A Universe of Data

Jonathan McDowell

Smithsonian Astrophysical Observatory

21st century astronomy has seen amazing new telescopes on Earth and in space

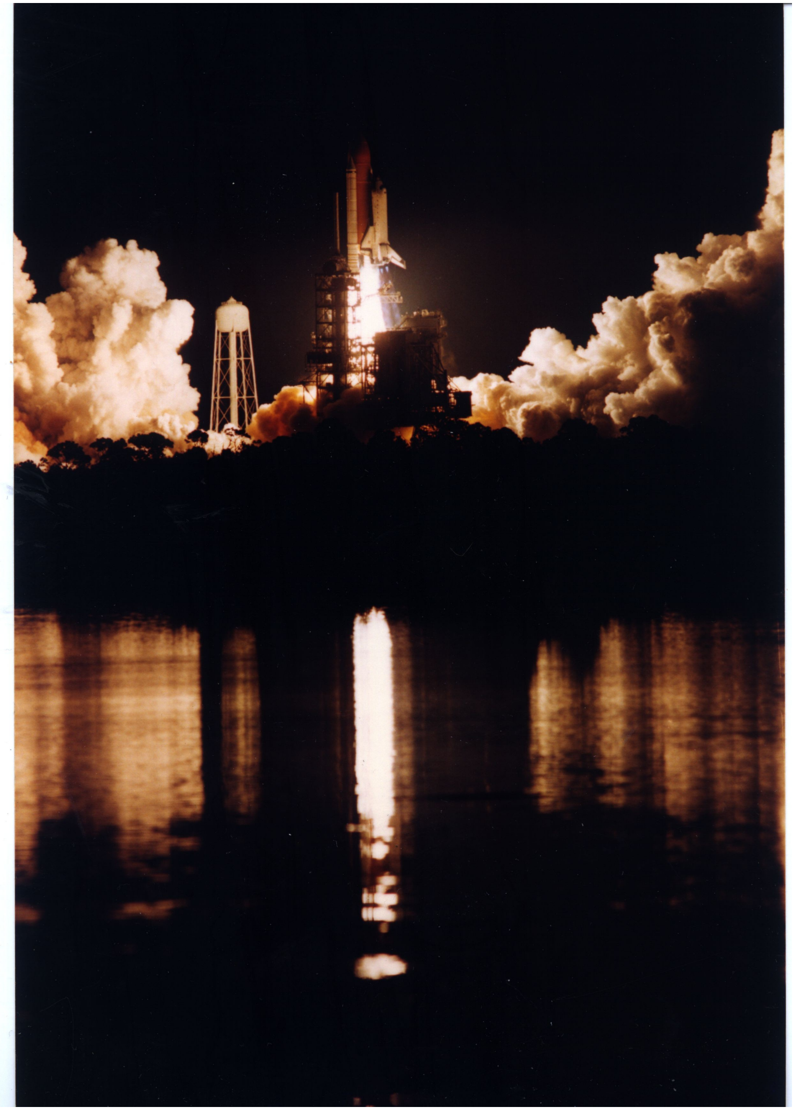
But the new discoveries we're making are not just because of better telescopes – there's a whole new way of using them

First I'll introduce the Chandra space telescope and tell you a little about its discoveries

Then I'll explain the challenges and possibilities of astronomy in the digital age, where the software systems are as important as the space technology.

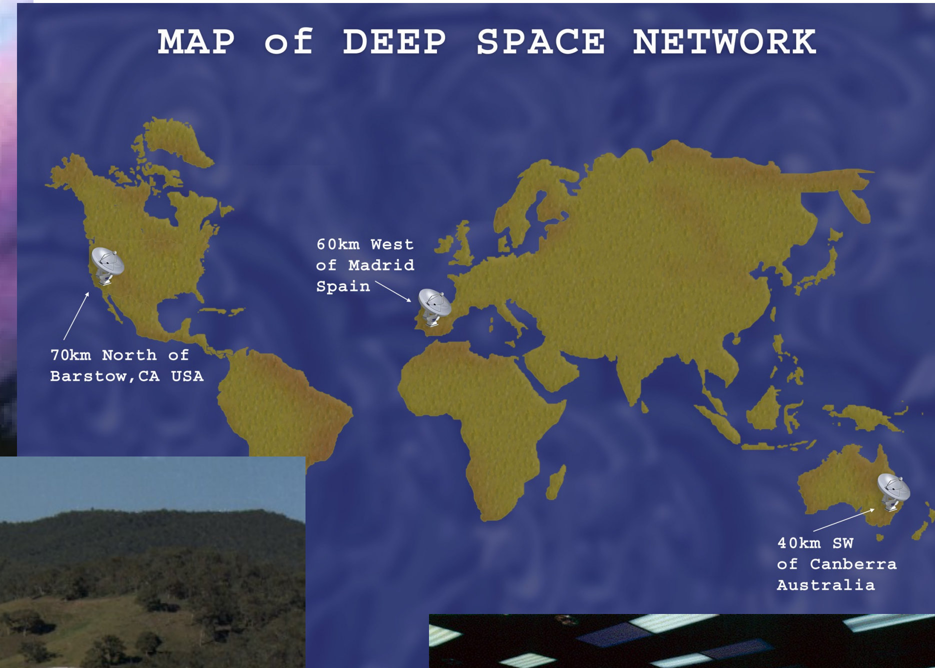
Part 1

Chandra reveals the invisible universe



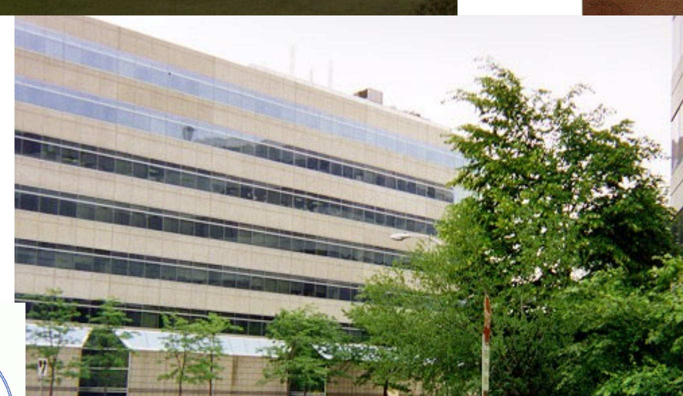
In 1999 NASA launched the Chandra X-ray Observatory into space to study the invisible energetic universe

Today, the Chandra space telescope continues to return amazing images of the cosmos – but not just images. I'll share some Chandra discoveries and then explore how astronomers use space data to unravel the stories of the heavens



DSN control
at Jet Propulsion Lab
Pasadena, CA

Chandra science center
Smithsonian Observatory, at
Harvard (Cambridge, MA)



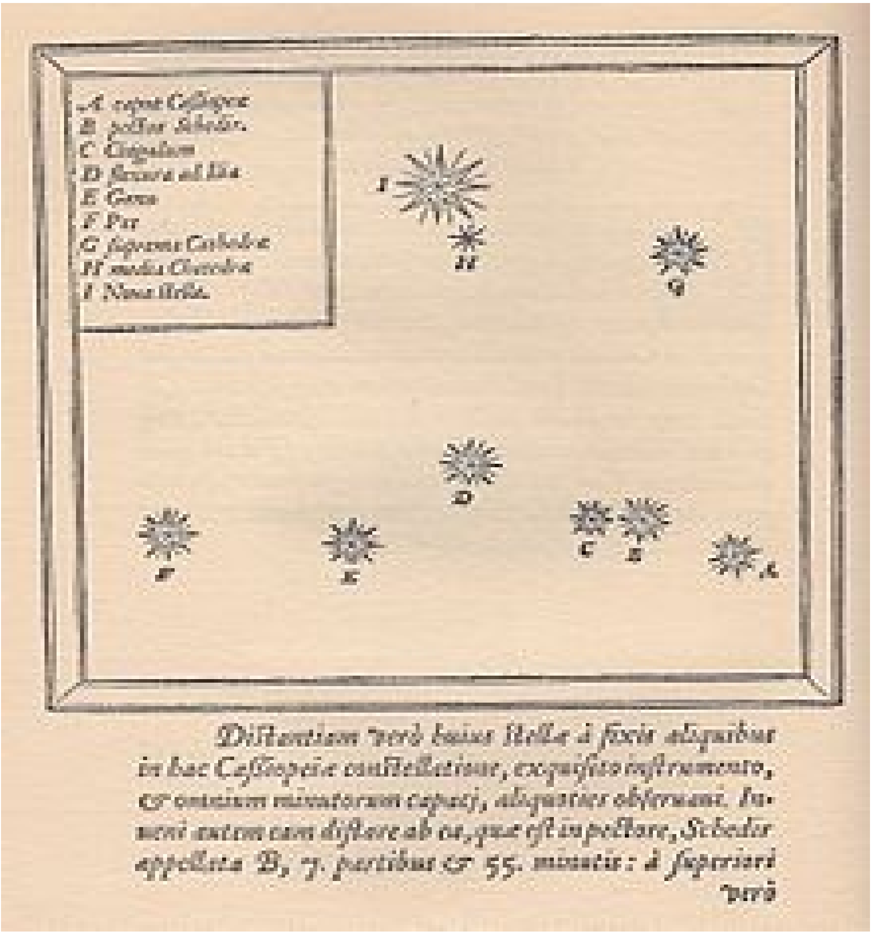
Chandra mission control
Near MIT in Cambridge, MA

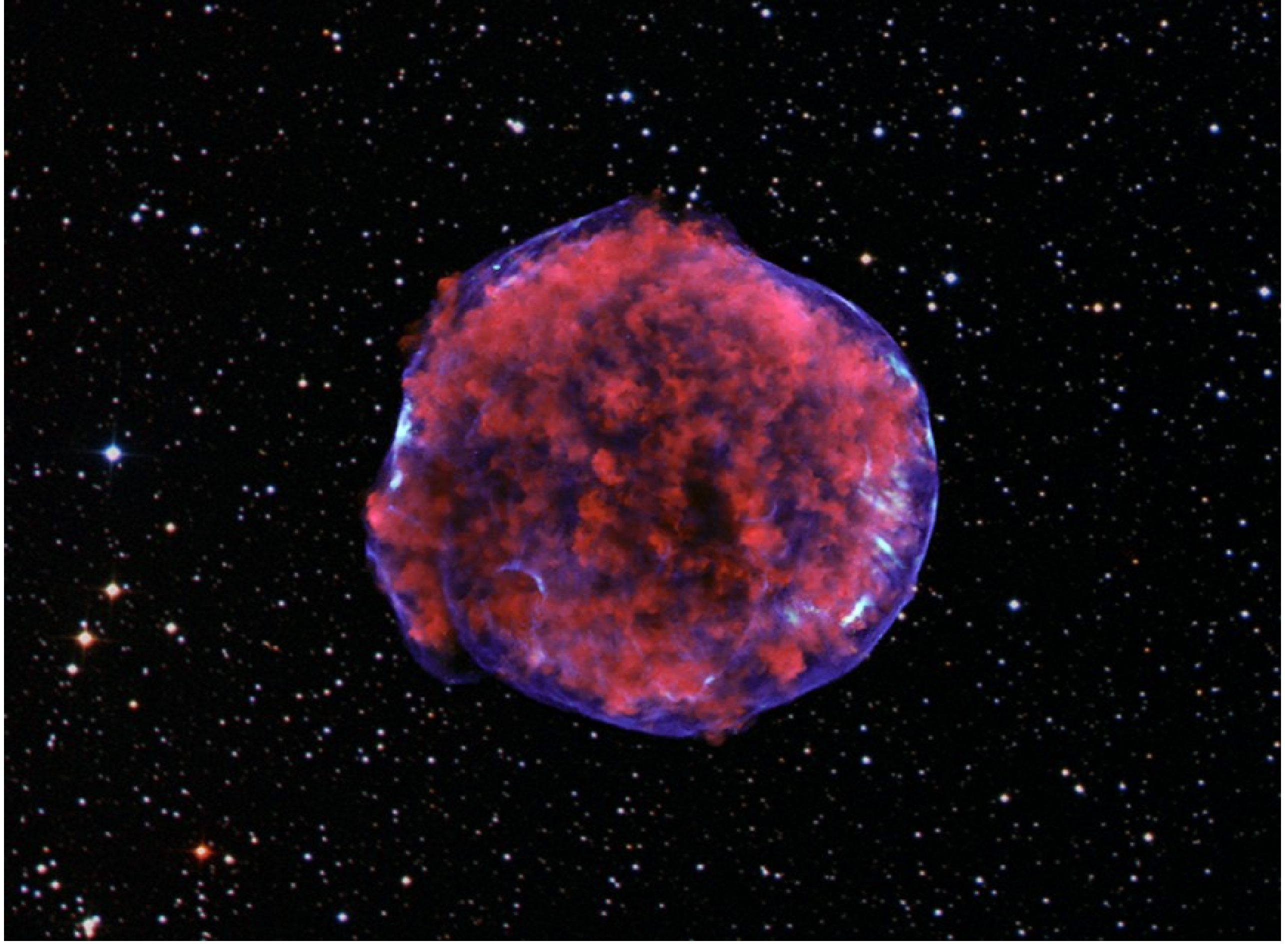
SAO operates
Chandra on
behalf of NASA

Let's visit the constellation Cassiopeia

In 1572, Danish astronomer Tycho Brahe recorded a 'new star' in the constellation Cassiopeia

It was visible to the naked eye until 1574, slowly fading from view..

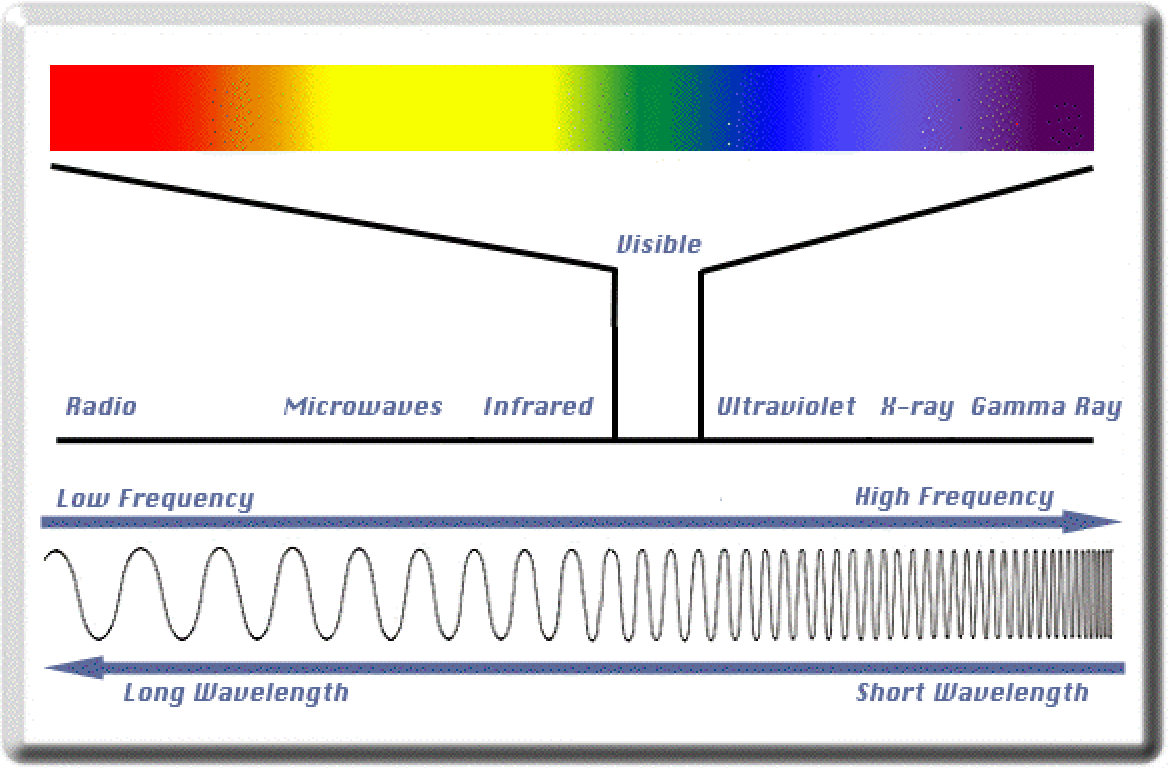




Credit:
Kris Eriksen et al, 2011 Ap J Letters 728, L28

The rainbow we all know is a sequence of colors from red through green and blue to violet – how our eyes make us perceive the different WAVELENGTHS or FREQUENCIES of light

Color in light is exactly the same as 'pitch' in sound waves
- blue is higher pitched light than red
- you can think of the invisible kinds of light like 'infrared' and 'x-ray' as different musical keys, many octaves away from the visible 'key' of light that we see



Digression: What's an X-ray?

A lot of people are familiar with, but confused by, medical X-rays

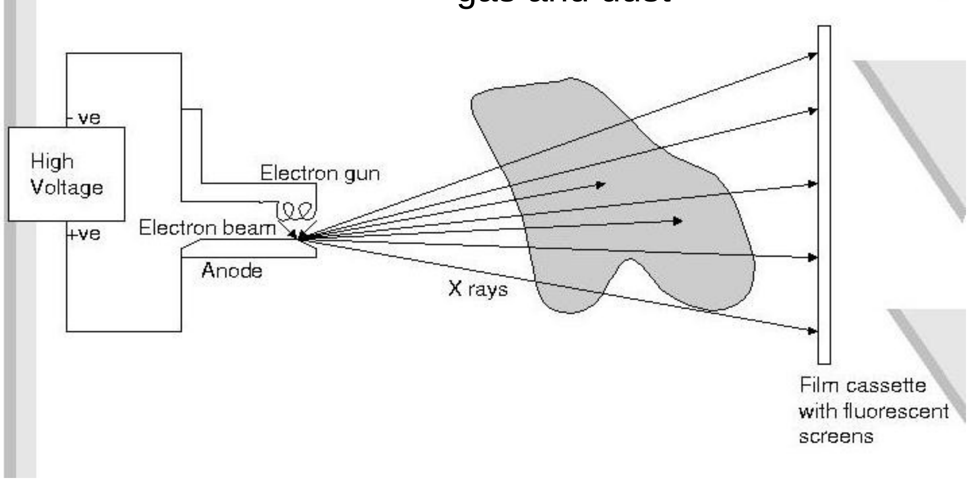
The photo at left is a picture of an X-ray light bulb, photobombed by someone's hand

The X-rays are the light bit. The dark areas are where there aren't any X-rays because the hand has blocked them.

X-ray bulb
= star,galaxy

Hand
= interstellar gas and dust

X-ray camera



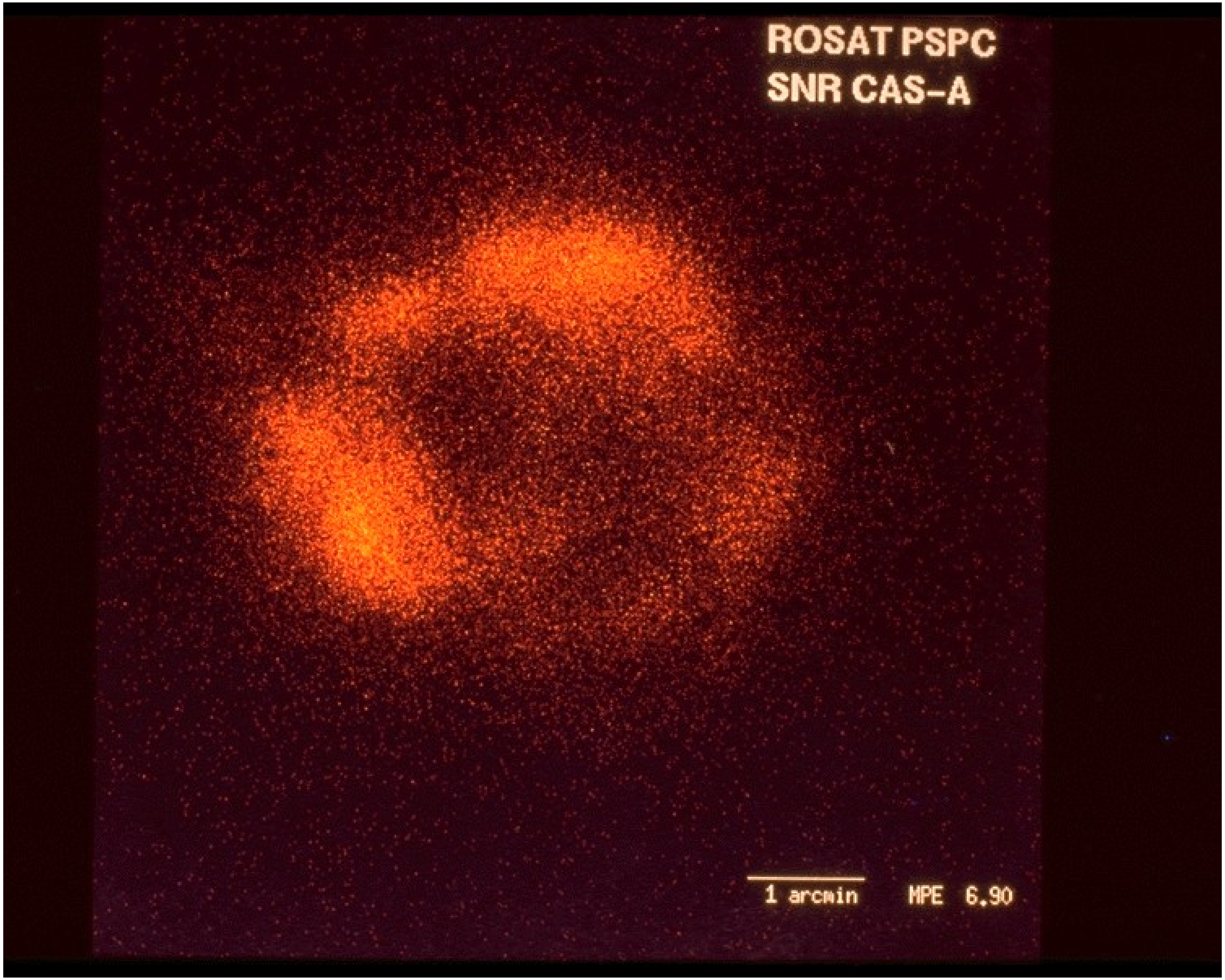
In X-ray astronomy we are usually taking a picture of the “light bulb” (the star making the X-rays) and not interested in the “hand” (stuff blocking the X-rays between the star and us)

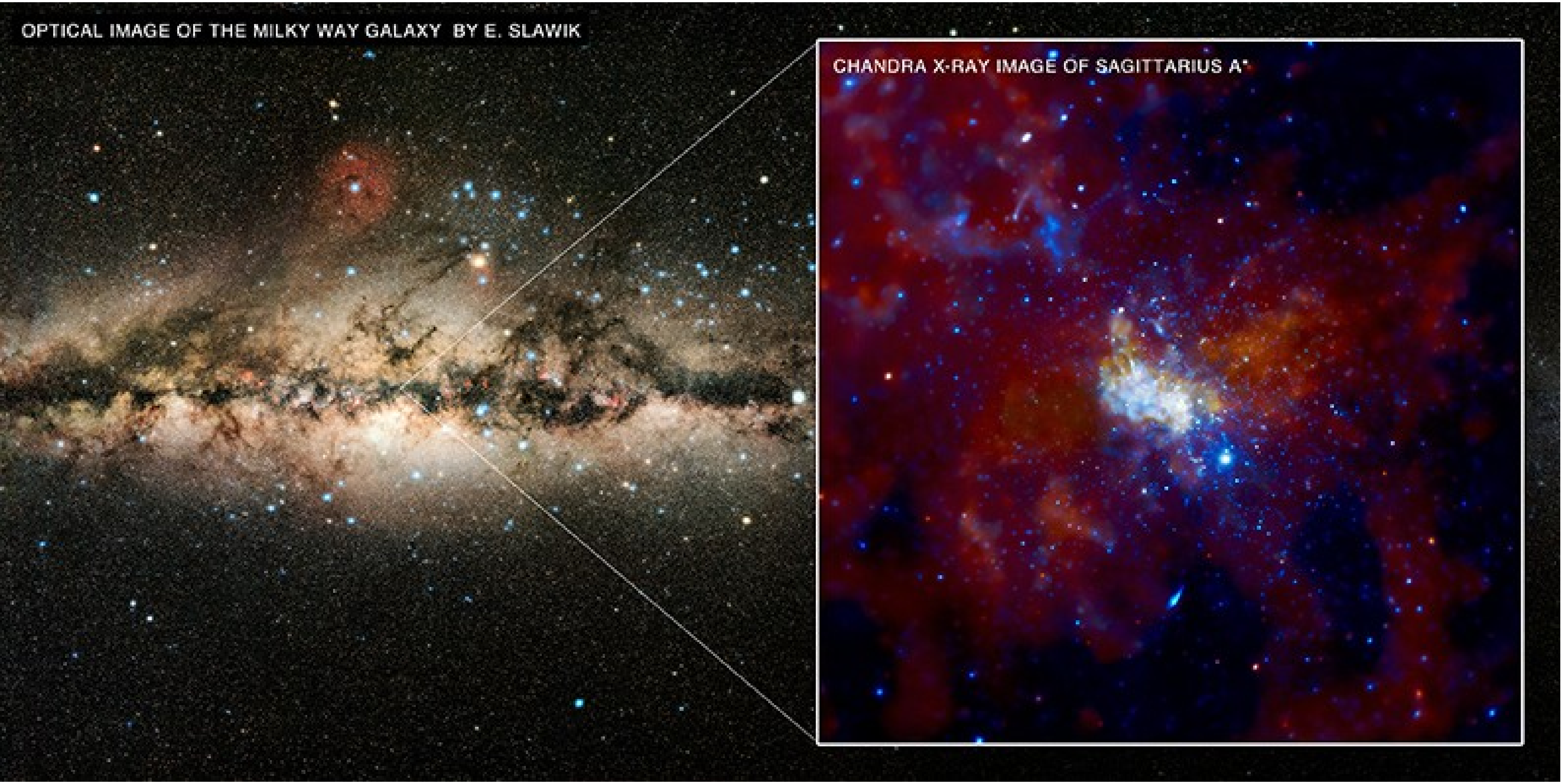
Visible-light photons are like raindrops
- each one is 'small' (has a small amount of energy)
- there are lots of them, but don't do any damage

