: no development

3. Reprocessing

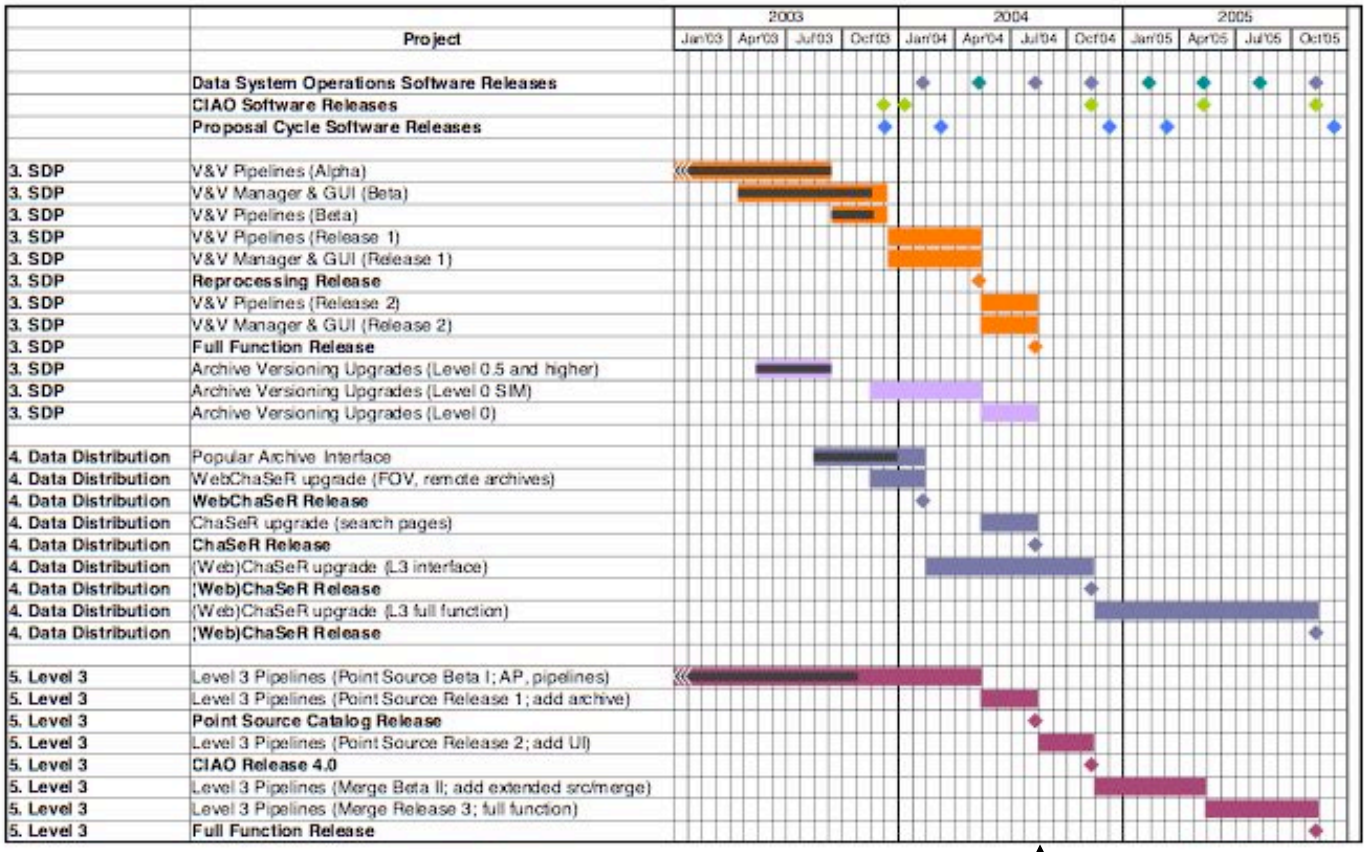


CXC Software Schedule 2003-2005: 2

3. V&V

4. Archive

5. Level 3



CXC Software Schedule 2003-2005: 3

6. MTA

7. CIAO

	Project	2003 Jan'03	Apr'03	Jul'03	Oct'03	2004 Jan'04	Apr'04	Jul'04	Oct'04	2005 Jan'05	Apr'05	Jul'05	Oct'05
	Data System Operations Software Releases					◆	◆	◆	◆	◆	◆	◆	◆
	CIAO Software Releases				◆	◆			◆		◆		◆
	Proposal Cycle Software Releases				◆		◆		◆		◆		◆
6. MTA	MTA II Trends Databases				■	■							
6. MTA	**MTA Trends II Release**					◆							
6. MTA	MTA III Trends Databases					■	■						
6. MTA	**MTA Trends III Release**						◆						
6. MTA	MTA Tool uprades						■	■					
6. MTA	MTA Source and Alignment Monitoring							■	■	■	■		
7. Data Analysis	Bug fixes				■	■							
7. Data Analysis	**CIAO Release 3.0.2**					◆							
7. Data Analysis	ACIS Bakeout				■	■							
7. Data Analysis	**CIAO Release 3.1**					◆							
7. Data Analysis	ChIPS redesign, / Speed, Scripting Languages					■	■	■	■				
7. Data Analysis	**CIAO Release 4.0**								◆				
7. Data Analysis	Enhanced functionality; bug fixes								■	■	■		
7. Data Analysis	**CIAO Release 4.1**									◆			
7. Data Analysis	Advanced Analysis, VO DM, Advanced GUIs								■	■	■	■	■
7. Data Analysis	**CIAO Release 5.0**												◆
7. Data Analysis	Enhanced functionality; bug fixes												■

CIAO 3.1 "ACIS bakeout"

CIAO 4

CIAO 5

Reprocessing 3

- Principle: *Support core software for Chandra users*
- Create uniform archive with latest calibrations
- Begin reprocessing: Spring 2004
- Pipeline Reprocessing enhancements:
  - ACIS
    - time-dependent gain
    - Accurate CC-mode event time-of-arrival
    - New hotpixel/afterglow tool
    - `destreak` applied in pipeline
  - HRC
    - Improved tap-ringing correction to event positions - better PSF, LSF
  - Gratings
    - Coordinate fixes - more accurate wavelengths
    - Pixel randomization turned off - better LSF
      - will be turned off for imaging too, when super-resolution algorithm is implemented

Level 3 Pipeline

- Principles: *1. support full spectrum of users, 2. multi-observatory analysis*
- Create catalogs of properties of all Chandra observations
- Phase 1 is point source catalog (single OBI only)
  - release May 2004 for internal debugging
  - Catalog creation begins August 2004, once archive connection is in place
  - Enables quick cross-checks with other catalogs
  - Include simple source properties
- Phase 2 (processing begins Oct 2005) options:
  - Combine multiple observations
  - Include extended sources
  - Allow upper limits to be derived
- Speed is an issue. *(~5000 OBIs so far in 4 years of Chandra operation)*
  - detect tool `sextractor` adapted to Poisson statistics
    - In test. Appears to be fast, accurate; gives speed improvement needed
  - Also need hardware (Beowulf cluster?)

CIAO 3.1

- Principle: *Support core software for Chandra users*
  - *I.e. S/C, instrument changes have priority*
- ACIS bakeout release, CIAO 3.1: Spring 2004
- CIAO 3.0 has structure to deal with ARF spectral/time dependence
- ARF spatial dependence will be in CIAO 3.1,
  - ready for ACIS bakeout
- Will include 2-part RMF in CALDB:
  - zero-CTI “ideal” (“launch”) matrix
  - CTI degradation matrix
  - speeds, simplifies calibration, better physical description
- `mkrmf2` will create RMFs from 2-part input matrices
- `acis_process_events` will include time dependent gain

Beyond CIAO 3

- Principles: *Support full spectrum of users, Chandra-unique capabilities*
- Chandra scientific needs assessment
  - User survey
    - In CIAO 3: scripting, access to architecture, timing
    - Others: visualization/graphics, speed, syntax, i/f (GUI)]
  - CUC comments, priorities
  - HelpDesk common problems, requests
  - Workshop feedback
  - Systematic review of analysis paths - details next
- Leads to priorities for CIAO 4 contents
  - Not just new tools:
    - speed, stability, scripts;
    - threads, guides, web services