X-ray photons are like hailstones
- each one is 'big' – lots of energy
- there are many fewer of them
- but each one packs a wallop

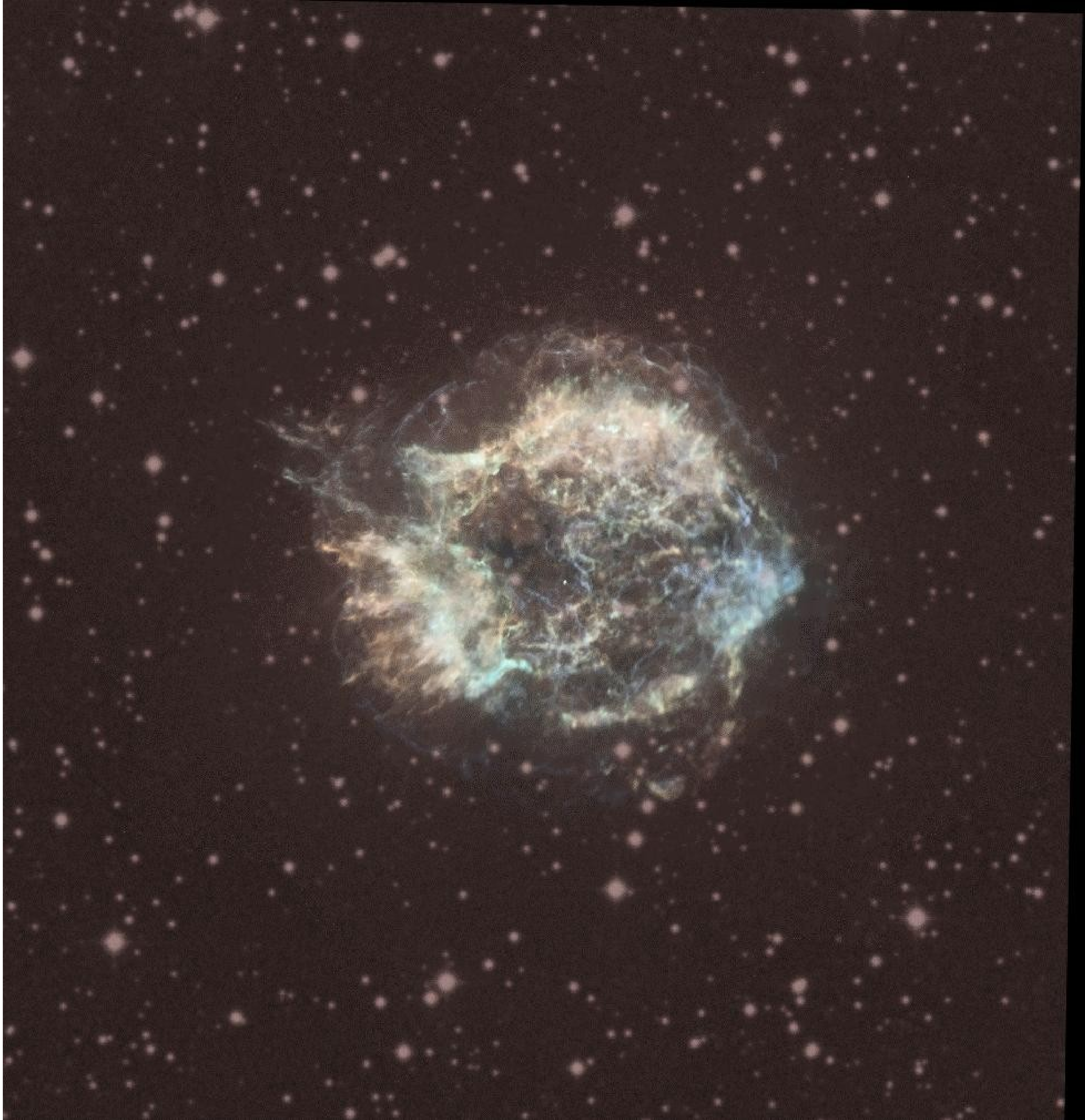
If you up the INTENSITY (number of photons) in a beam of light you increase the total energy you get but not the energy per 'packet'
If you want to get a tan (or worse) you have to increase the energy per photon, not just the number of photons.
We have a word for the energy of a photon: “COLOR”
(well, “COLOUR” but I'll defer to the local sensibility)

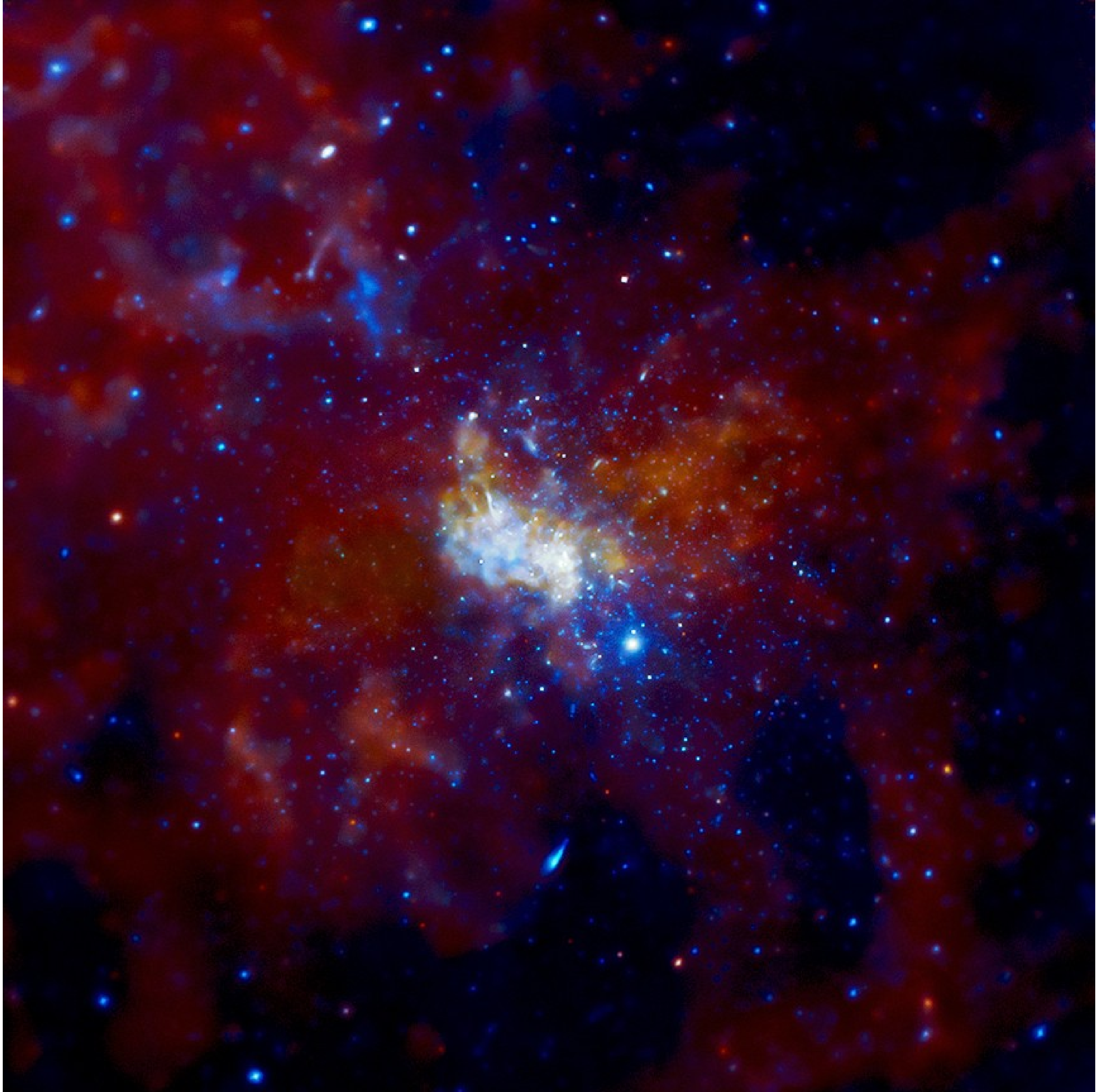
Milky Way galaxy: Supernova remmant (X-ray)
This was the best X-ray picture before Chandra



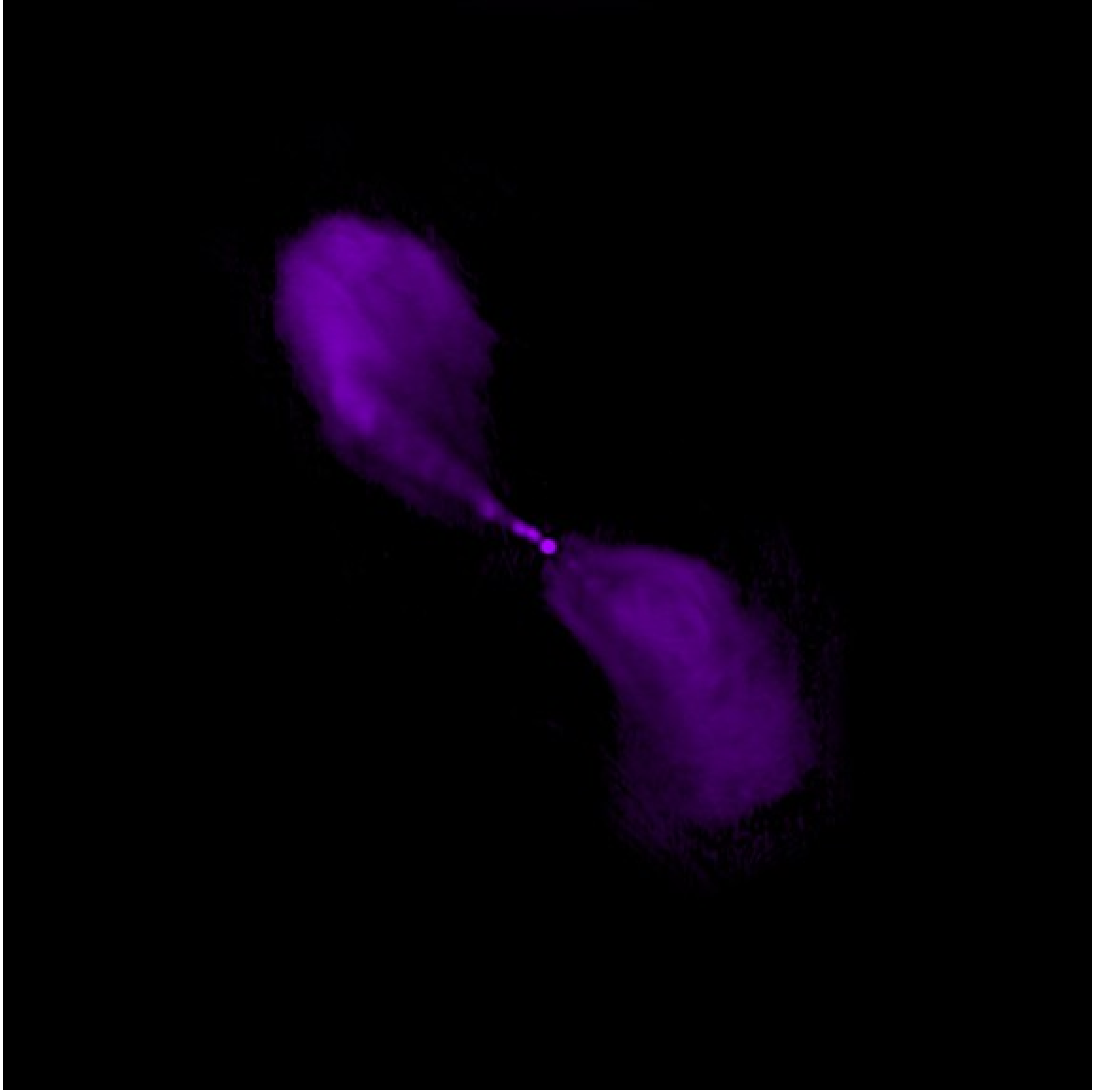


Milky Way Galaxy: Galactic Center

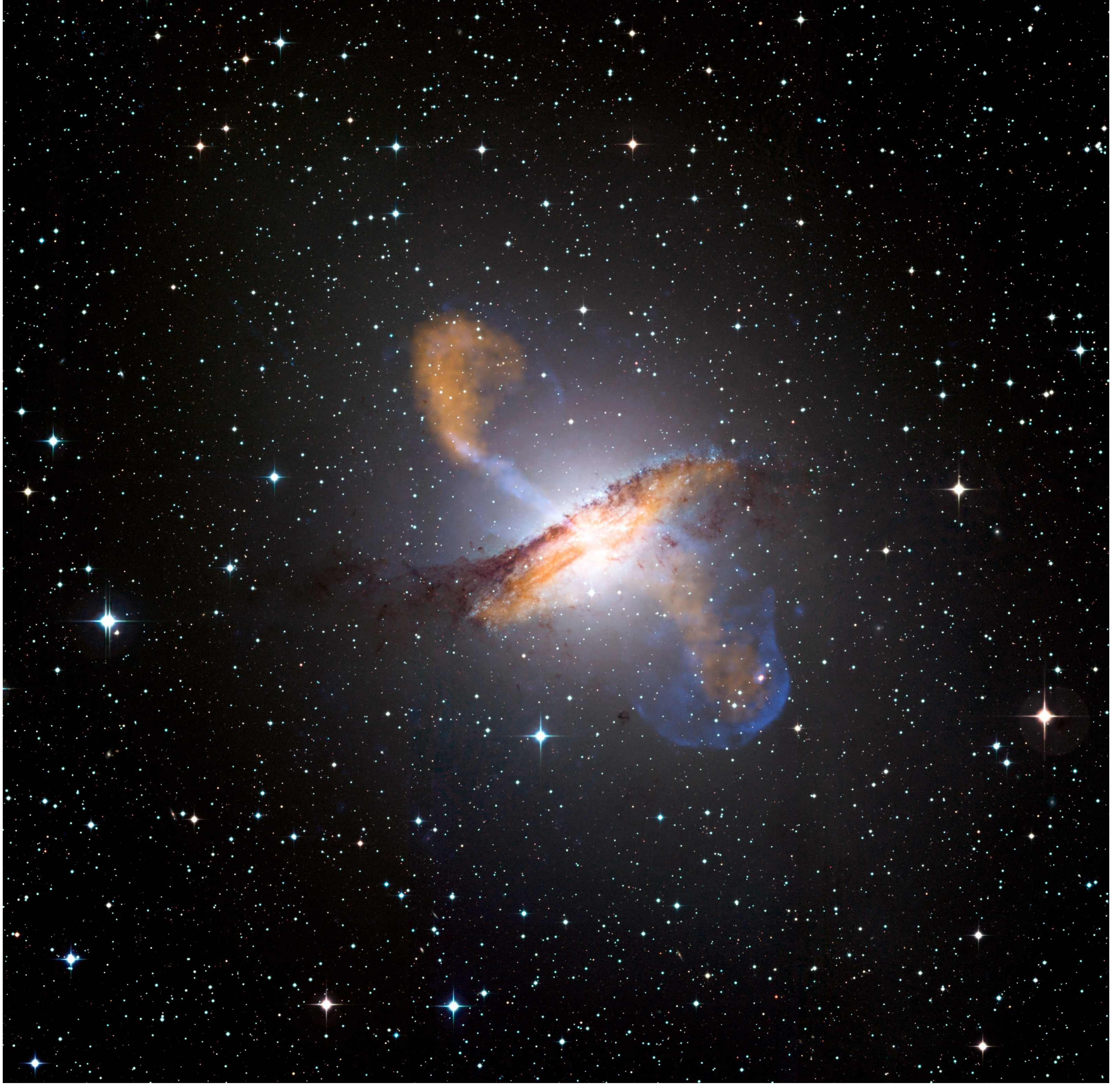





Galaxy Centaurus A (NGC 5128) - 12 million light years away



Extragalactic Universe: Active Galaxy (X-ray)

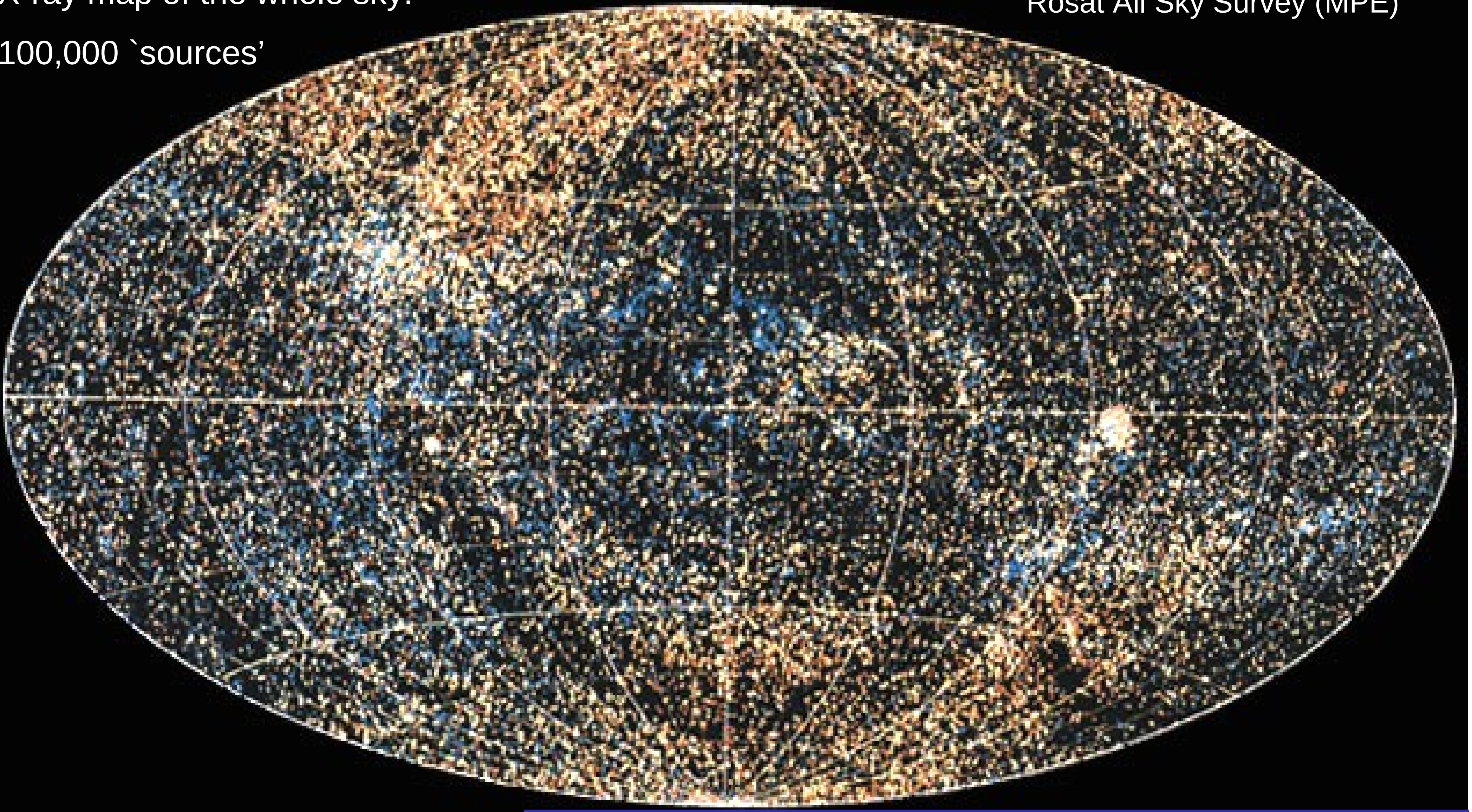


Powerful sources of X-rays

X-ray map of the whole sky:

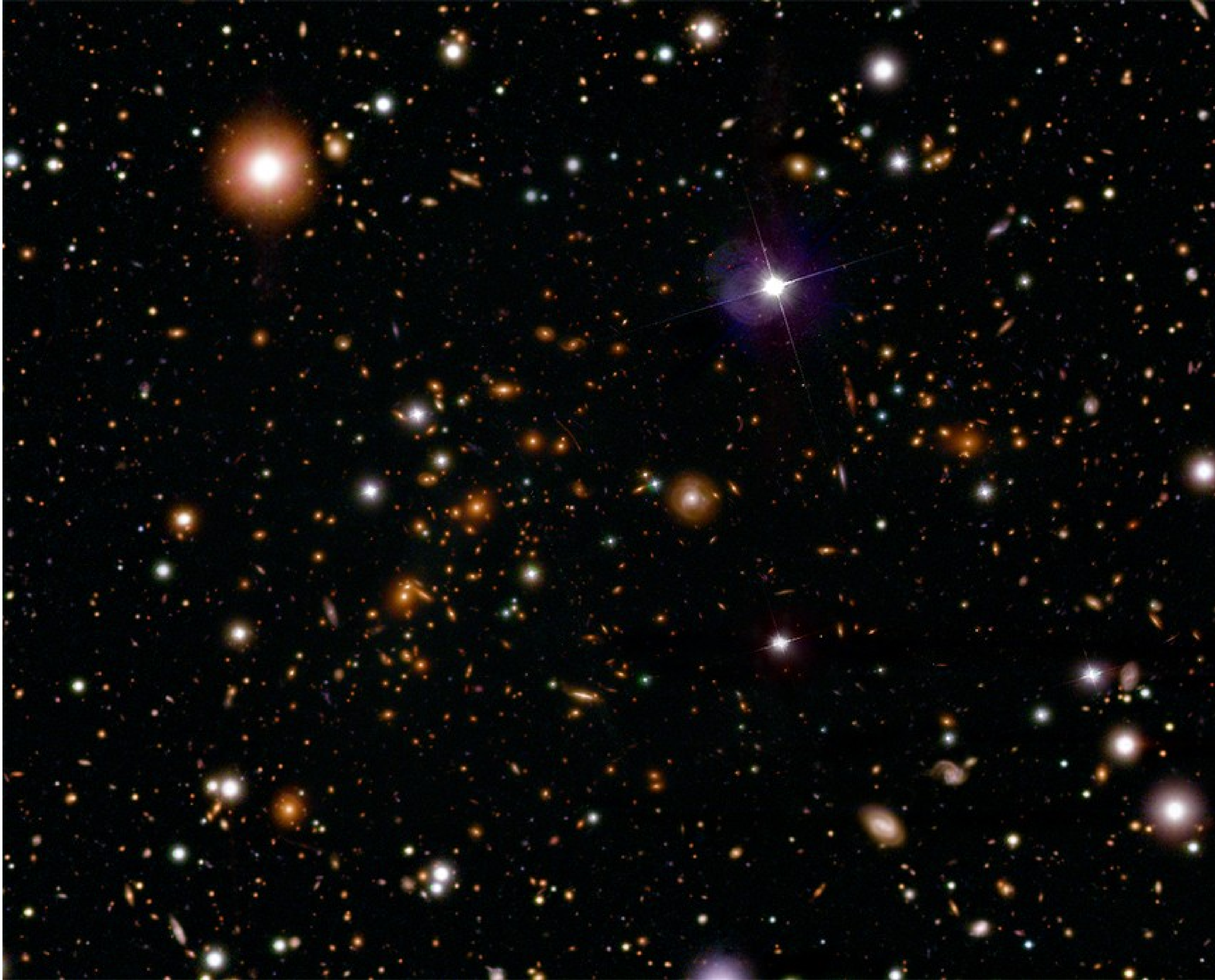
100,000 `sources'