Chandra Scientific Needs Assessment

- Create set of outline analysis paths, ~30 cases
- Aim to span analysis space:
  - Major analysis modes: SPATIAL, SPECTRAL, TEMPORAL
  - Minor modes: IMAGING, GRATING
  - Source type:
    - Point: SINGLE, FEW, MANY
    - Extended: SMALL (~PSF), LARGE, VERY LARGE (~FOV)
  - Major modes can mix:
    - Spatial-spectral: IMAGING, EXTENDED
    - Spectral-temporal: IMAGING, POINT (SINGLE -> MANY)
    - Spectral-temporal: GRATING, POINT (SINGLE)
    - Spatial-temporal: IMAGING, EXTENDED (SMALL)
  - Each instance has a
    - Detector mode: ACIS-I, ACIS-S, HRC-I, HRC-S
    - Grating mode: NONE, HETG, LETG
    - Only some combinations are common: HETG/ACIS-S, LETG/HRC-S, LETG/ACIS-S

Chandra Scientific Needs Assessment

- For each analysis path:
  1. make an outline thread detailing steps
  2. Put CIAO versions of steps where available
  3. Identify alternative steps (e.g. IDL, FTOOLS) if available
  4. Where not software available put in algorithm
  5. If no algorithm enter `research needed'
- Current status:
  - ~ 2/3 done at outline level
  - ~ 1/3 done at CIAO, alternate level
  - Completion scheduled Jan 31 2004
- Prioritize
  - How common each analysis is
  - How much of a show stopper each missing feature is
    - *Using helpdesk, Workshop feedback, CUC reports & SDS knowledge as inputs*
- Create schedule

Chandra Scientific Needs Assessment

- Expect that CIAO 3.0 extensibility will allow most new features to be scripted quickly by SDS scientists.
- *New approach*
  - But not entirely: CIAO 3 and even CIAO 2.3 have scripts of this kind

    E.g. `lc_clean, analyze_ltcurv, merge_all, spectrum, acisspec, acisabs, chart_spectrum, sherpa_utils`
- Expect differing levels of integration depending on demand
  - E.g. specialized script; general script; script with .par file, DM I/F.
  - Full integration as a CIAO tool has a significant overhead

    coding standards, DM, regression tests, error condition handling, .par file, ahelp, bugs listing, threads, analysis guides
- Not every feature will be full integrated
  - Would lose flexibility, quick response advantage of CIAO 3, and limit scope
- The most valuable features will become CIAO 4 tools

CIAO 4

- Some features are long lead items and clearly needed
- So CIAO 4 preparation has begun
- **Upgrade to Visualization, Graphics** (ChIPS) USER SURVEY REQUEST
  - Problems with SM as base plotting package:
    - GPL licensing
    - Publication quality line graphics
    - Combining imaging and line graphics
    - Scriptability in CIAO
  - Now assessing alternate plotting packages against detailed specs
  - Phased implementation being plotted out
    - E.g. publication quality line graphics first
- **Speed enhancements** USER SURVEY REQUEST
  - Identifying main offenders, acceleration methods

CIAO 5

- Prime directive will guide allocation of resources:

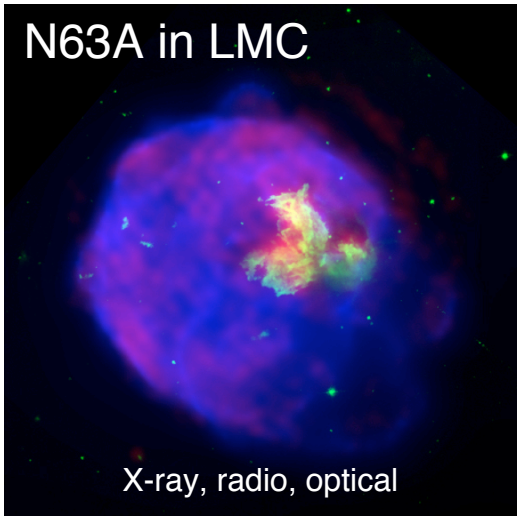
To enable the best science to be extracted from Chandra data

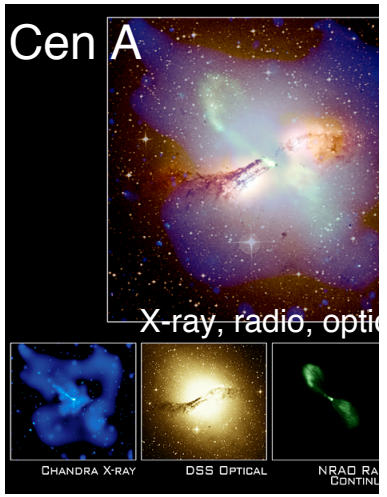
- Ultimate performance of Chandra should be realized
  - physical ACIS model, sub-pixel imaging with ACIS
  - Flight HRMA model for core and wings
  - Imaging pile-up model
  - Grating wavelength scales to 50km/s
  - Optimal deconvolution methods (example shown at June 2003 CUC meeting)
  - Some of these are in early development
- VO will become a reality in FY05
  - CXC must be ready
- Release: Fall 2005

Virtual Observatory & CIAO

- CXC Principles:

  Support multi-observatory analysis, Support Chandra-unique capabilities
- CUC says:
  - *"…[VO] should be accorded a lower priority than other themes until an overall vision is endorsed by the committee"*
  - *"…long term efforts be coordinated with other large astrophysics data analysis efforts."*
- **Conclusion: CXC cannot ignore VO**
- What we are doing:
  - Long term collaboration with HEASARC (follow, enhance standards)
  - Coordination with XMM (e.g. FITS keywords)
  - Participate actively in NASA ADEC coordination process (eg GLAST)
  - Chair international data model definition group for VO (J. McDowell)
  - Chair IAU FITS commission (A. Rots)
  - Participate in NASA Centers joint committee (ADEC) (G. Fabbiano, A. Rots)
  - Added ADS interoperability to Chandra Archive
- VO uses only Science time of CXC staff, plus 2 FTE NSF-funded developer positions (thru 8/04)





Scripting Languages & CIAO

- CUC report of June 2003:

  "…concerned about the choice of Slang…"

  1. "*cost in effort to those in the community who need to learn yet another system*"
  2. *"questionable long term support for Slang in the community"*
  3. *"the small size of its user community"*

3. "the small size of the user community"

- Not important at this stage
  - Slang is used by *all* CIAO users (under the hood)
    - No need to learn Slang
      - *C.f. XSPEC, ds9, MSWord, all of which have scripting languages, though few users are aware of this*
  - CIAO scripting community will grow
    - as demand warrants CXC will
      - Develop documentation, examples
      - hold training sessions

2. "questionable long term support for Slang in the community"

- Slang 1.0 is stable and frozen
  - under development since 1991: 5 releases in 2 years
- Slang is written in ANSI standard C
  - Ports wherever C ports: 16-, 32-, 64-bit systems. PDAs?

    msdos, 32bit-msdos(djgpp), Windows (3.x, 95, 98, xp, NT), OS/2, VMS, various unix (solaris, linux, *BSD, AIX, Ultrix, HP/UX, QNX, OSF, …), unix-like systems (BEOS, CYGWIN, Mac OSX).
  - Most ports done by Slang community.
  - Slang is small (20k lines of code. ~1/4 code size of Python?) Encourages porting.
- Cannot predict technology 5 years from now
  - Parallelization?
  - 128 bit machines?
  - Any language could become obsolete

1. "cost in effort to those in the community who need to learn yet another system"

- **Learning curve for those who wish to should not be large**
  - Slang is "IDL-like". IDL is the only widely used scripting language in astronomy
  - SDS will provide sample scripts
  - Any choice of language would have a learning curve *(e.g. variable access calls "get_X")*
- **CIAO 3 allows CXC to incorporate new features faster**
  - So far most new Chandra features came from CXC, GSFC or PSU
  - Has been hard for smaller centers to contribute
  - Module incorporation (C or Fortran code) fairly easy, minimizes Slang coding
- **Scripting languages are in a state of flux with no clear front runner**
  - Only IDL has a large user base in astronomy, but cannot be distributed in a data analysis package
  - Astronomer-in-the-street uses C, IDL, or Fortran
  - Python, Tcl/Tk, perl, Slang scripts mostly confined to major centers.
    - Tcl/Tk was front runner a few years back *(used by XSPEC)*
    - CXC adopted Slang in CIAO 2.0 *(released Dec 2000);* CIAO 3 opens Sherpa to Slang
    - NRAO adopted Glish (c. 1999?) then abandoned for Python (2003)
    - STScI adopted Python in Pyraf *(released 2003).* Today's front runner
    - Perl is widely used *(even in CXC) but not by a major center*
    - *ROOT (from CERN, a `command line C') used by INTEGRAL, GLAST?*

Scripting Languages & CIAO

- Slang provides great value in CIAO
- No urgent need, or resources, to replace Slang
- Pursuing, at low level, ways to be multilingual
- Watching developments
- Will report back to CUC