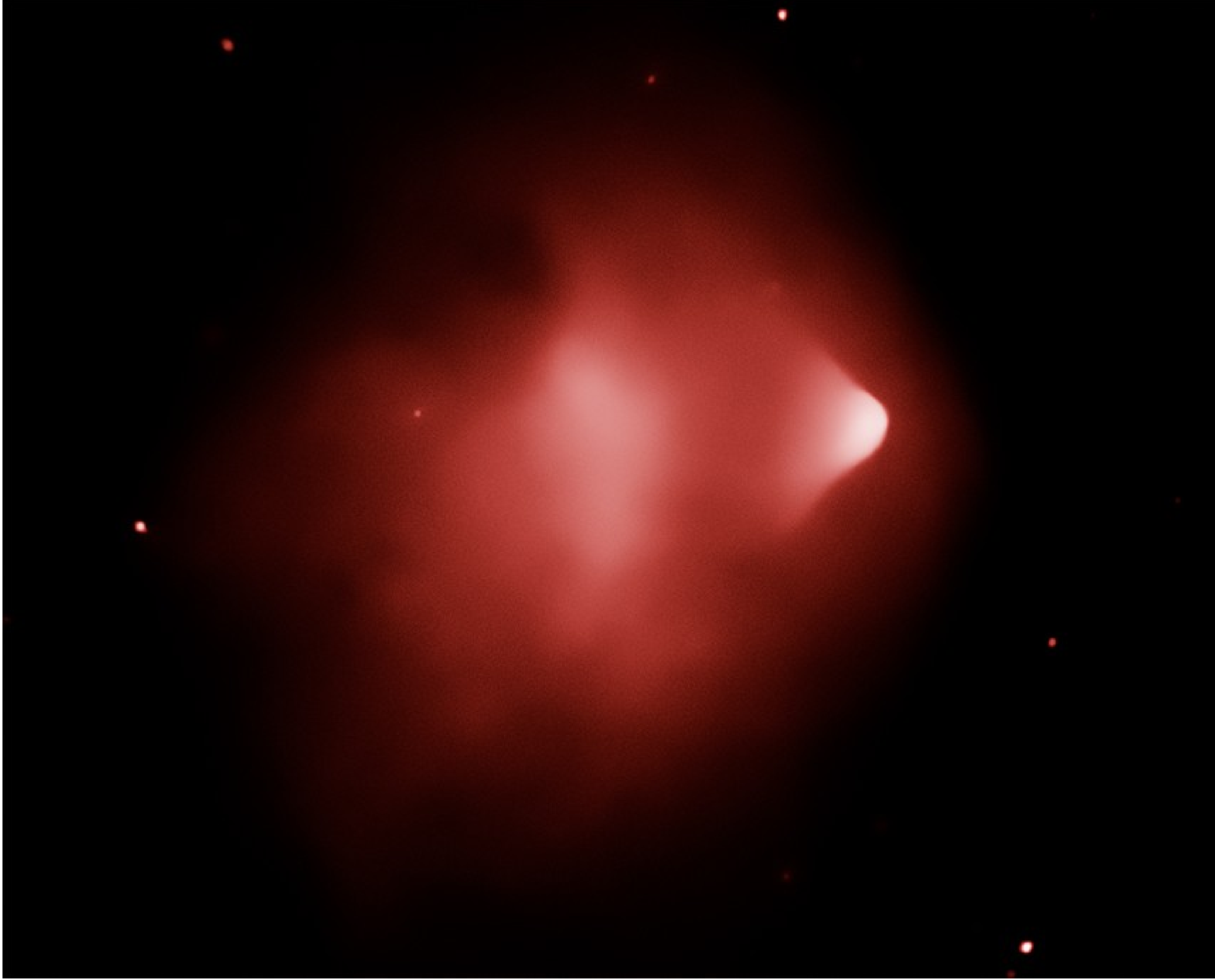
Rosat All Sky Survey (MPE)

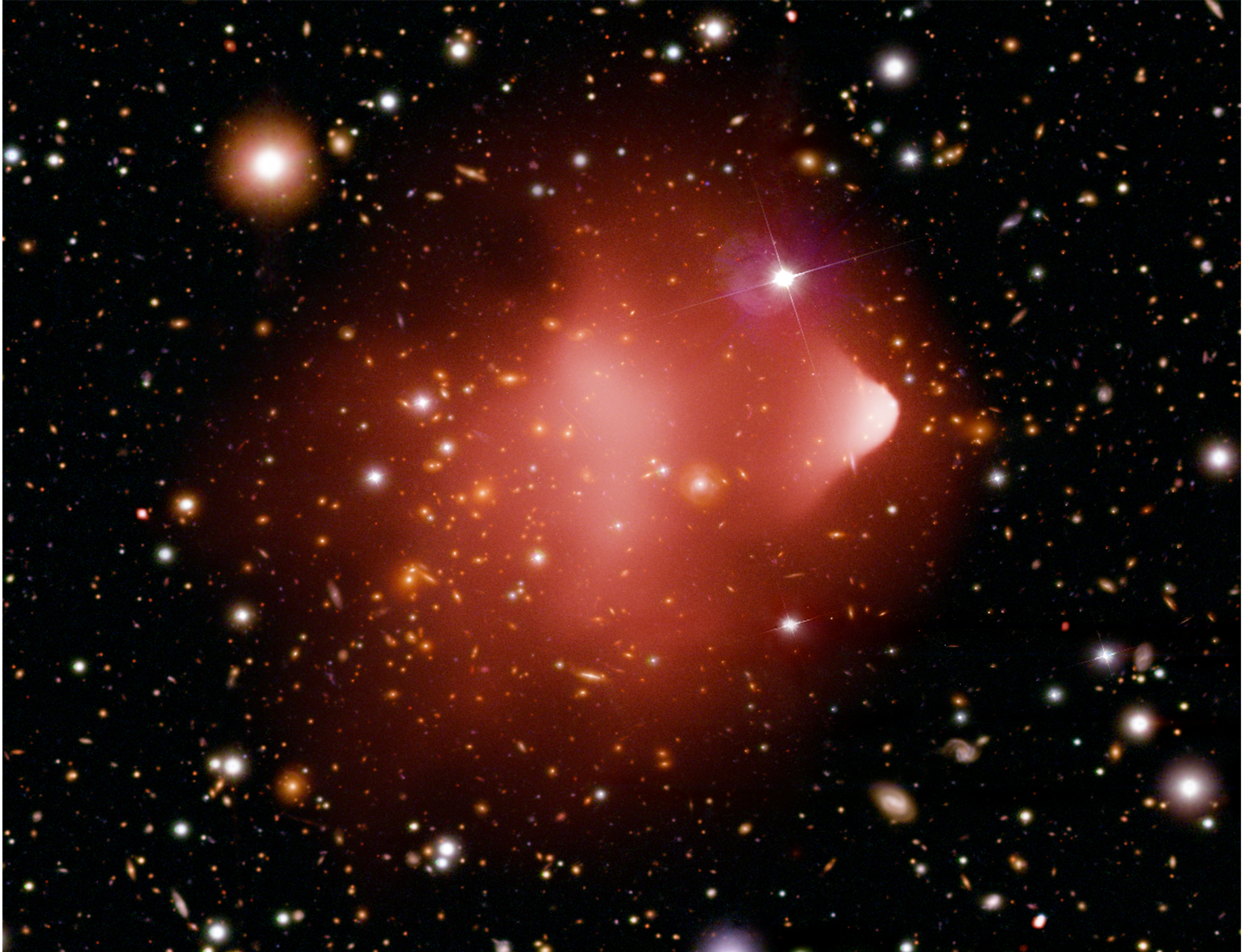


A power source entirely different from the nuclear fusion that drives the Sun and stars

…and much more efficient







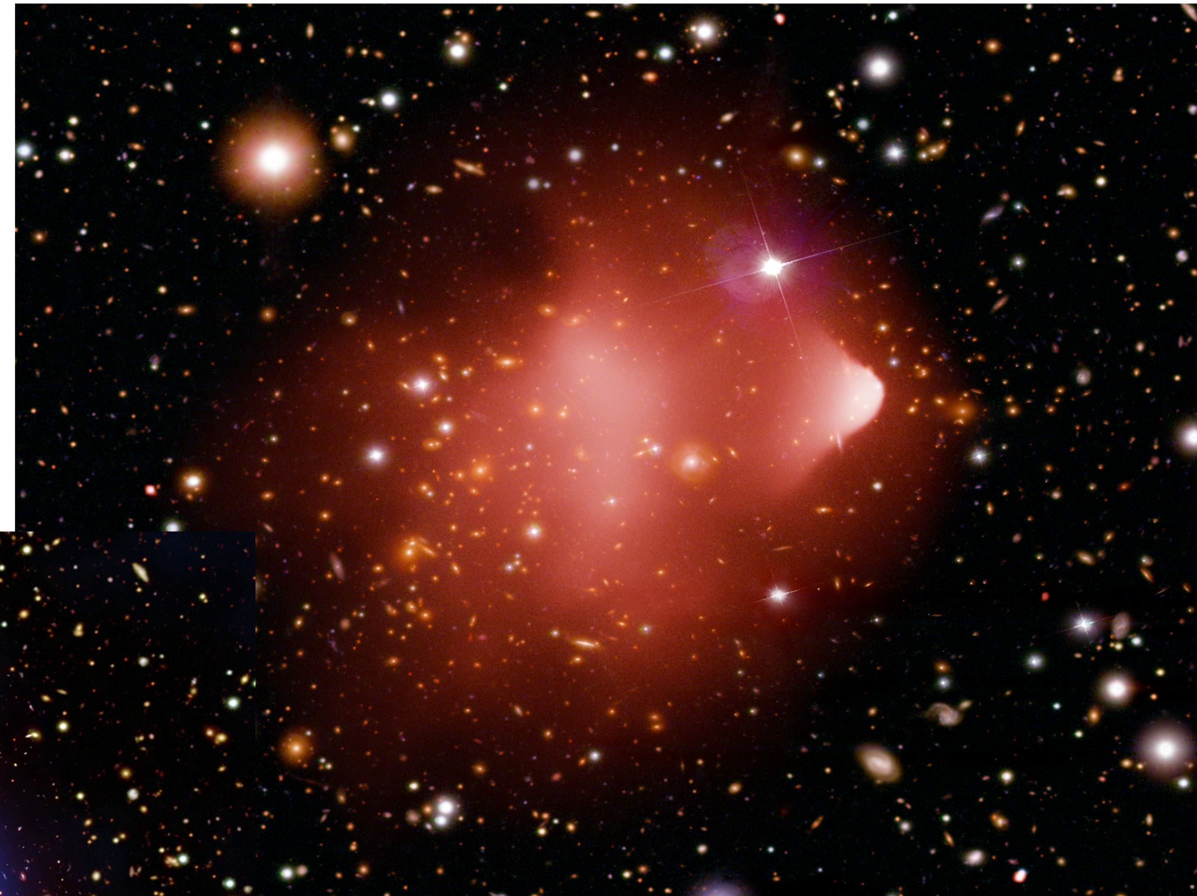
Extragalactic universe:
Cluster of galaxies (X-ray, visible and dark-matter model)

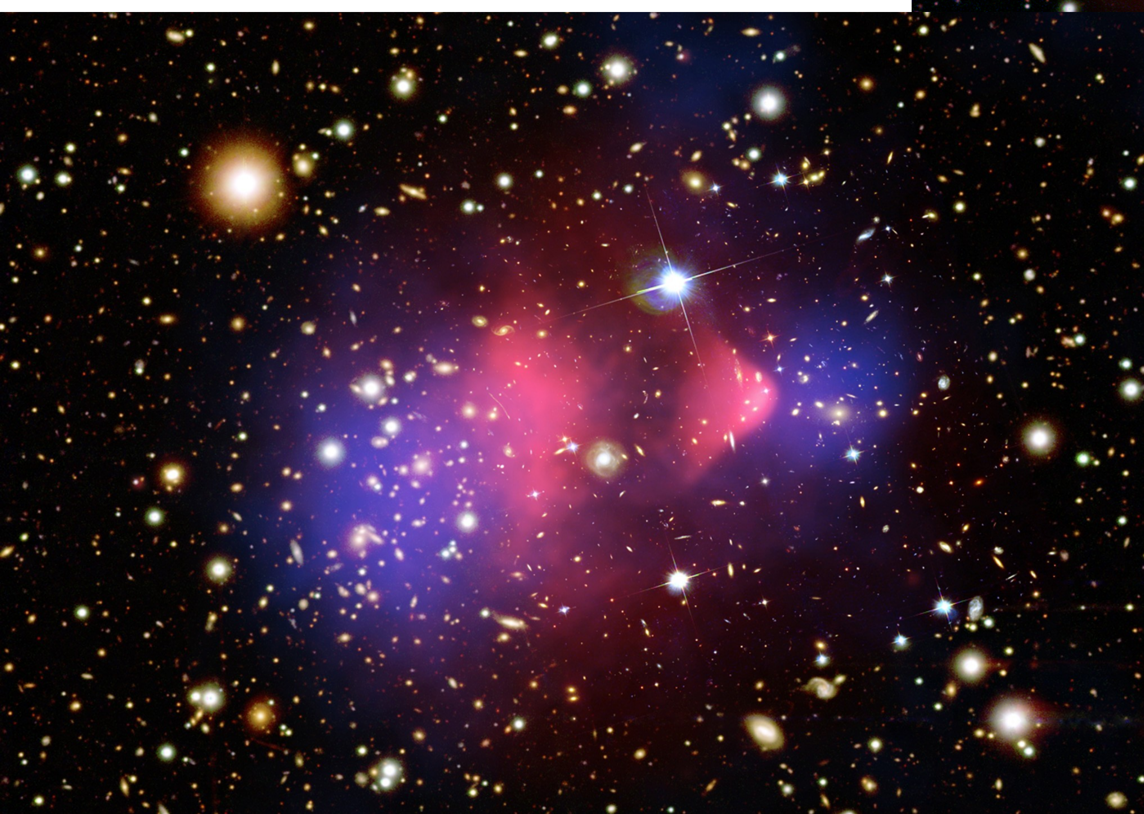
The Bullet Cluster, 1E0657-56

Two clusters in collision: studying this object let us measure the dark matter

Right: what we see directly in X-rays (red) and optical

Below: blue shows the matter distribution we infer





Distance: 3.3 billion light years

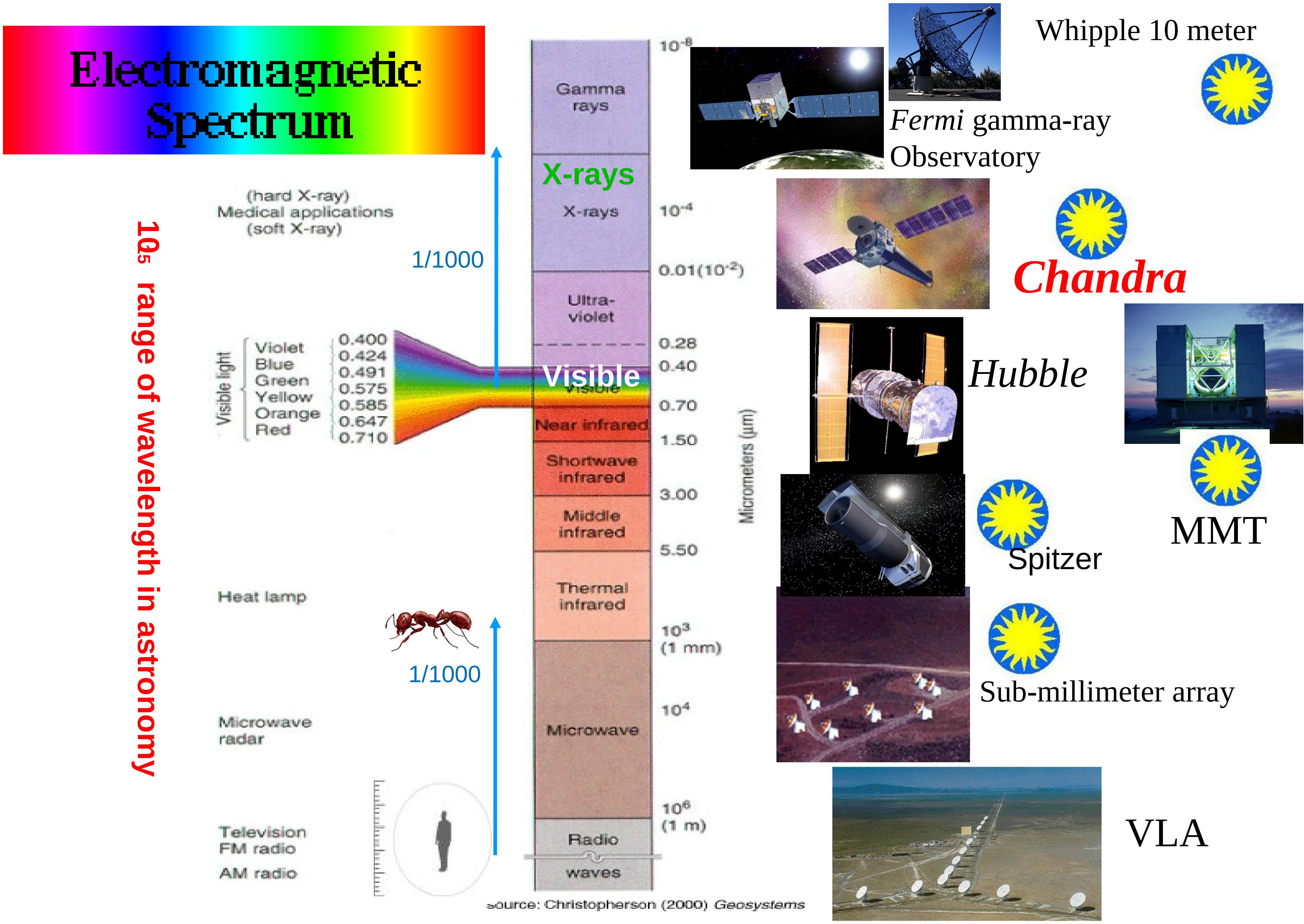
Size: 3 million l.y.

Data: Maxim Markevitch et al.

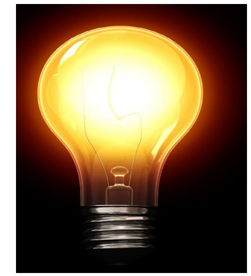
Part 2
A Universe of Data

How do we take these amazing space pictures?
How do we make scientific measurements with them?

We are now in the era of multiwaveband astronomy



Beyond the Pretty Picture

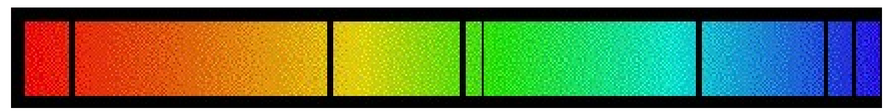


When? Timing

Where? Astrometry

How Bright? Photometry

What Color? Spectroscopy



and

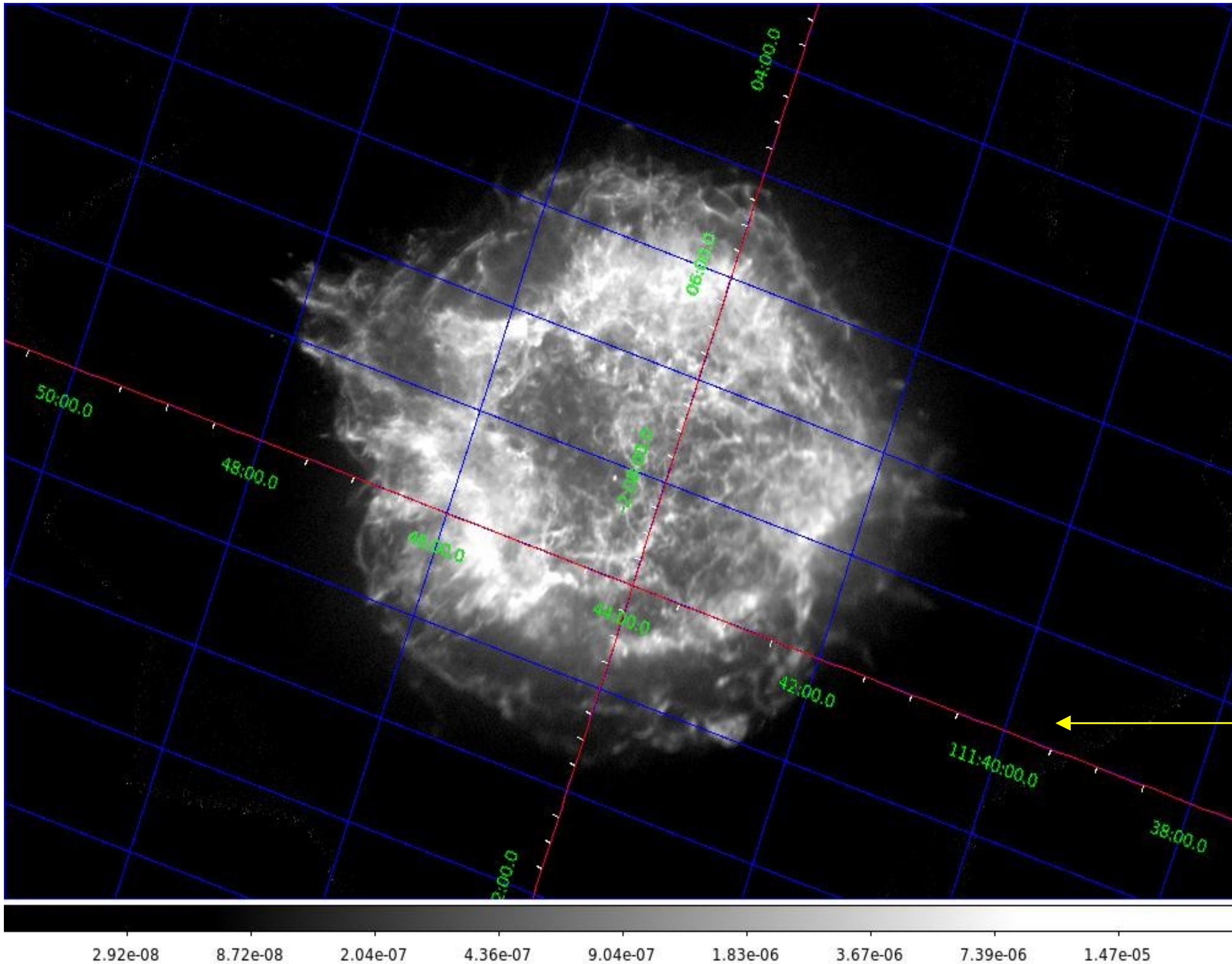
How am I being fooled?

- the camera is lying to me
  bad pixels, sensitivity changes...
  instrument background...

- the software is lying to me
  processing artifacts
  calibration issues
  mismatched data

- the universe is lying to me
  Are these two objects touching each other or just in the same direction?

Different Kinds of Data 1 - Images



HEADER

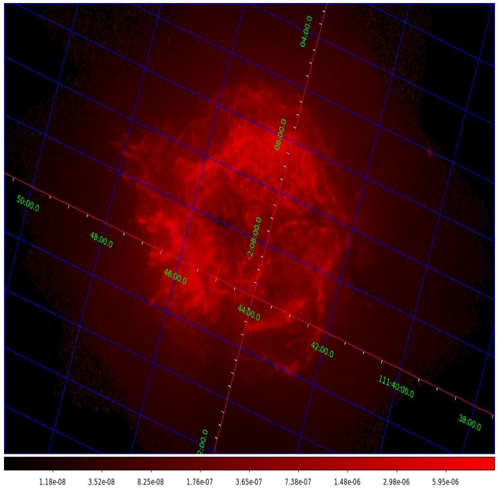
Date and time
Wavelength range
Which telescope and where
etc etc etc

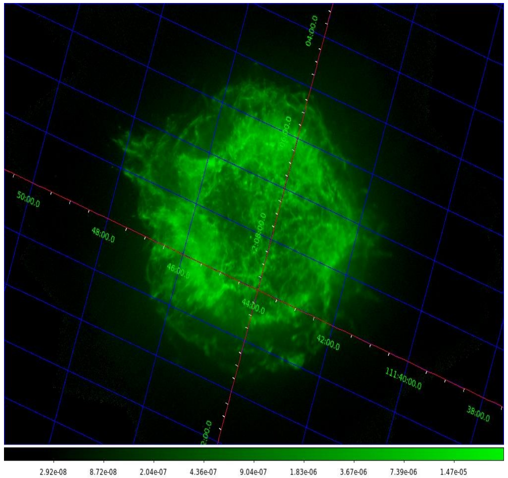
Galactic latitude and longitude grid

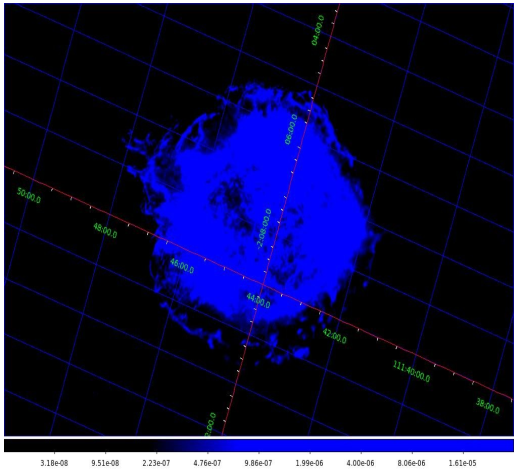
Need 3 of these (“red”, “green”, “blue”) for a color image

Calibrated brightness in milliwatts/sq m
- how bright is it in this particular wavelength range

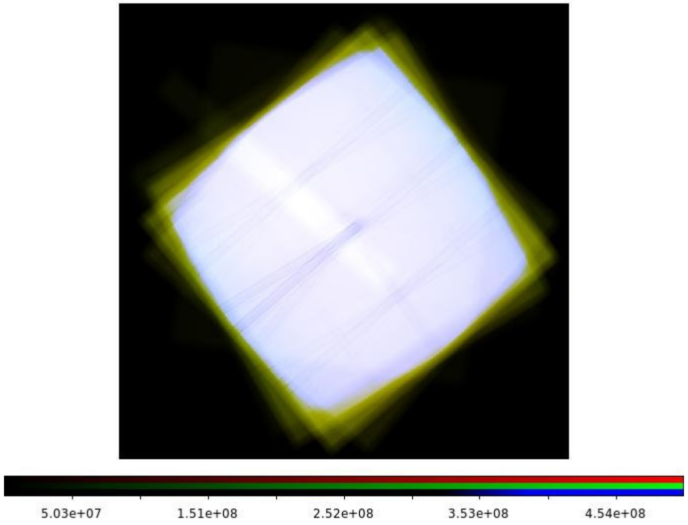
Different Kinds of Data 1(b) – Lots of Images







Data images in different color ranges

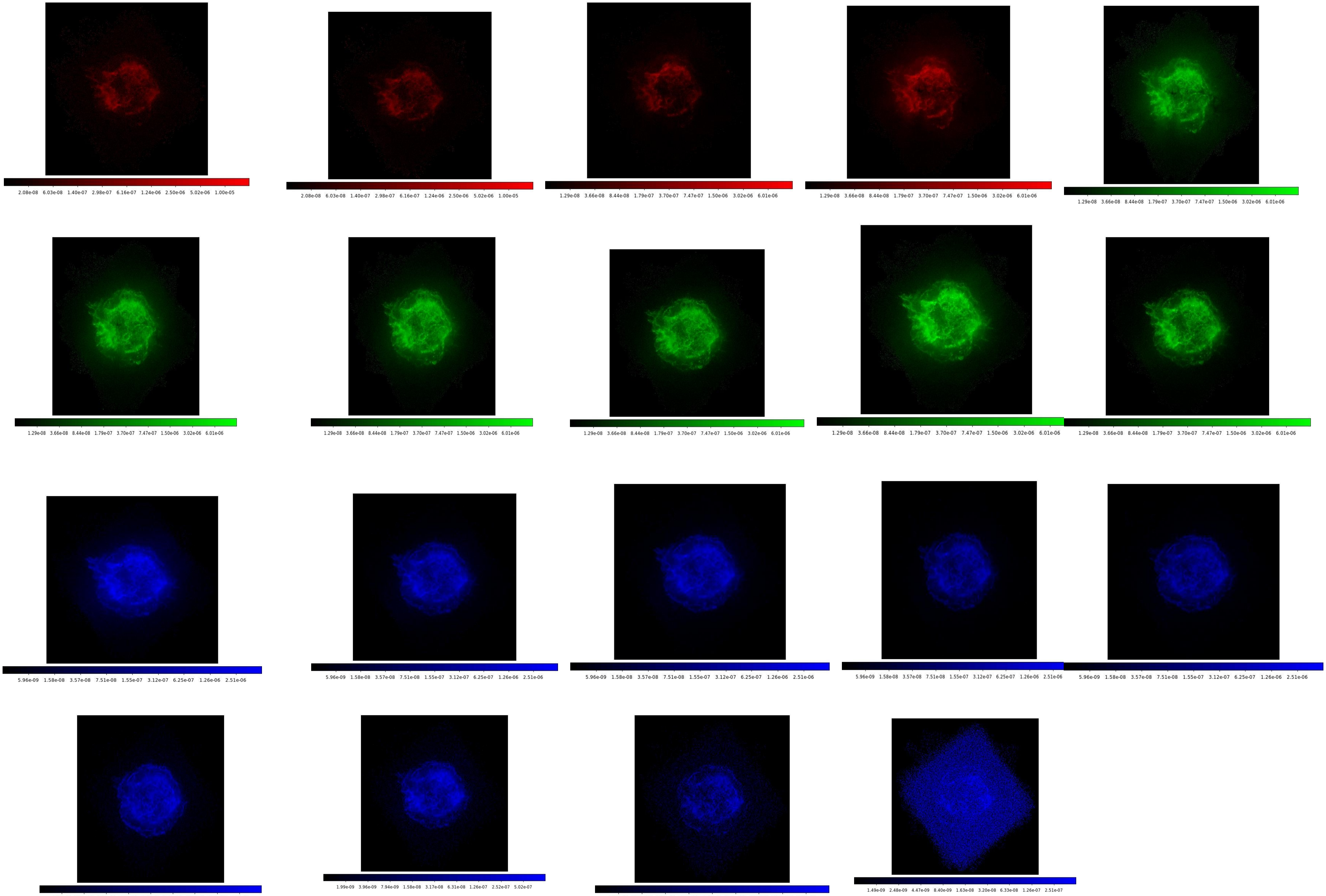


Camera sensitvity
- different for different colors

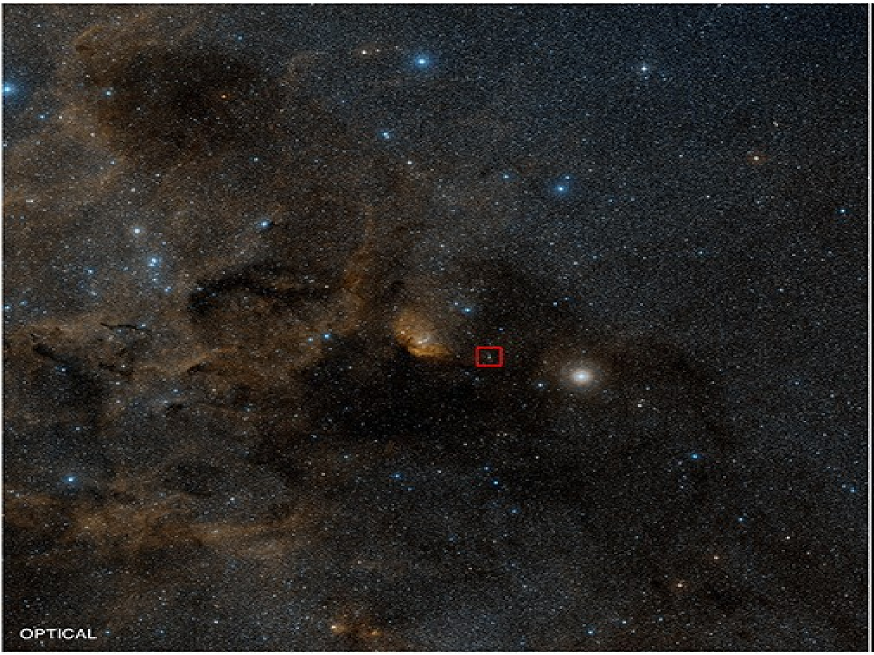
Also:

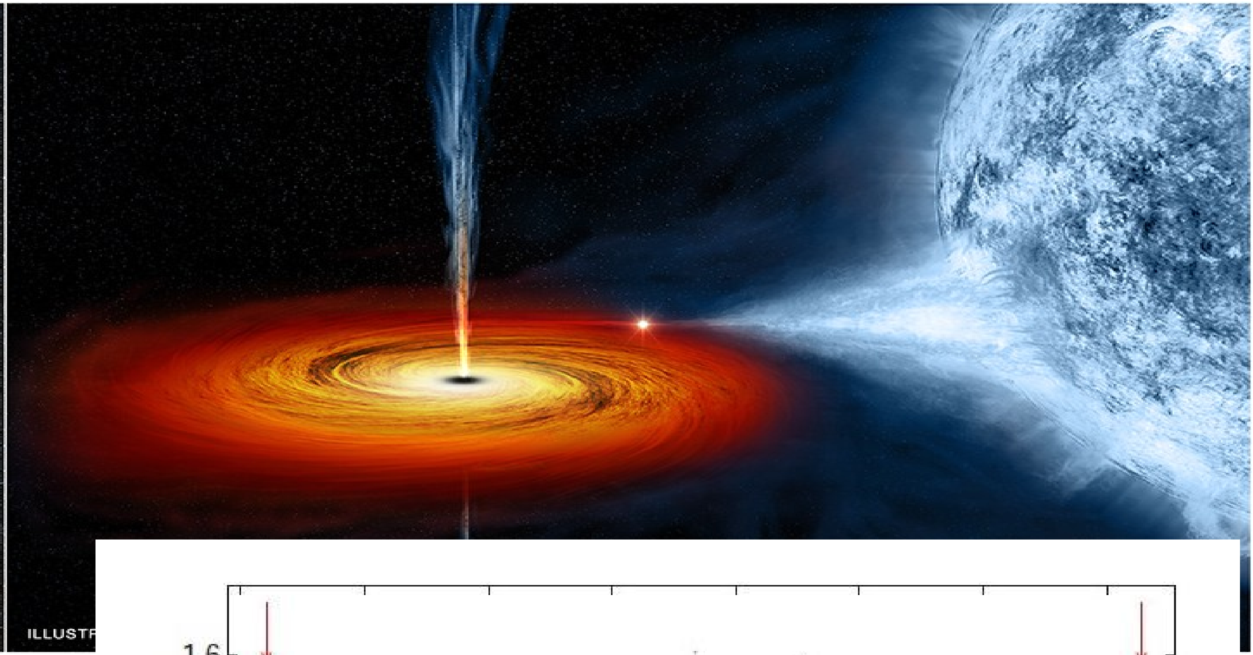
- sky background images
- bad pixel maps
- color 'gain' maps

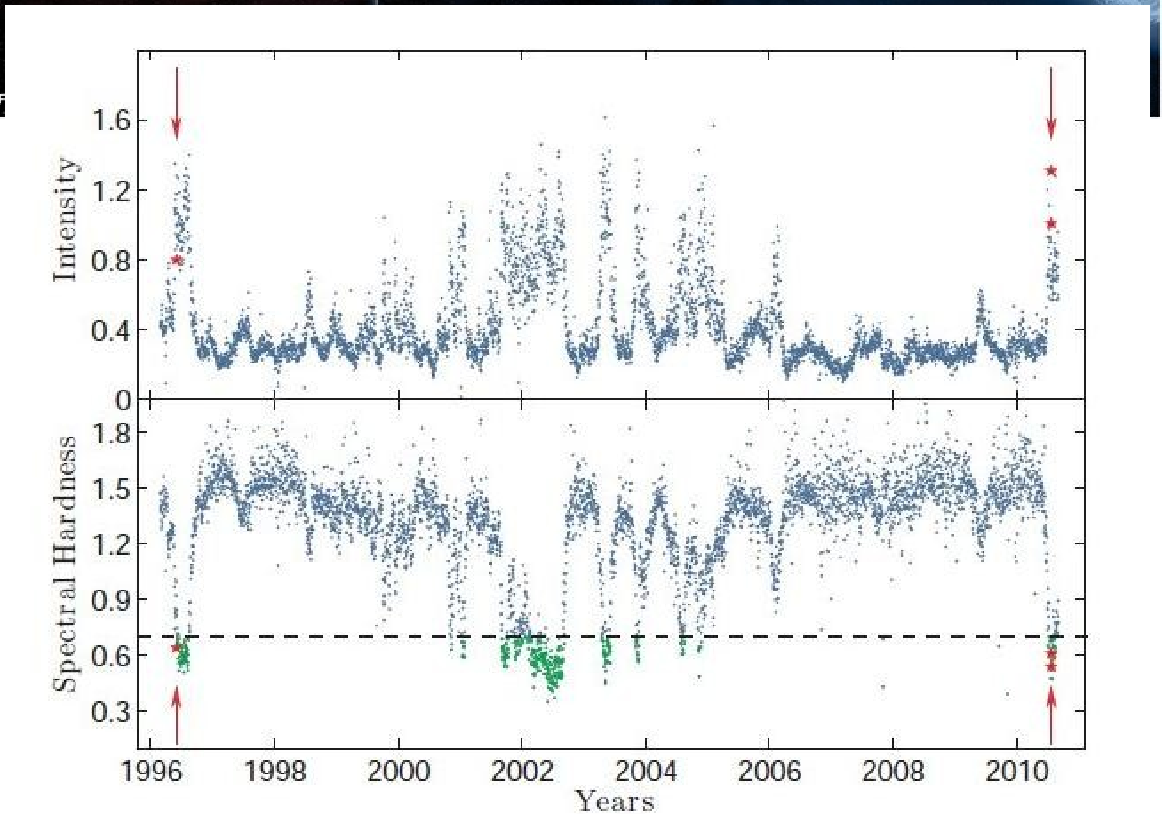
different problems with the camera, different calbration maps to fix them


2.08e-08 6.03e-08 1.40e-07 2.98e-07 6.16e-07 1.24e-06 2.50e-06 5.02e-06 1.00e-05
2.08e-08 6.03e-08 1.40e-07 2.98e-07 6.16e-07 1.24e-06 2.50e-06 5.02e-06 1.00e-05
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
1.29e-08 3.66e-08 8.44e-08 1.79e-07 3.70e-07 7.47e-07 1.50e-06 3.02e-06 6.01e-06
5.96e-09 1.58e-08 3.57e-08 7.51e-08 1.55e-07 3.12e-07 6.25e-07 1.26e-06 2.51e-06
5.96e-09 1.58e-08 3.57e-08 7.51e-08 1.55e-07 3.12e-07 6.25e-07 1.26e-06 2.51e-06
5.96e-09 1.58e-08 3.57e-08 7.51e-08 1.55e-07 3.12e-07 6.25e-07 1.26e-06 2.51e-06
5.96e-09 1.58e-08 3.57e-08 7.51e-08 1.55e-07 3.12e-07 6.25e-07 1.26e-06 2.51e-06
5.96e-09 1.58e-08 3.57e-08 7.51e-08 1.55e-07 3.12e-07 6.25e-07 1.26e-06 2.51e-06
1.99e-09 3.96e-09 7.94e-09 1.58e-08 3.17e-08 6.31e-08 1.26e-07 2.52e-07 5.02e-07
1.99e-09 3.96e-09 7.94e-09 1.58e-08 3.17e-08 6.31e-08 1.26e-07 2.52e-07 5.02e-07
1.99e-09 3.96e-09 7.94e-09 1.58e-08 3.17e-08 6.31e-08 1.26e-07 2.52e-07 5.02e-07
1.49e-09 2.48e-09 4.47e-09 8.40e-09 1.63e-08 3.20e-08 6.33e-08 1.26e-07 2.51e-07

Different Kinds of Data 2- Light Curves



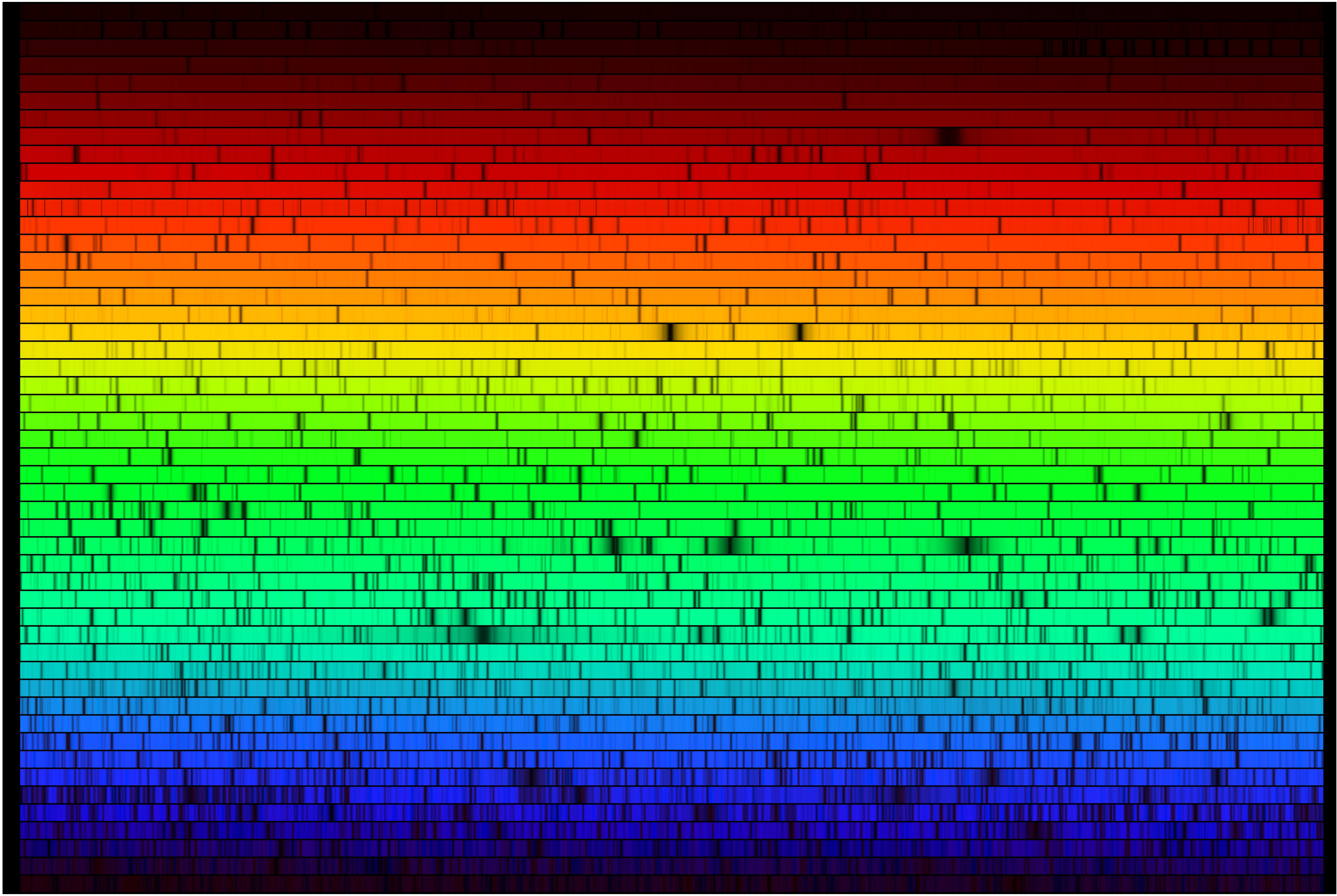




Cygnus X-1
A massive blue star slowly being eaten by its companion black hole
When the stream from the blue star hits the material swirling around the hole X-rays are produced

The Rossi XTE satellite monitored the brightness of Cyg X-1 over 14 years

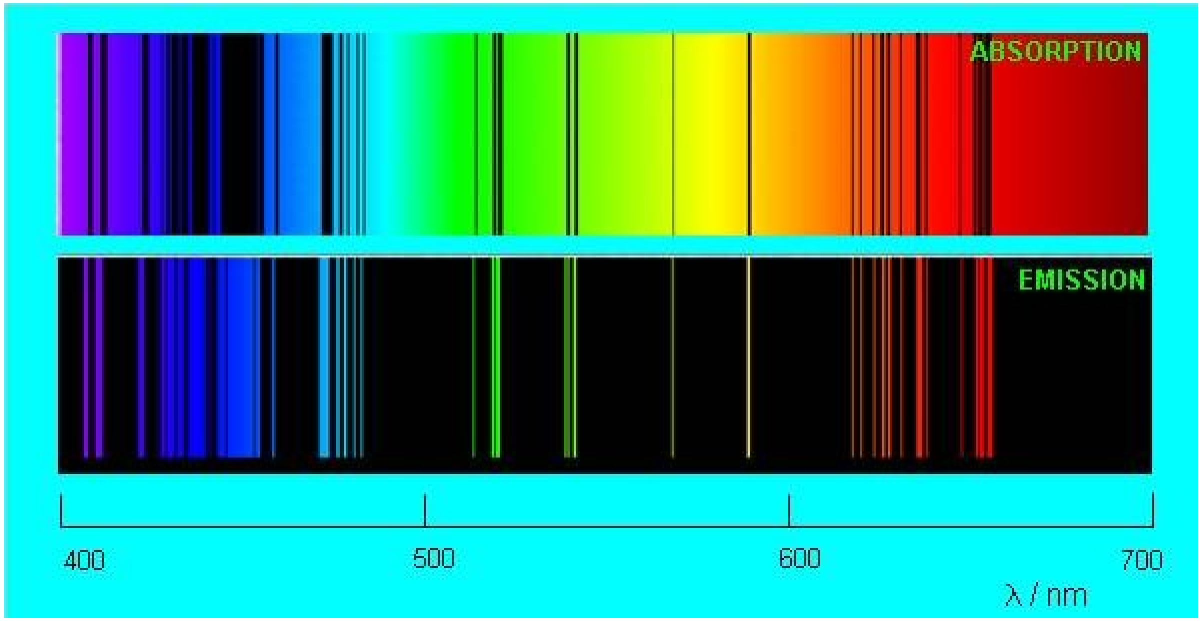
Data: Gou et al 2011, ApJ 742, 85



Data: Bob Kurucz et al (SAO); Image: Nigel Sharp. NOAO; Telescope: KPNO-McMath

Solar spectrum, 2960-13000 Angstroms

What we can learn from a spectrum:



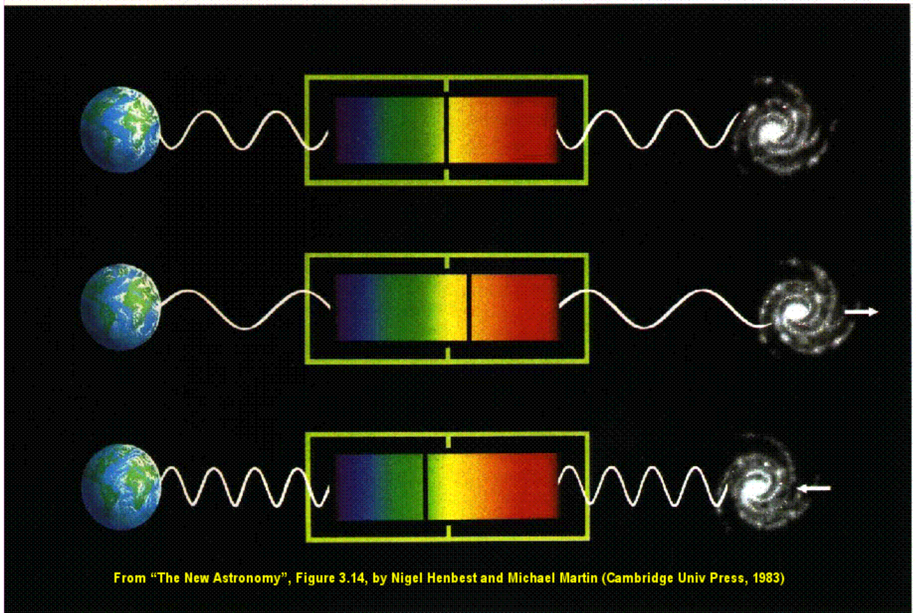
What is the light source made of?