CIAO 3 Scripting Presentations

1. Spectra
   - **A.** Using Slang functions and modules in Sherpa.
     Aneta Siemiginowska
     - Better analysis from improved flexibility
   - **B.** Better modeling of line, continuum components in coronal spectra
     Nancy Brickhouse
     - Better physics from improved flexibility
2. Timing/Spectra
   Mike Nowak
   - Easily examine thousands of sources via scripting
3. Imaging spectroscopy
   John Houck
   - Mapping by repeated analysis steps via scripting
4. Grating spectroscopy
   David Huenemoerder

Sherpa in CIAO 3

Aneta Siemiginowska

- Sherpa CIAO 3 Highlights:
  - Greatly improved robustness and reliability of 2D image fitting
  - Support for multiple components in one instrument model expression
  - Flexible configuration of plots
  - Data access via Sherpa S-Lang functions
- Sherpa S-Lang module function types
  - `get` => retrieving data & settings e.g. get_data, get_fluxed_spectrum
  - `ls` => query, e.g. is_subtracted
  - `list` => retrieving string-based info. E.g. list_par, list_proj
  - `load`=> run computational tasks, e.g. run_regproj
  - `run` => put and set the data into Sherpa, e.g. set_data, set_back
    - and set plotting options, e.g. set_erroff, set_log
  - `set` => restore default settings, e.g. restore_proj

Scripts on the CIAO Web Site

Aneta Siemiginowska

- Examples of Sherpa S-lang scripts now on CIAO Web site:
  - Calculate k-correction
  - Light curve filtering
  - Modify default Sherpa plots
- http://cxc.harvard.edu/ciao/download/scripts/
- `CIAO_script.tar` contains all currently available scripts
- Easy to install:
  - Untar in directory where CIAO has been installed: `/soft/ciao`
- Documentation and help available on Web and via `ahelp`

Modelling Grating Coronal Spectra

Nancy Brickhouse

- Example of New Flexibility in CIAO 3:
- Continuum Modeling with APEC:
  - Better physics for coronal sources
- Slang Script Courtesy of Ronnie Hoogerwerf

Grating coronal spectra: Overview

Nancy Brickhouse

- Goal: increase flexibility for analyzing spectral data using plasma models
- Example: fit continuum model to line-free regions
- Before CIAO 3.0:
  - Global fit method
    - use entire spectrum
    - works ok to derive a few parameters
    - not ok for diagnostic lines
  - Line-based analysis
    - uses the most interesting information
    - not accurate for weak lines because of blending with other lines and continuum
    - often uses local continuum (fit by “eye-ball”)
- Demo will show *Iterative approach*:
  - Standard global fit to Capella HETG spectra using XSAPEC model in Sherpa
  - Import into Sherpa a SLang script by R. Hoogerwerf with the following commands:
    - Find line-free regions
    - Plot spectrum indicating line-free regions
    - Notice only the line-free regions
  - Fit XSAPEC to line-free regions and show where those regions are
  - Interpolate to find continuum model for regions containing lines

Grating coronal spectra: Status

Nancy Brickhouse

- We have shown new modeling flexibility available in CIAO 3 with contributed Slang script in Sherpa.
- We have demonstrated the path to new capabilities for users at all levels: to iterate between global fitting and line-based analysis.
- Line-free scripts developed for particular applications will be generalized, documented, tested, and brought into CIAO script library, depending on SDS priorities.
- Generalization could include: testing on ACIS spectra; other algorithms to determine line-free regions, e.g. bin-by-bin line-to-continuum ratio test; other forms of continuum, e.g. power law fits to absorption line-free regions; model-independent continuum regions, e.g. based on sigma rejection/acceptance.

Timing/Spectra

Mike Nowak

Imaging Spectroscopy

John Houck

Grating Spectroscopy

David Huenemoerder

Conclusions

- CIAO 3 is a powerful system, a big step forward
  - Enables SDS development of features via scripts, modules
    - Speed, flexibility
- 2-year plan based on prime directive
  - ***To enable the best science to be extracted from Chandra data***
- 4 principles establish SDS priorities:
  1. Core Chandra features: keep up with instrument, S/C performance
  2. Fully support Chandra-unique capabilities
  3. Support *all* types of user
  4. Enable multi-observatory analysis
- Reprocesssing, Level 3 address these principles
- CIAO 4, 5 based on systematic science needs assessment

SDS Presentation

Questions?