- this is the "fingerprint' of sodium

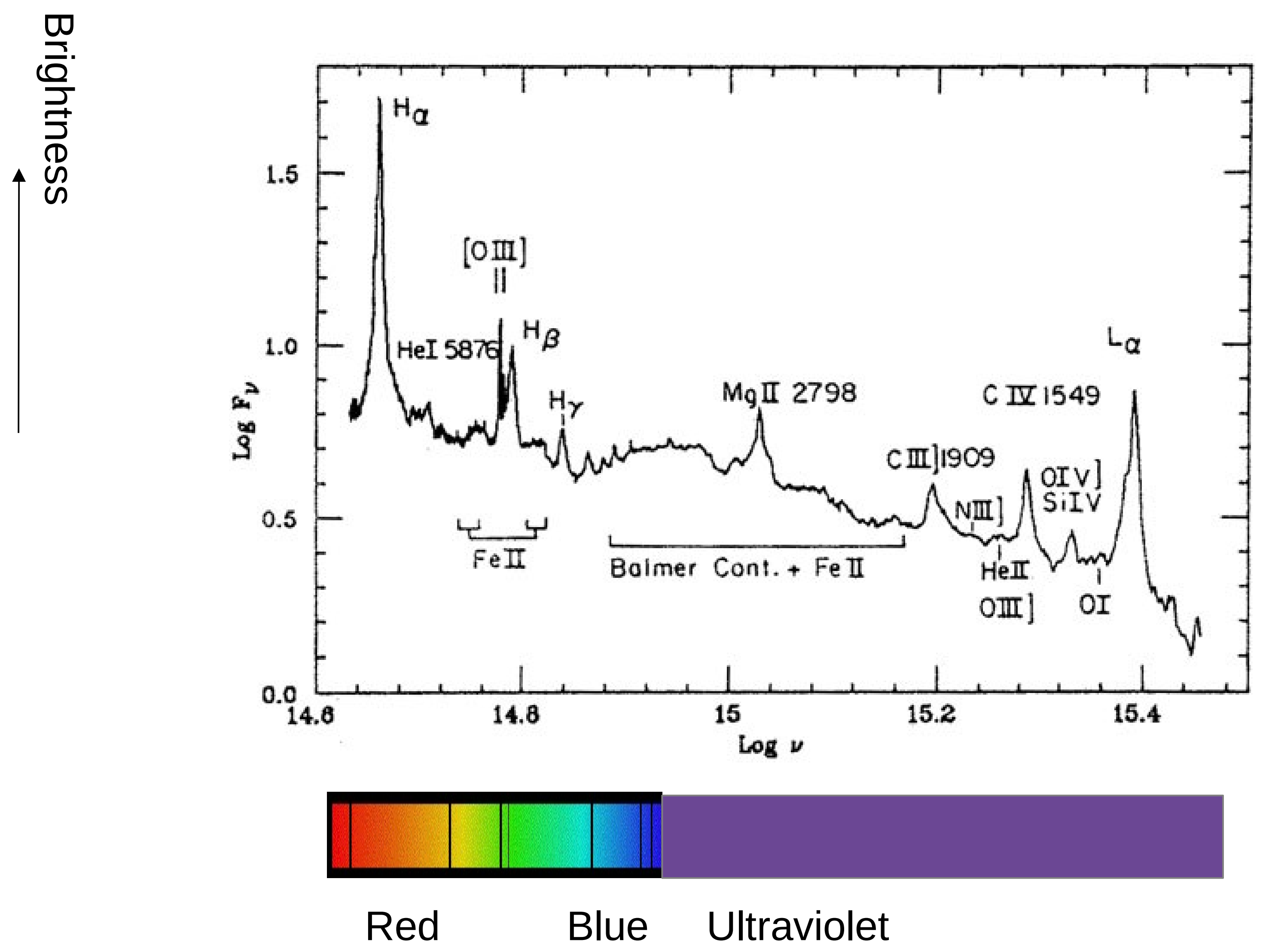
What are the physical conditions like?

- relative brightness and thickness of different lines indicates temperature and density

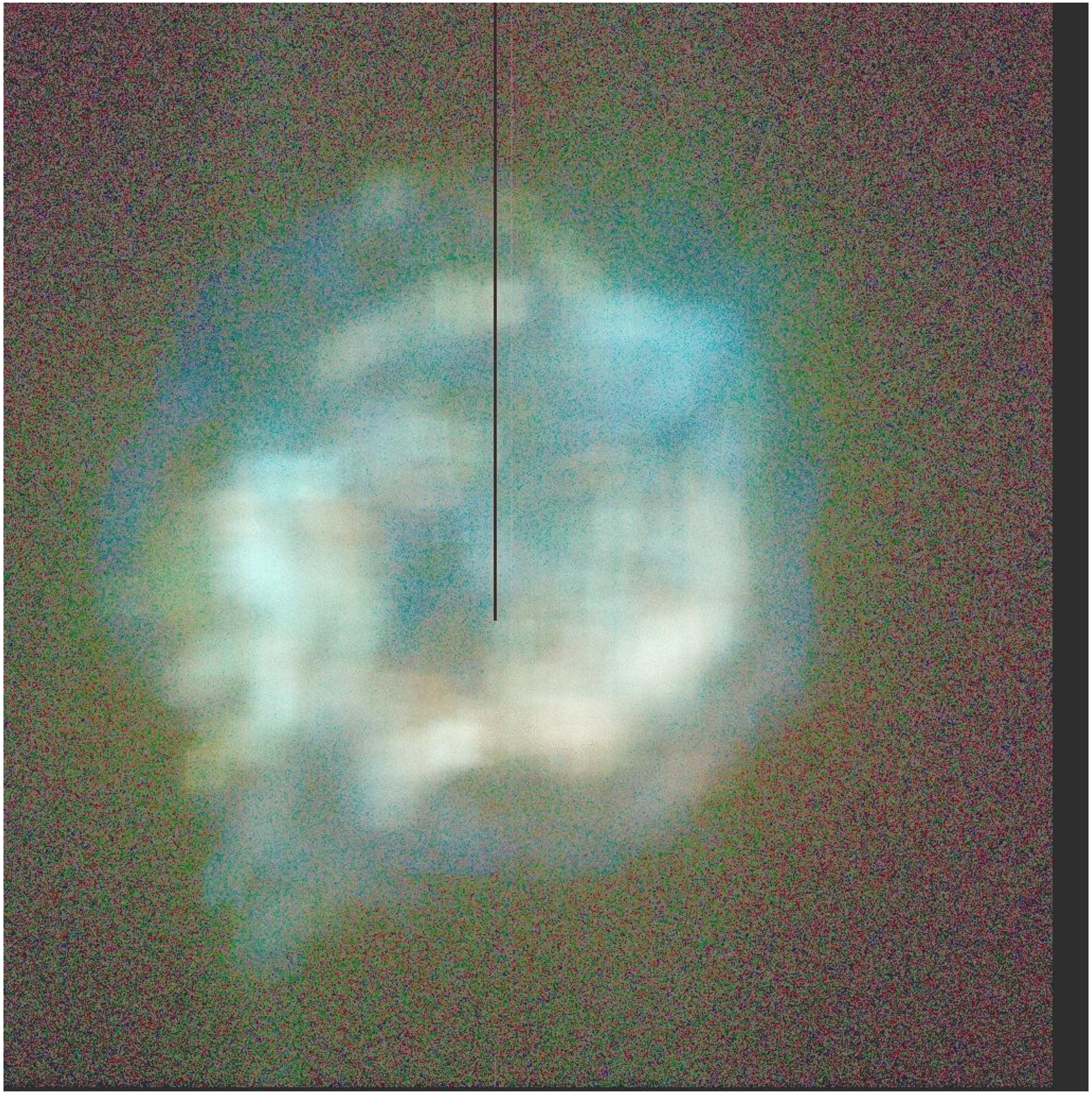
How fast is it moving?
"Doppler Shift" stretches or squeezes the spectrum: read off the speed



Different Kinds of Data 3 - Spectra



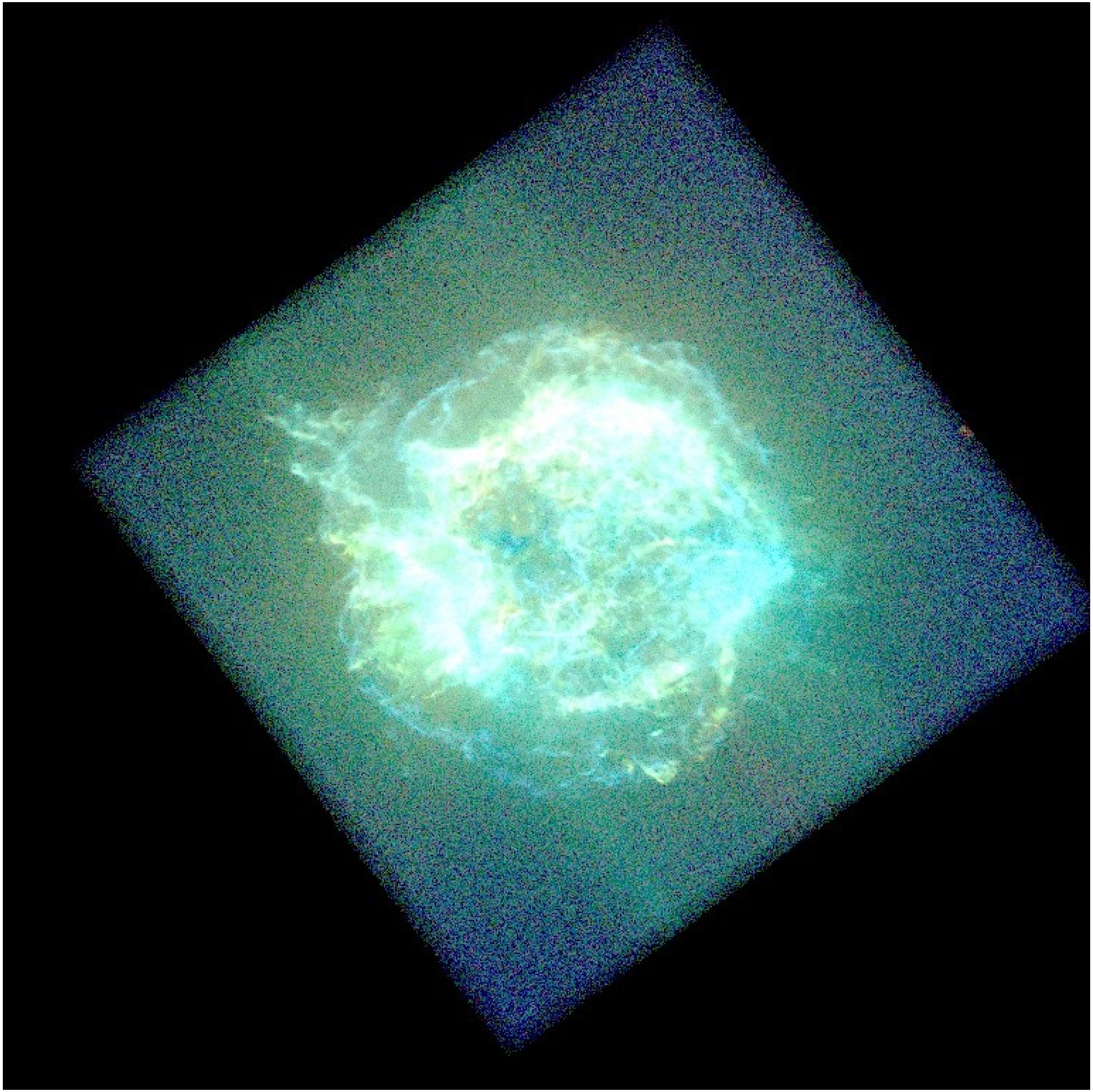
Data; J. Baldwin via NED IPAC (Caltech)



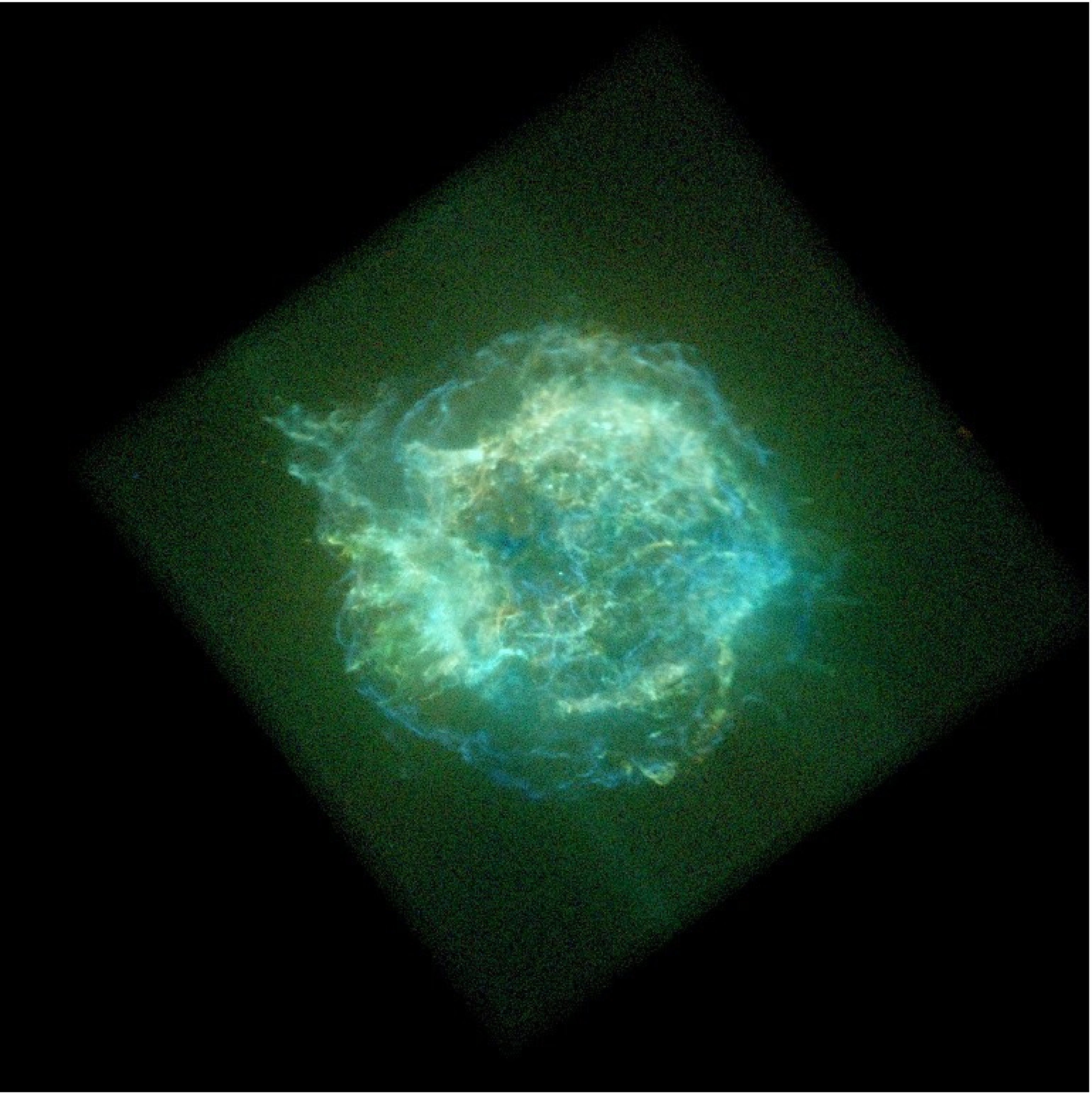
Observation 114
S. Holt, 2000
Downloaded from public archive for reanalysis

Our first image, as the spacecraft takes it

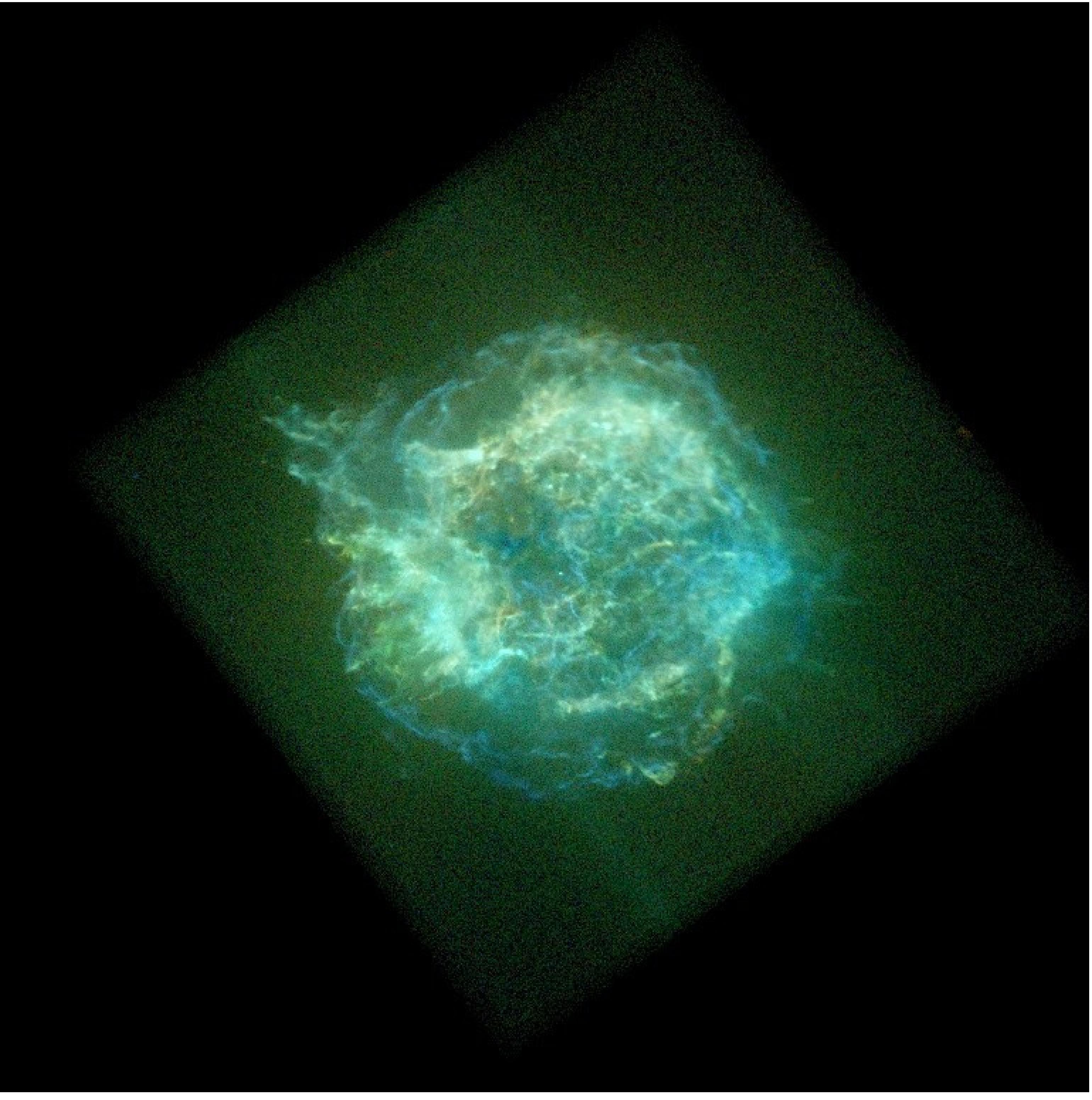
wrong (and unknown) way up
blurry
lots of background noise



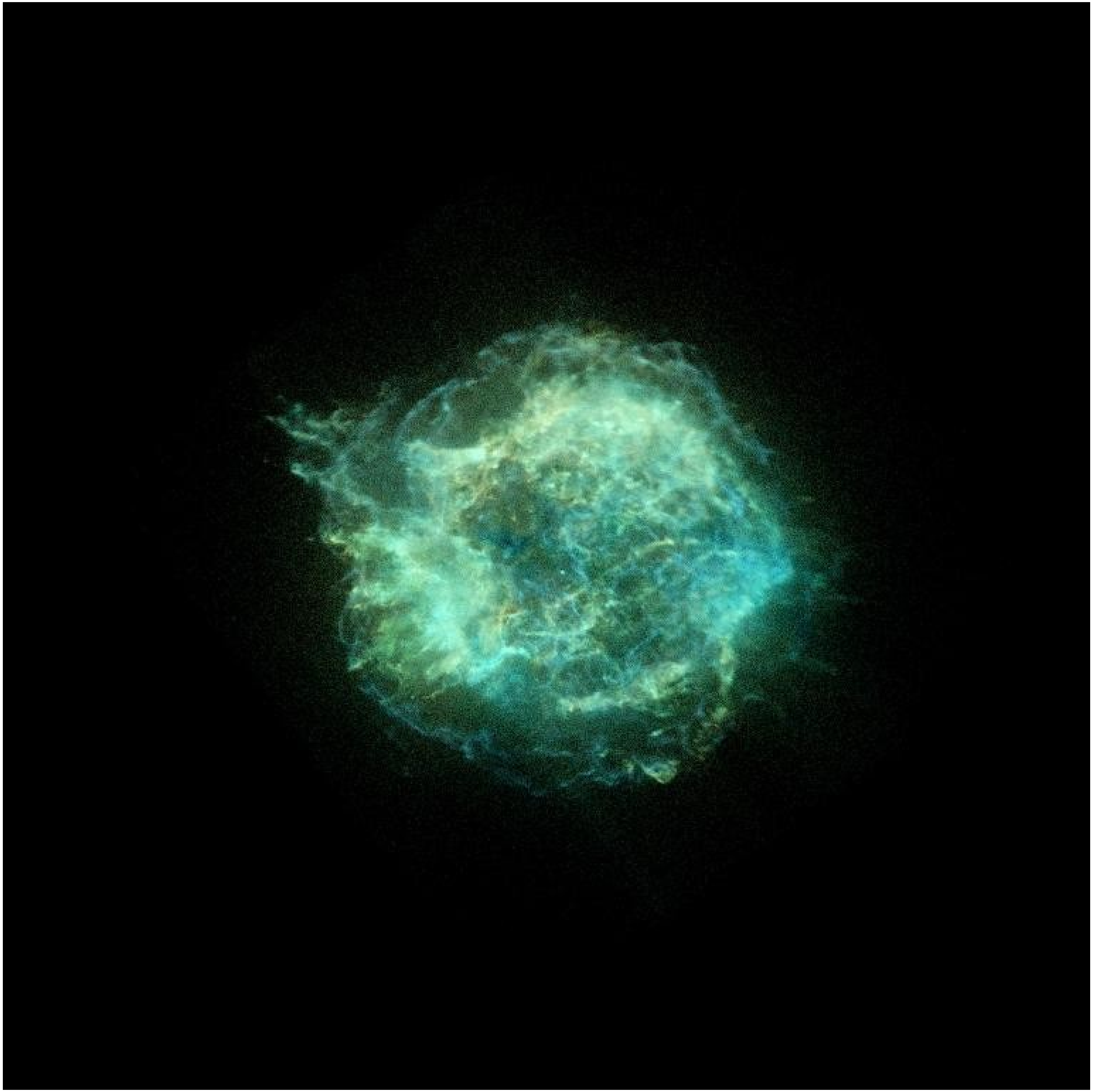
Correct for telescope motion



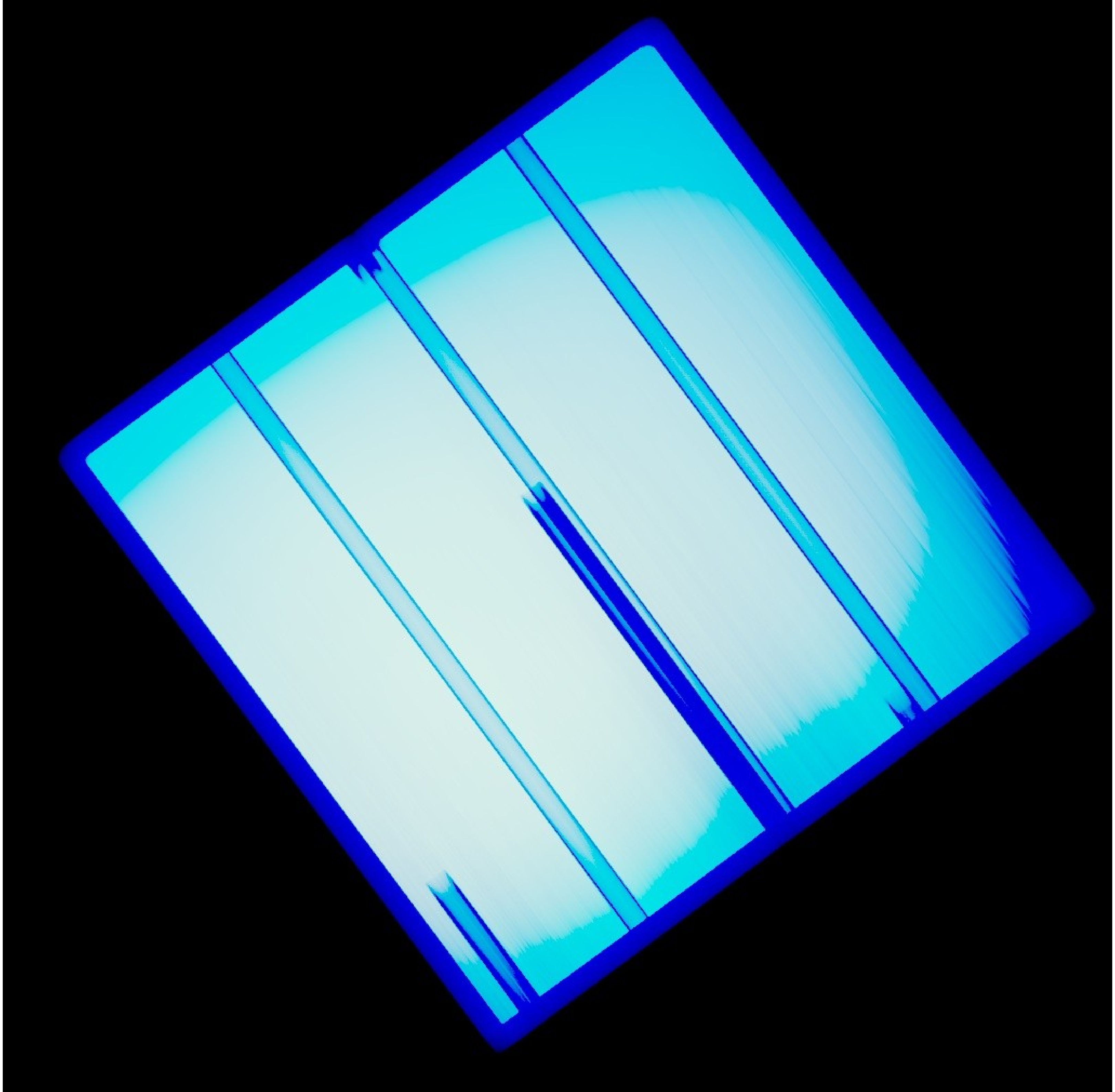
Remove bad
energy ranges



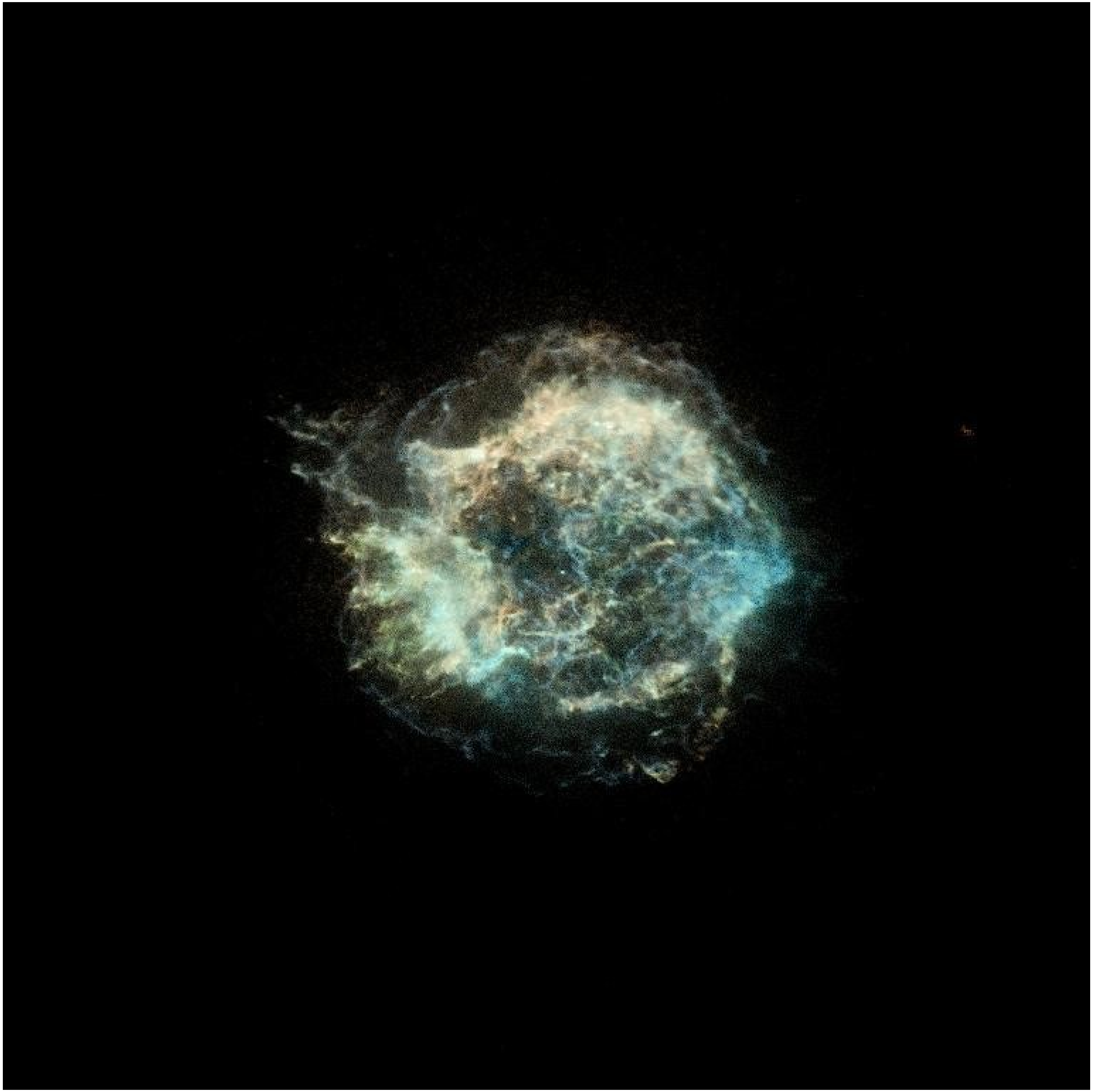
Remove bad
energy ranges



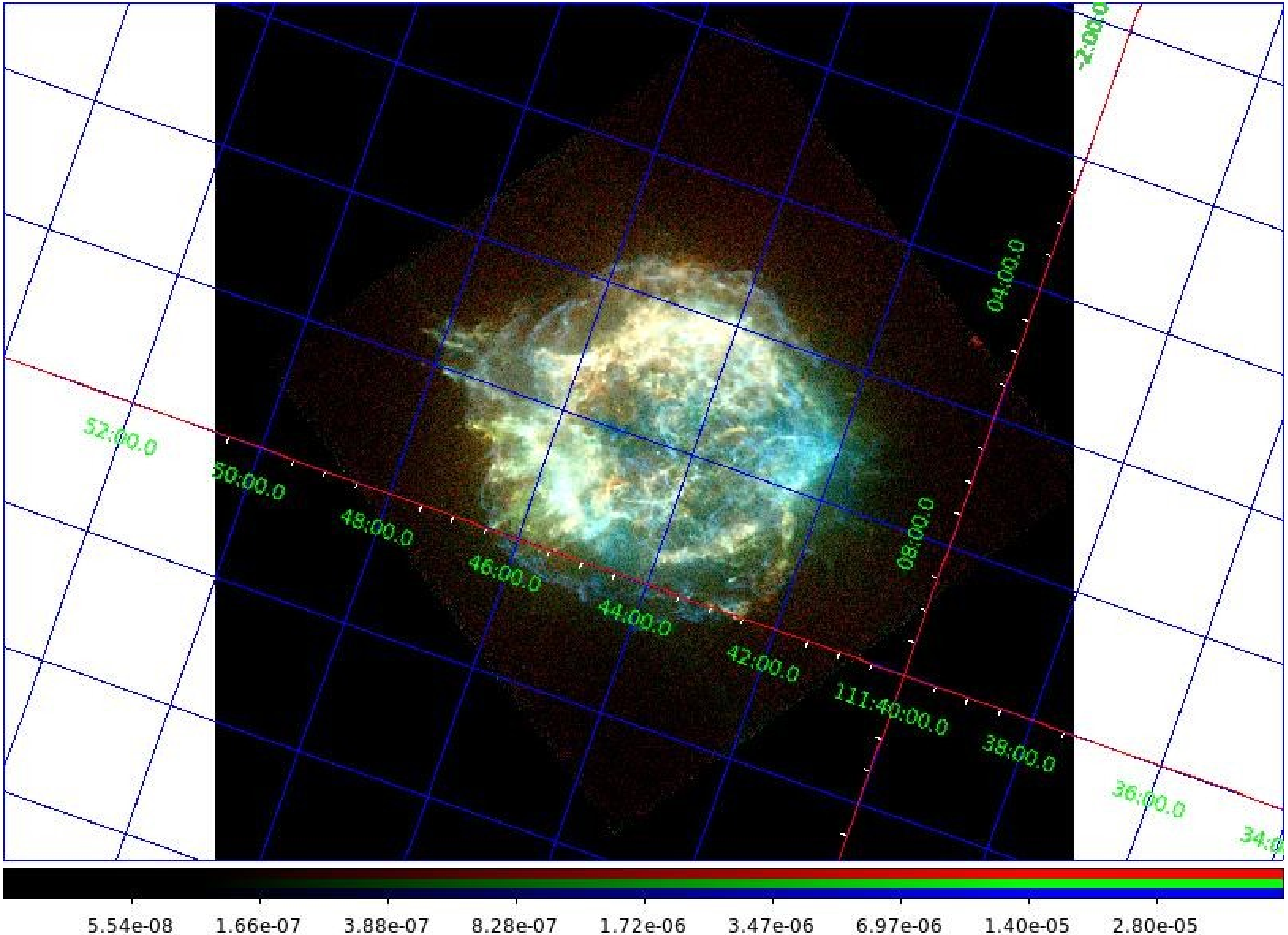
Remove bad 'grades'
(probable cosmic rays),
bad time intervals
(solar flares)



Correct for detector sensitivity

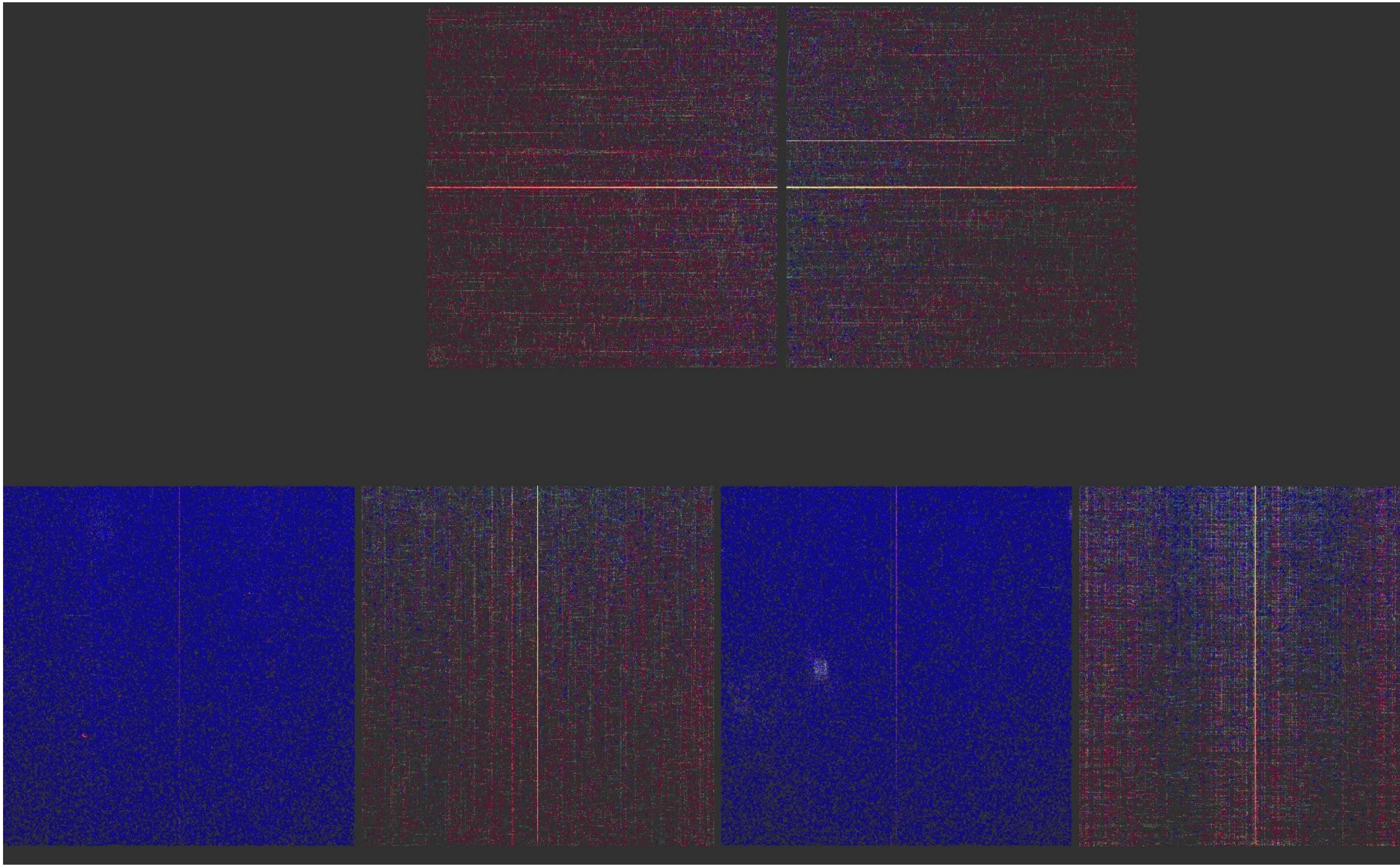


Correct for
detector
sensitivity

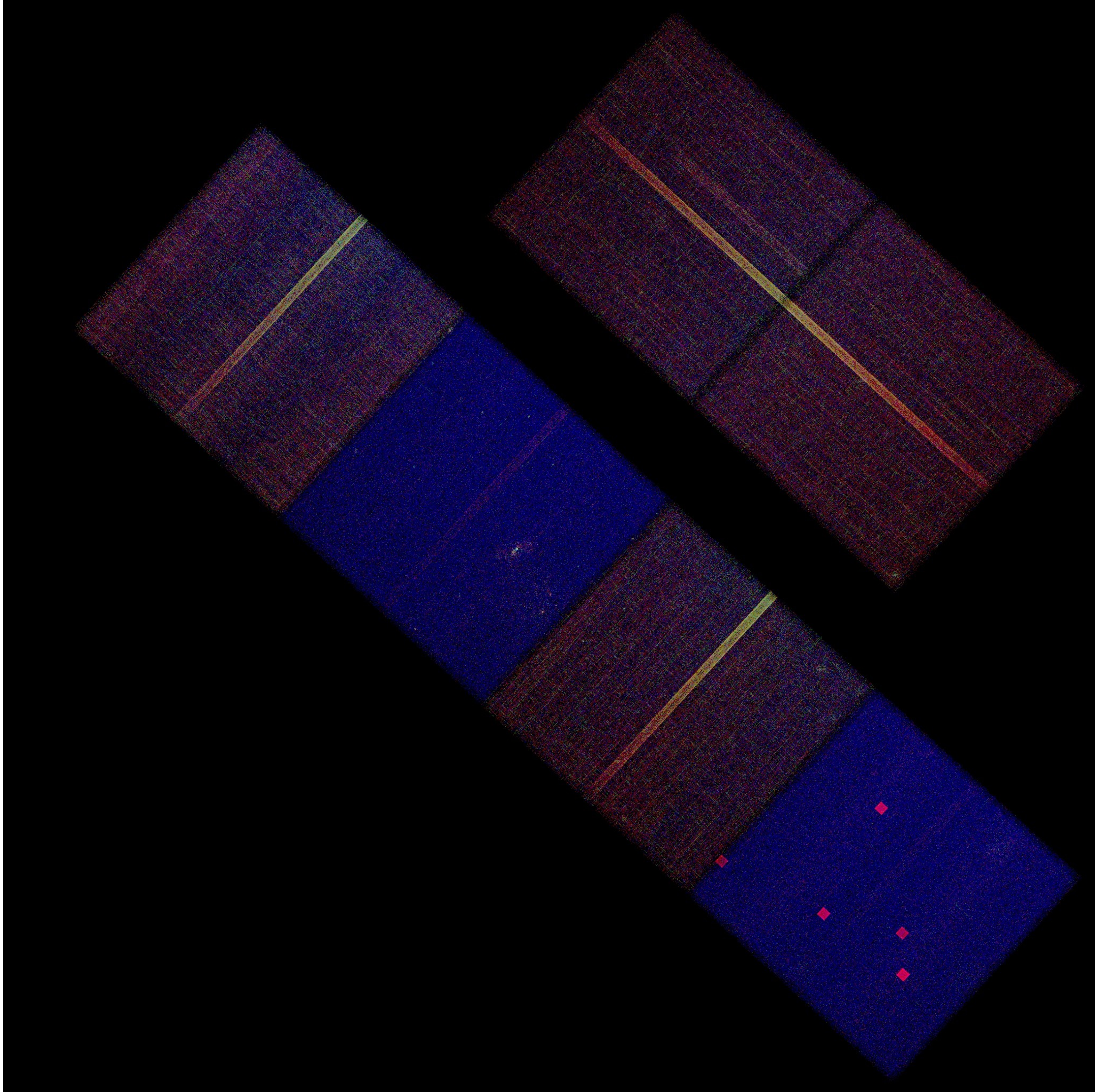


Latitude
and
longitude
grid
etc.

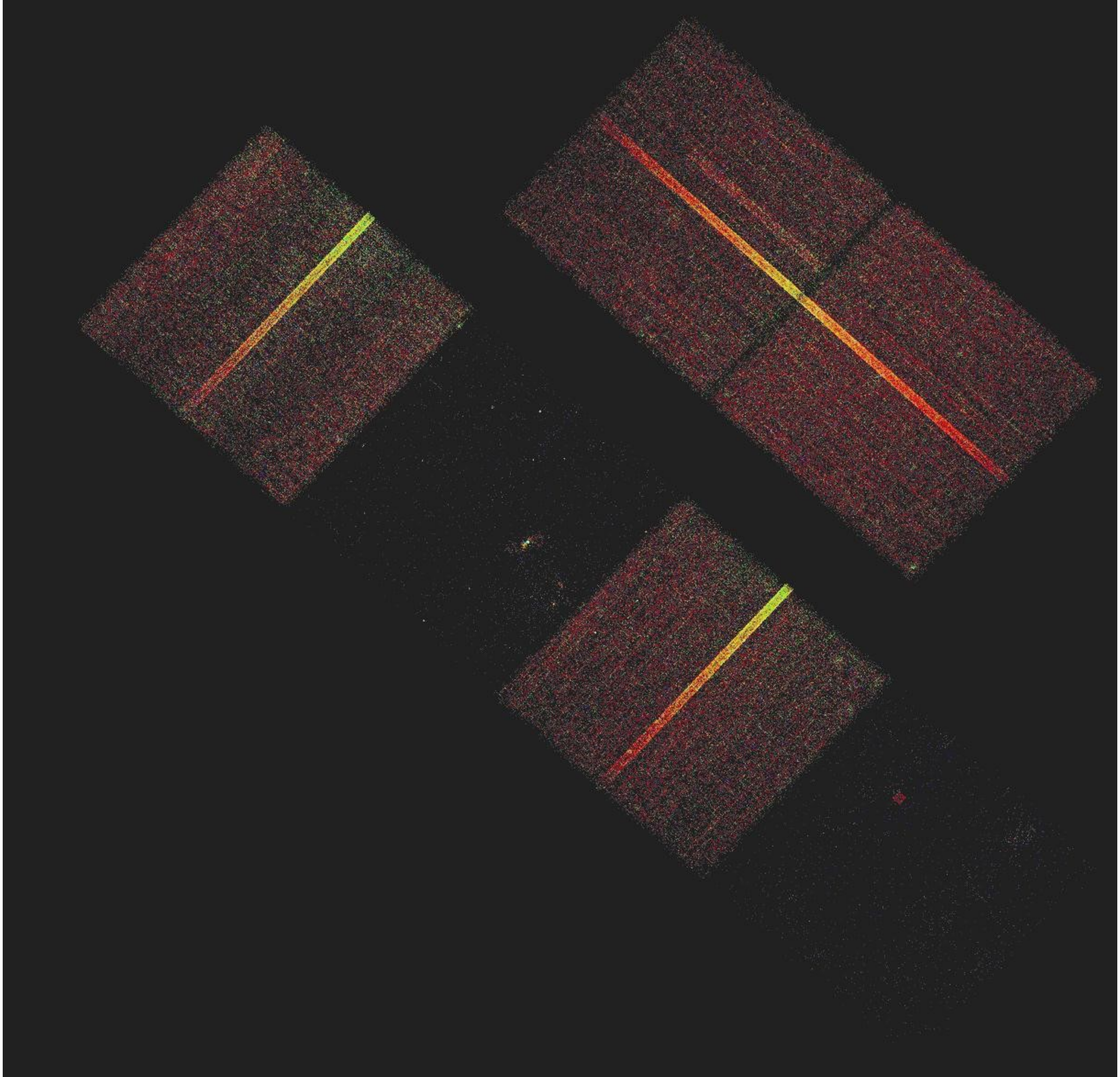
And now, to the constellation Serpens...



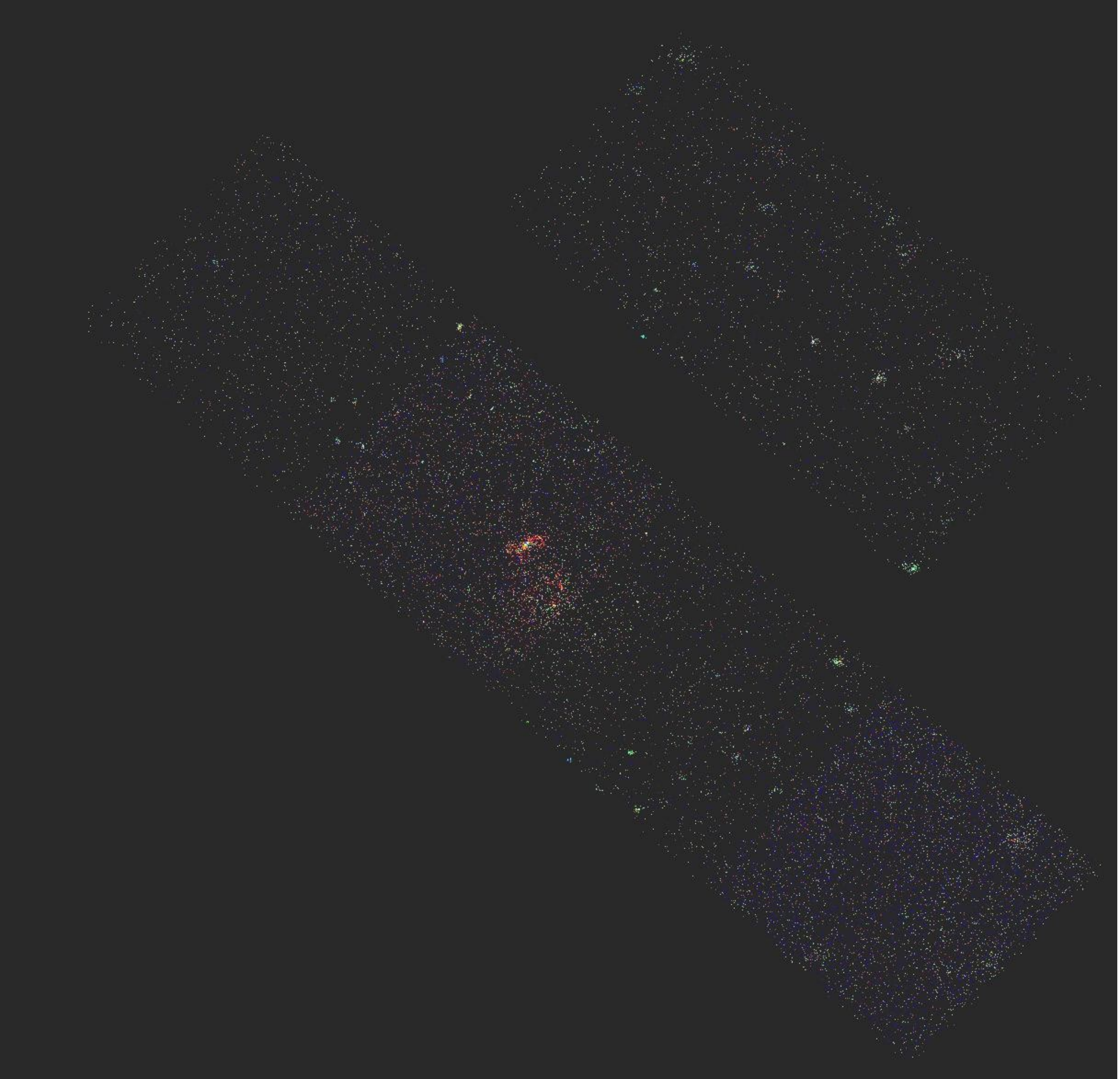
Our first image, as the spacecraft takes it
wrong way up
blurry
lots of background noise, bad columns - not obvious if we have got anything!



Corrected
for
telescope
motion



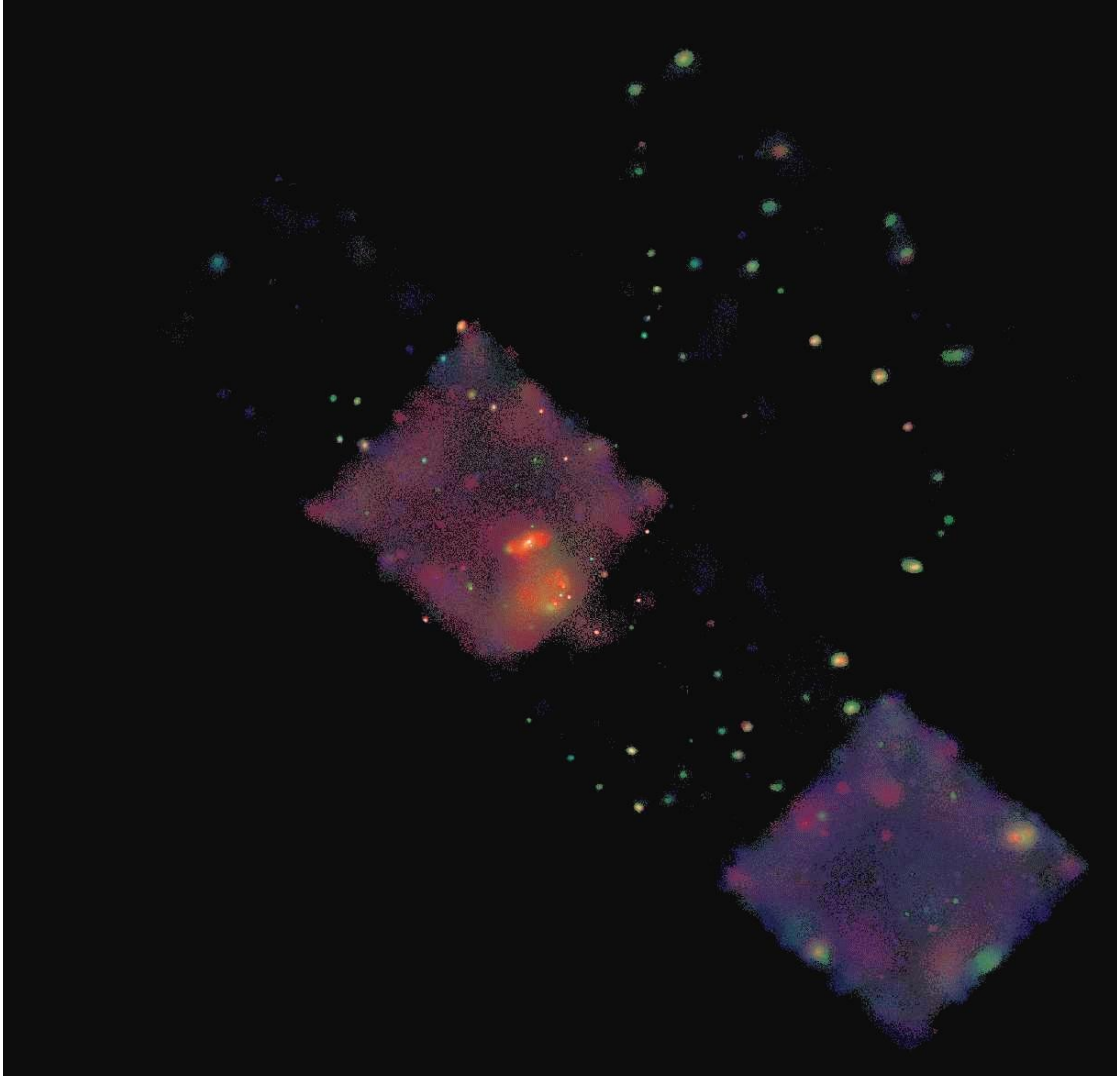
Clip out bad color ranges – removes some bad pixels and some of the cosmic ray background



Get rid of bad columns, pixels

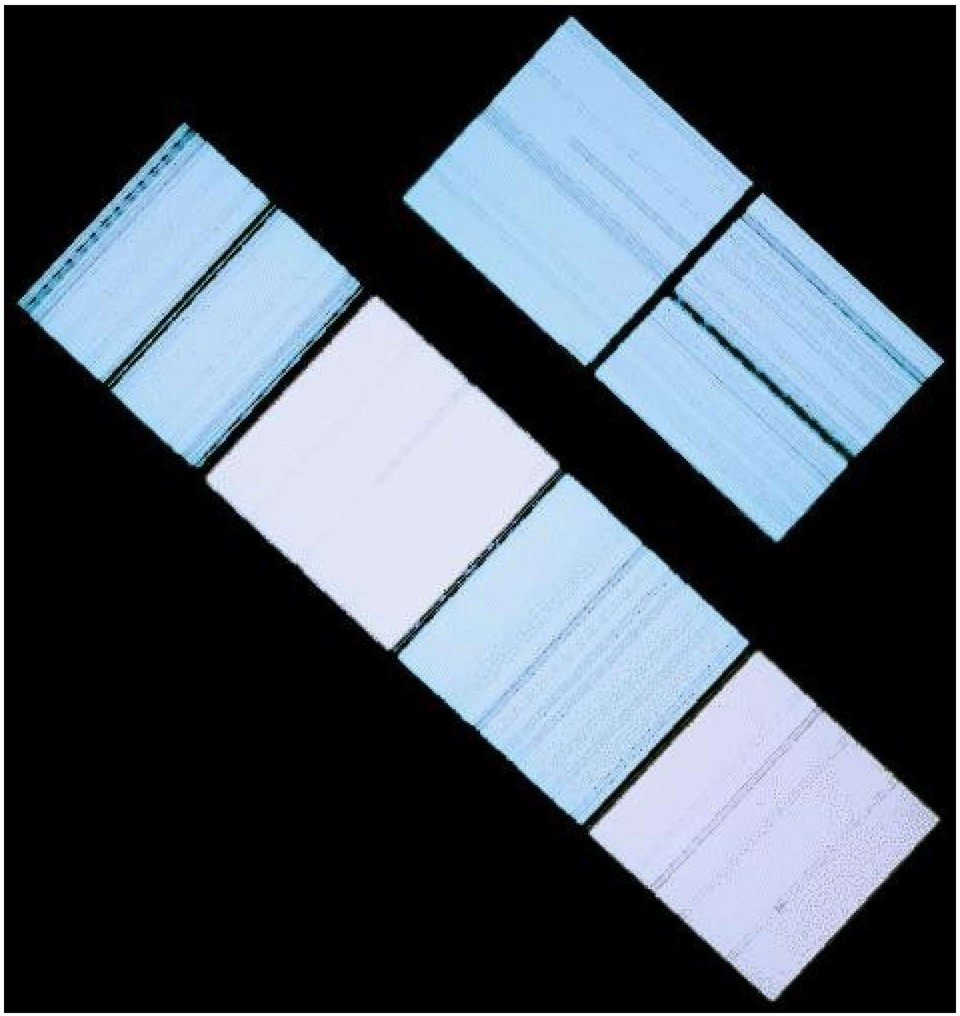
Get rid of some more cosmic rays (“grade filtering”)

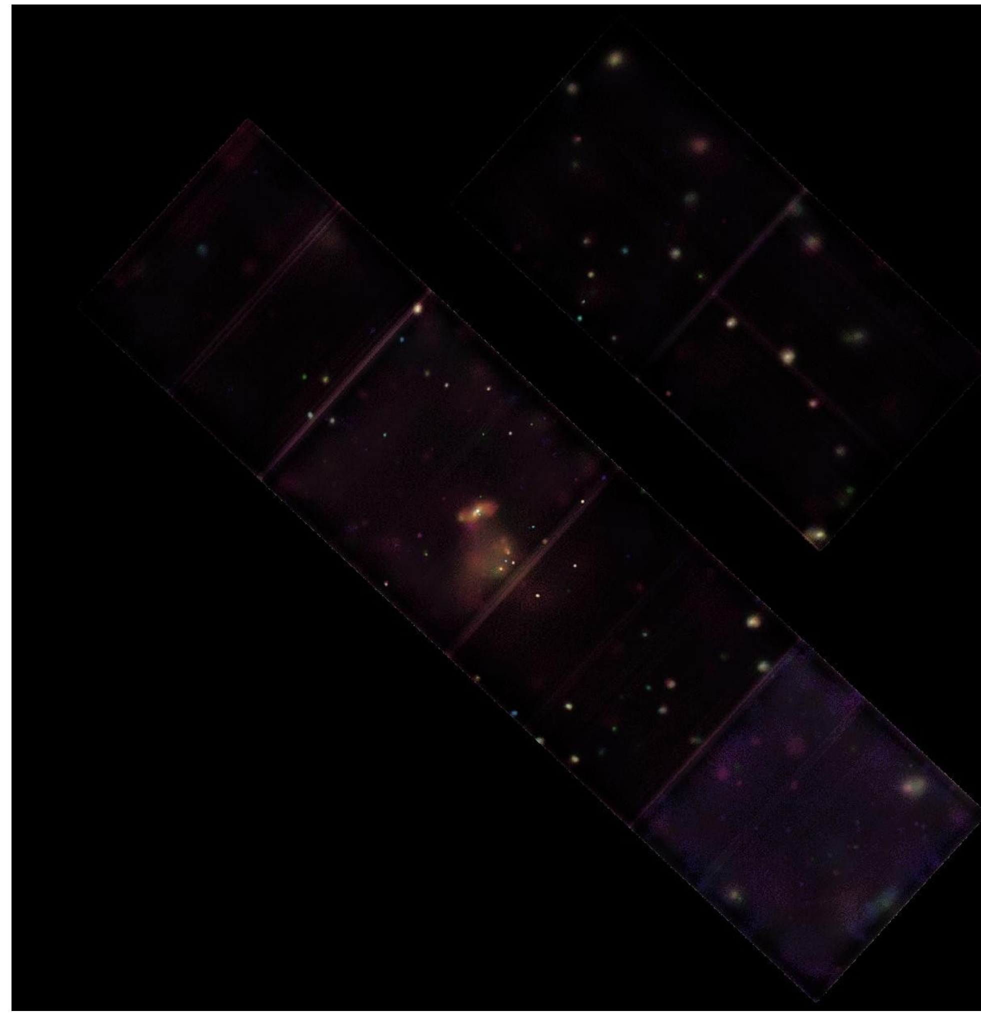
Get rid of bad time intervals

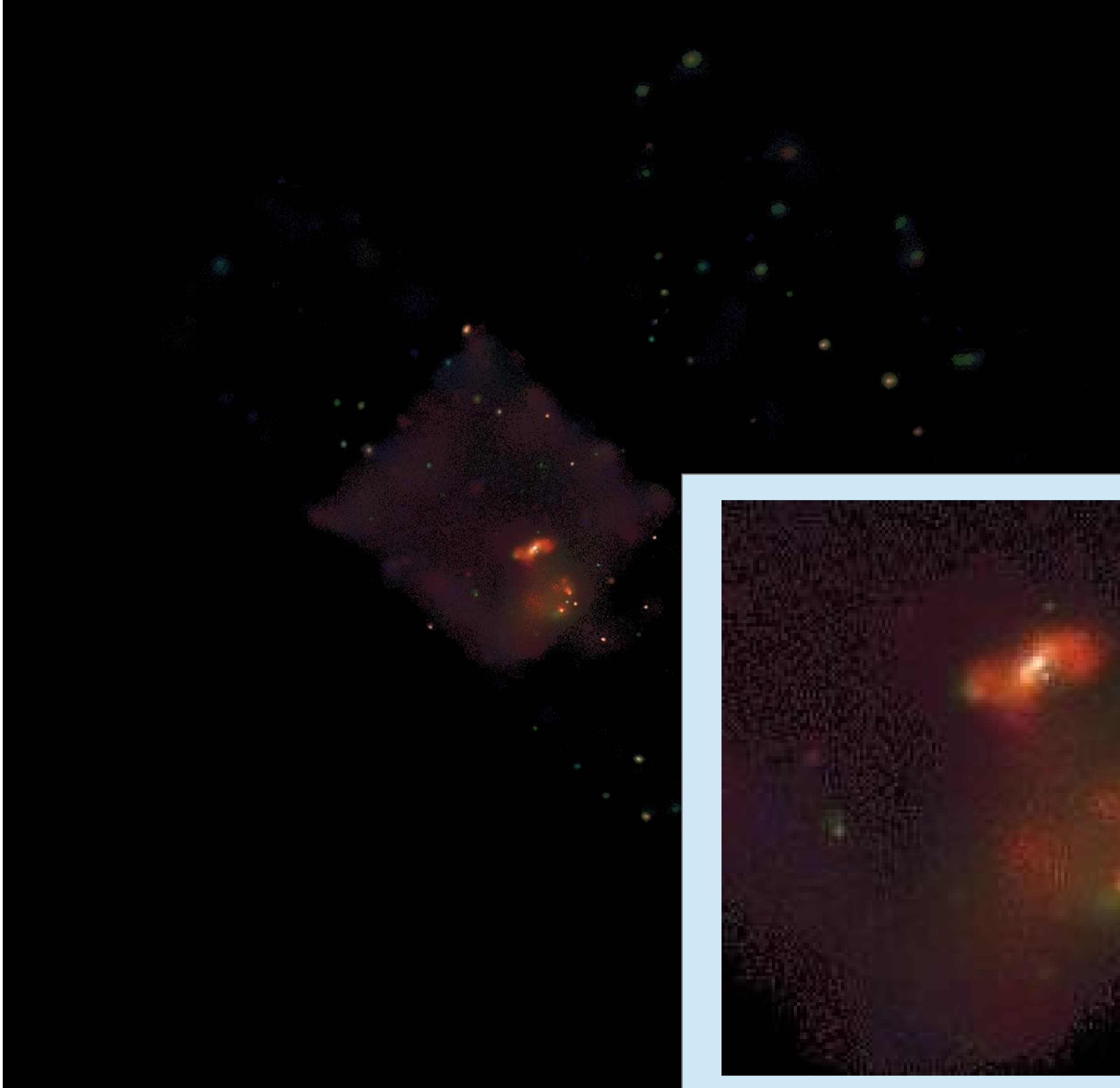


Adaptive Smoothing: desperately try and compensate for the fact that we have only a handful of X-rays.

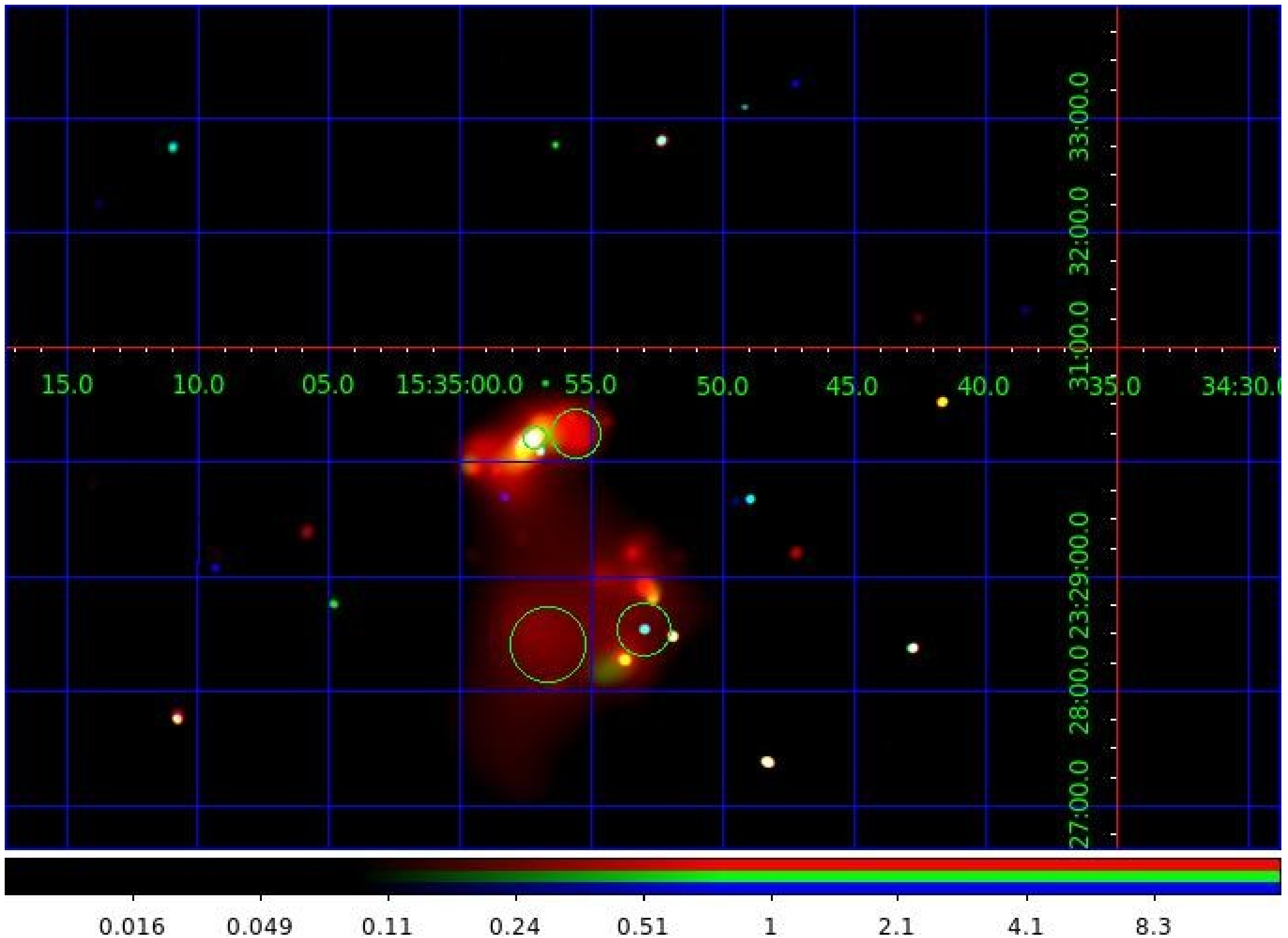
Now we correct for the varying exposure across the image caused by imperfections in the camera and the observing process







Isolate the region of interest



Put on the coordinate grid
Find the galaxies you are interested in
Define circles to isolate them
Now we can look at the light from those regions and play the spectrum game:
 - how hot? what's it made of? How fast is it moving?

Papers: Clements, McDowell, et al 2002 – Astrophysical Journal 581, 974
McDowell, Clements, Lamb et al 2003 – Astrophysical Journal 591, 154

Part 3
Astronomy Data in the New Millenium

Historical Digression

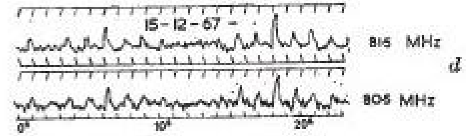


Fig. 1. *a*, A record of the pulsating radio source in strong signal conditions (receiver time constant 0·1 s). Full scale deflexion corresponds to

In the 1960s it was almost all done by hand...
- photographic plates
- light tables

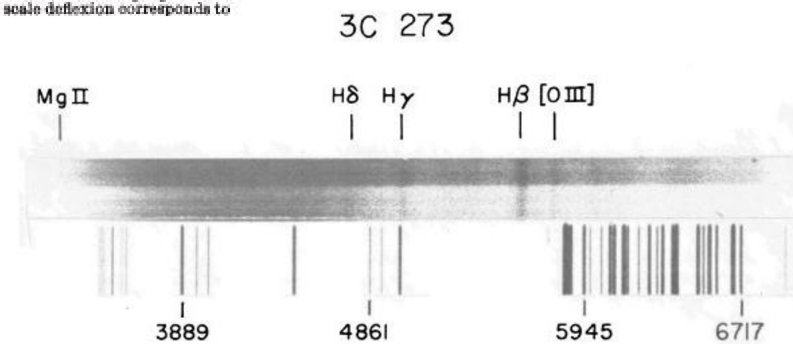


Fig. 2.—Spectrum of the quasi-stellar object 3C 273B, 400 Å/mm original, 103a-F, January 23, 1963. The comparison spectrum is H + He + Ne. Exposure over the upper half of slit was three times that over the lower half. Redshifted emission lines of H and [O III] are indicated; also the barely visible line of Mg II, confirmed on denser exposures.

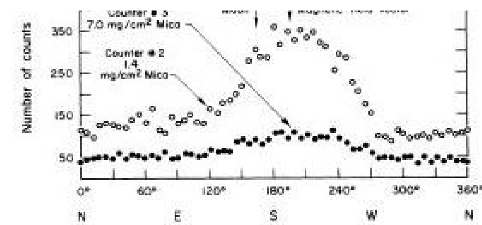


FIG. 1. Number of counts versus azimuth angle. The numbers represent counts accumulated in 350 seconds in each 6° angular interval.

even lower back-lobe level.

From a combination of the above, we compute the remaining unaccounted-for antenna temperature to be 3.5° ± 1.0° K at 4080 Mc/s. In connection with this result it should be noted that DeGrasse *et al*. (1959) and Ohm (1961) give total system temperatures at

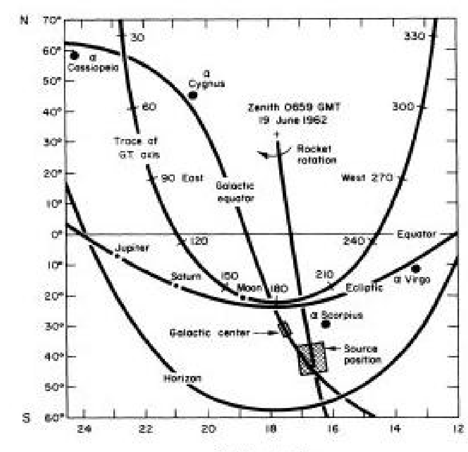
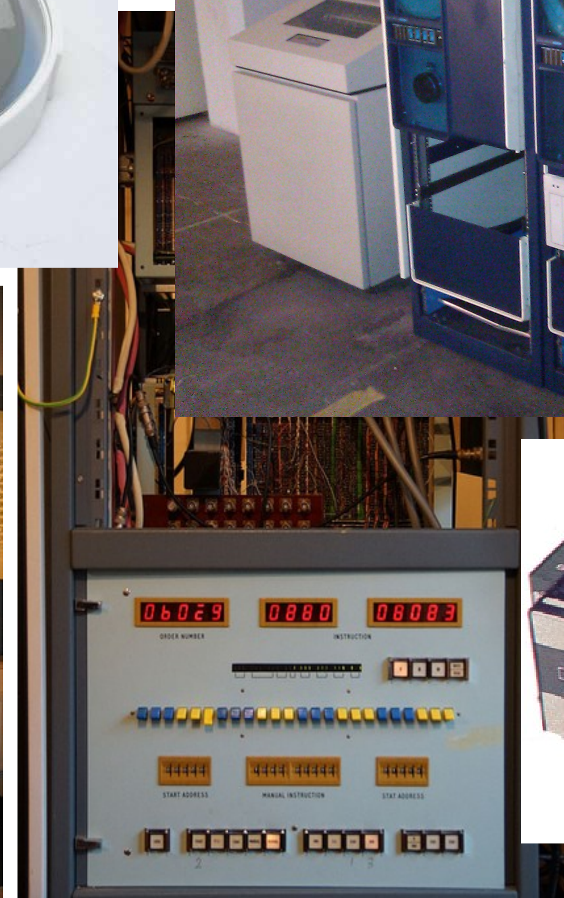


FIG. 2. Chart showing the portion of sky explored by the counters.

In the 1970s we introduced computer controlled telescopes
Data was written to magnetic tape, or photographic plates digitized by primitive scanners

Each telescope and camera had its OWN software – and was usually used by only one or a few astronomers. The software often only ran on ONE computer (not one TYPE of computer!)



In the 1980s the first generic software systems appeared

FIGARO by Keith Shortridge (1982, Caltech then AAO)

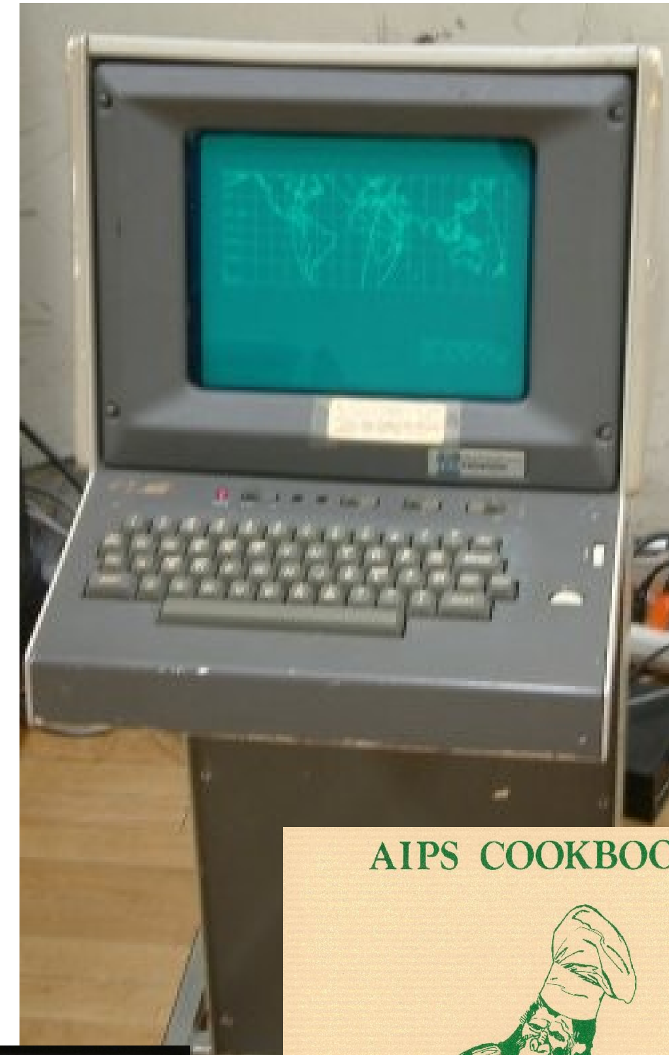
IRAF by Doug Tody (1984 at NOAO)

AIPS by Eric Greisen et al (1980 at NRAO) for radio astronomy

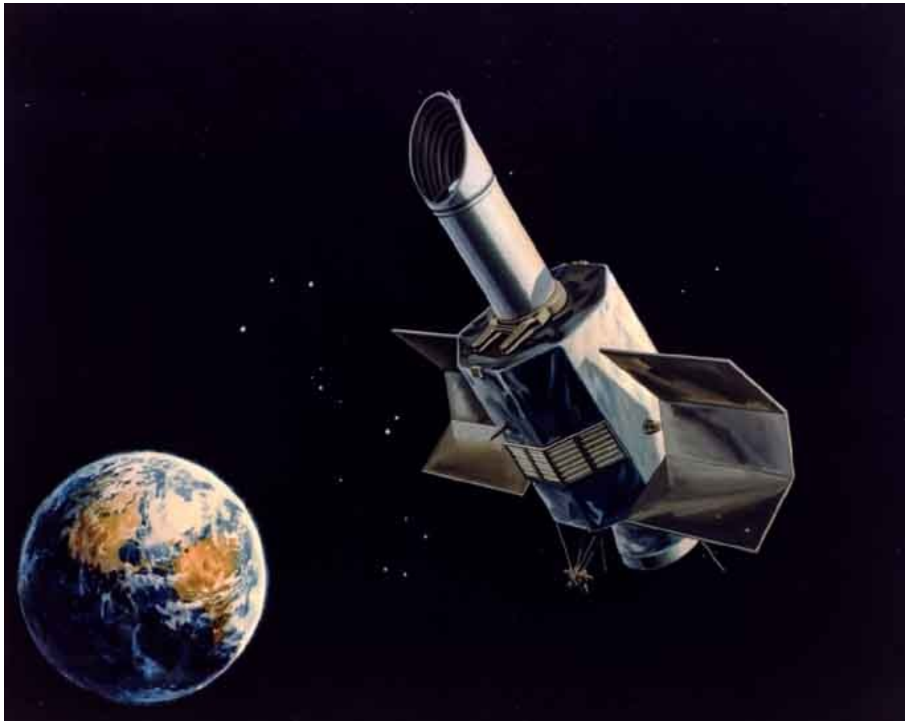
XSPEC by Rick Shafer, Keith Arnaud (c.1986) for X-rays

IDL (David Stern, commercial package) especially for analysis of IUE satellite data, late 1970s

Now astronomers could spend time using software instead of writing it
(Imagine if every time you needed to write a letter, you first had to invent your own equivalent of Word..)



Another crucial change in the late 1970s and early 1980s:
the public archive and the beginnings of the move to open source/open data



HEAO-2 1978-80
IUE 1978-1996

With HEAO-2 (Einstein) and IUE, astronomers ("Guest Observers") across the world could request data from the telescope and come to mission control to analyse it
This drove us to standard 'data products' and formats, and user-friendly software

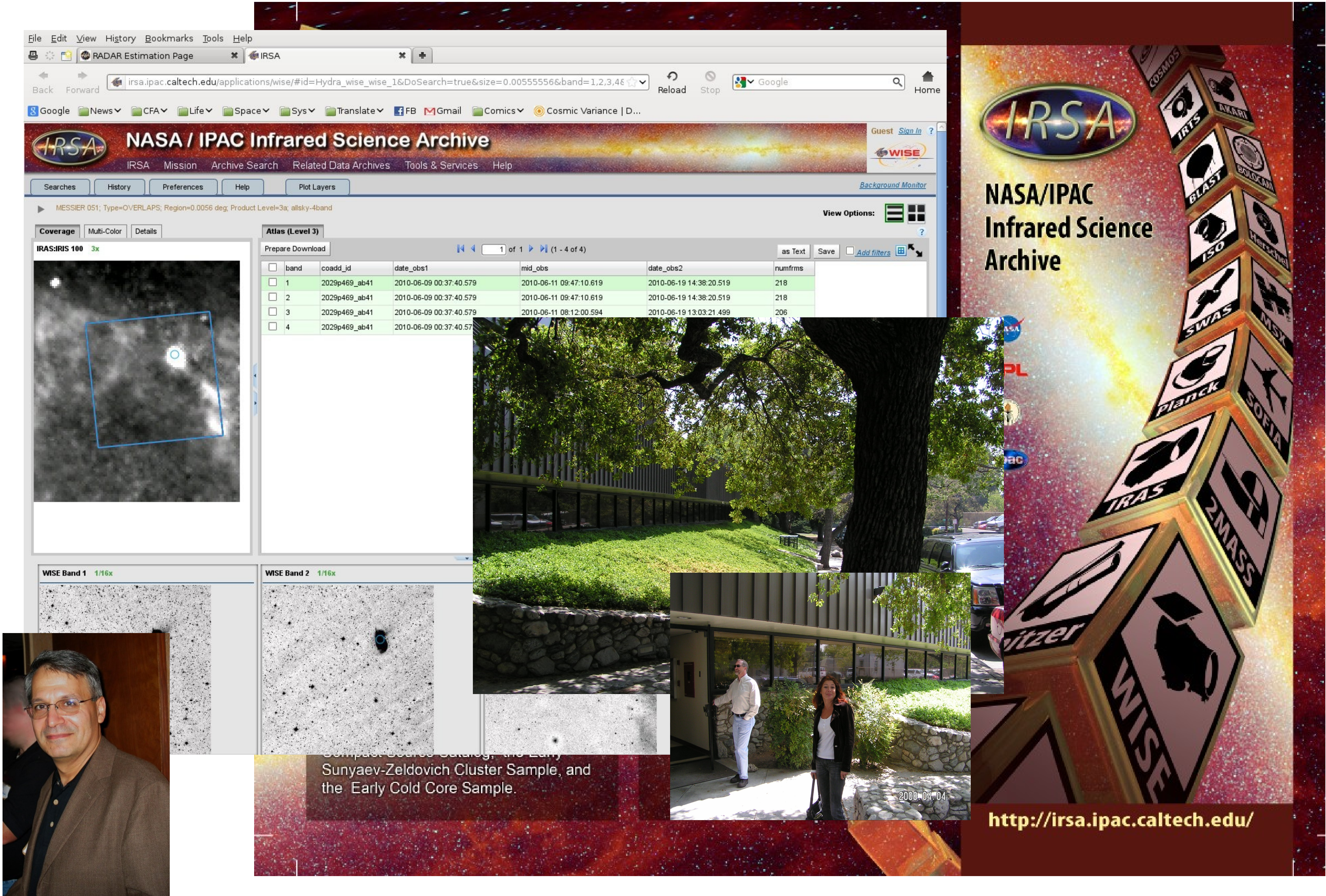
Public data archives and archival research

- in the late 1980s and early 1990s it became practical to download science data over the internet.
- That forced us to be a lot more detailed in describing what the data is, so that astronomers who didn't make the observation can do something with it without making mistakes
- We even went back and read the old data tapes from old missions – gathering mold in warehouses – reformatted them and wrote the files to CD-ROM, put them online
- Expertise for different wavebands ended up in special centers

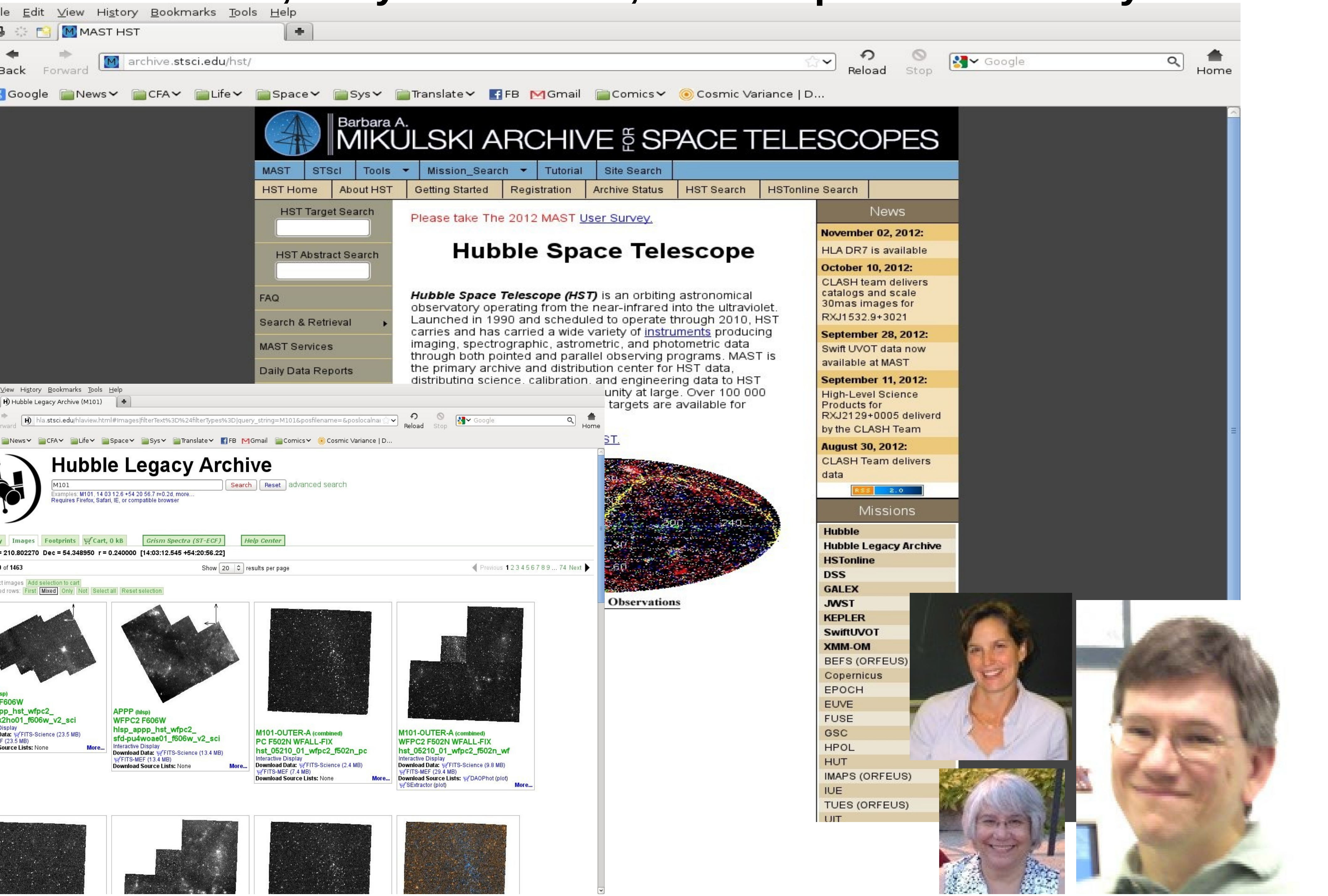
Today members of the public can download space telescope data and software for free

- learning how to use the software may be a bit more tricky!

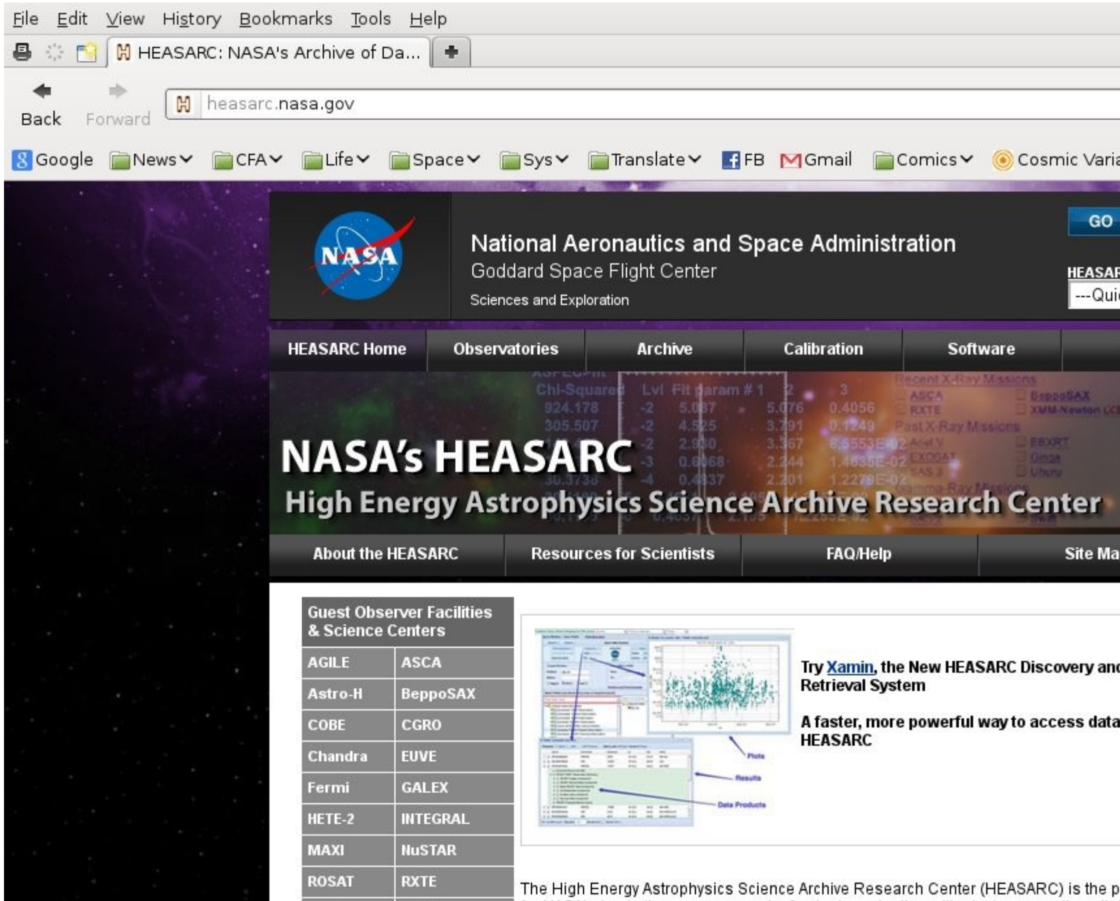
Pasadena, California: IPAC, for Infrared Astronomy



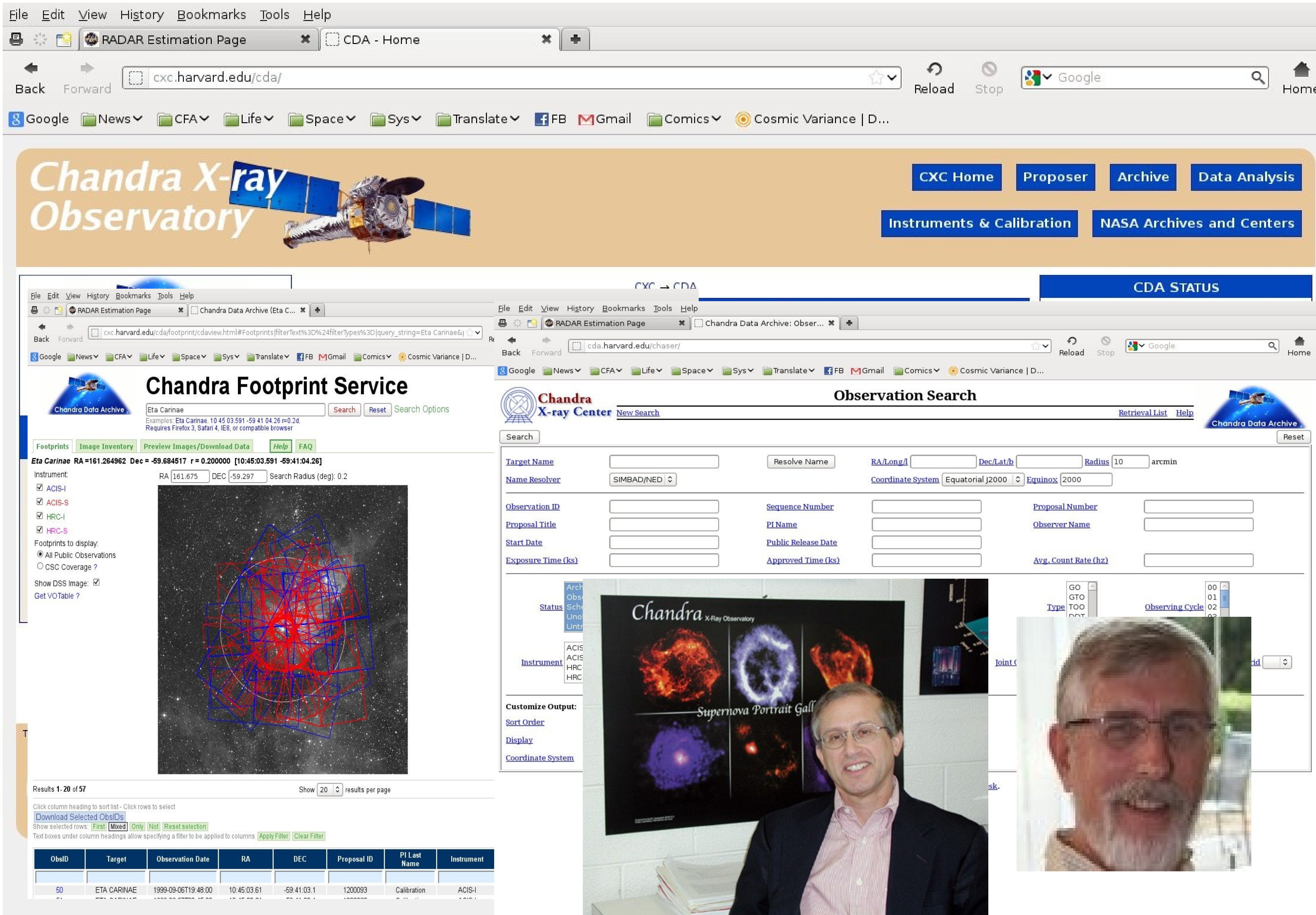
Baltimore, Maryland: MAST, for UV/Optical Astronomy



Greenbelt, Maryland: HEASARC, for High Energy Astronomy

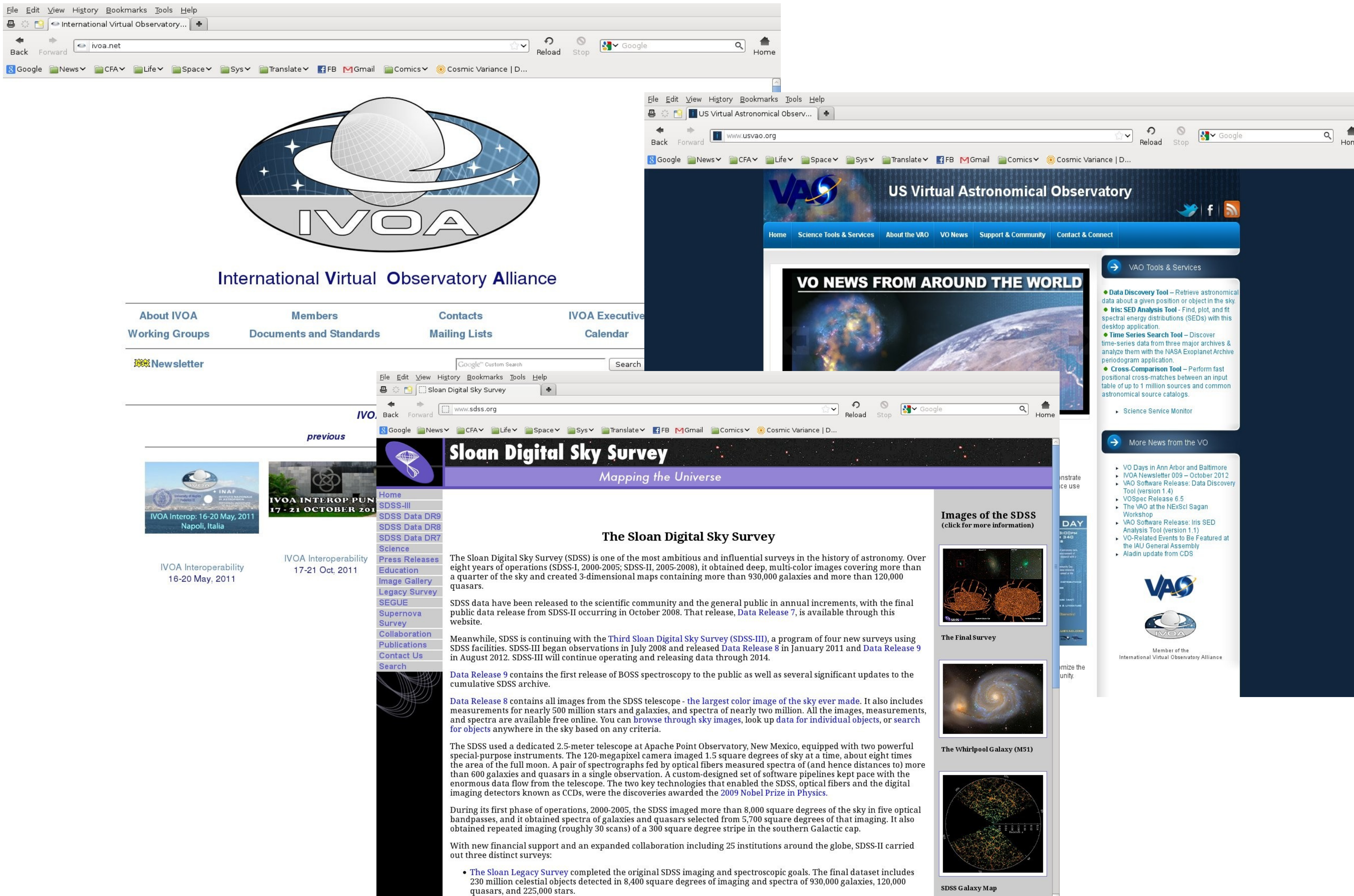


Cambridge, Mass: Chandra Observatory

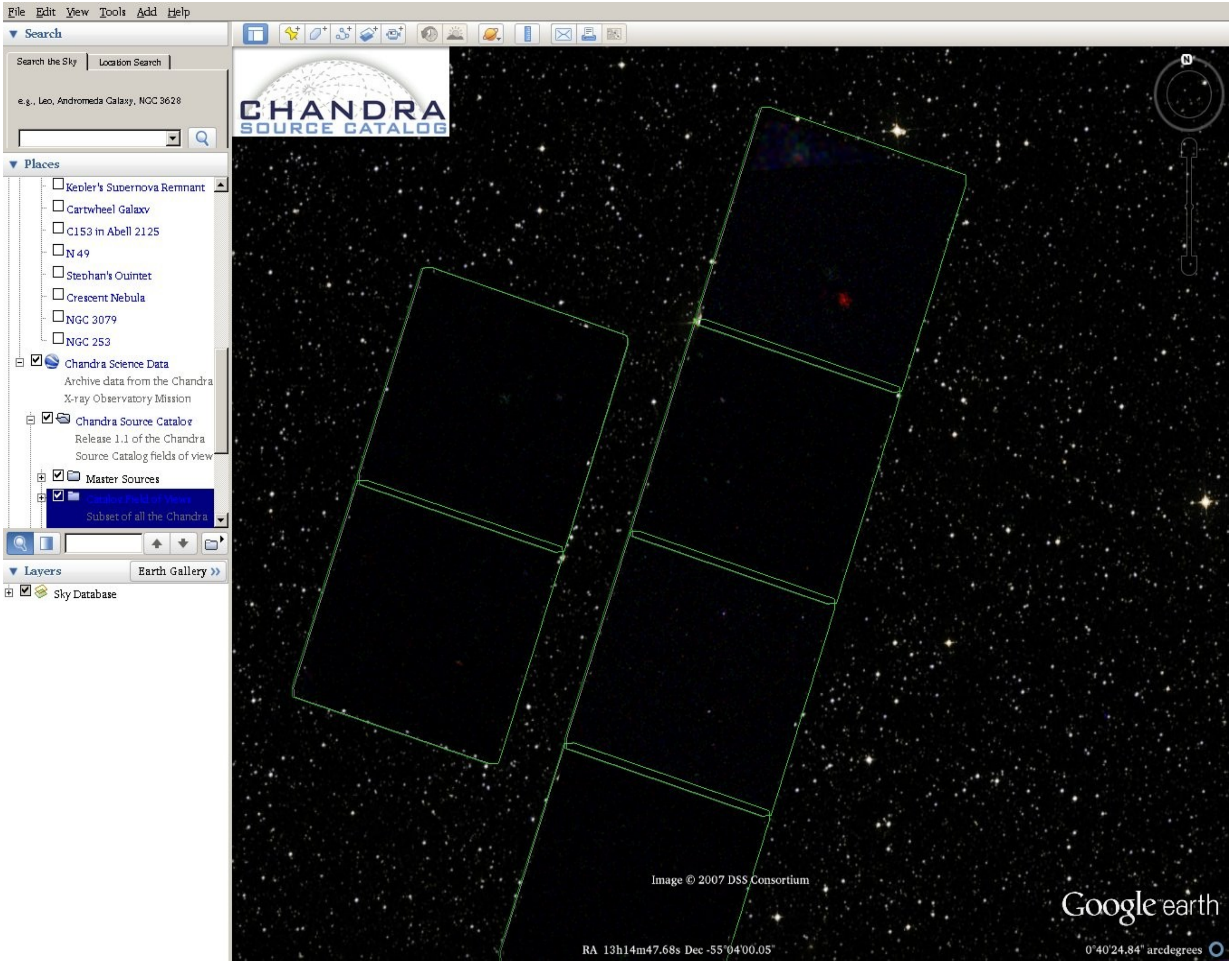


The 21st Century Challenge: Tying It All Together

How do I find all the data for one particular galaxy? Easily combine Hubble, Chandra, ground-based images?

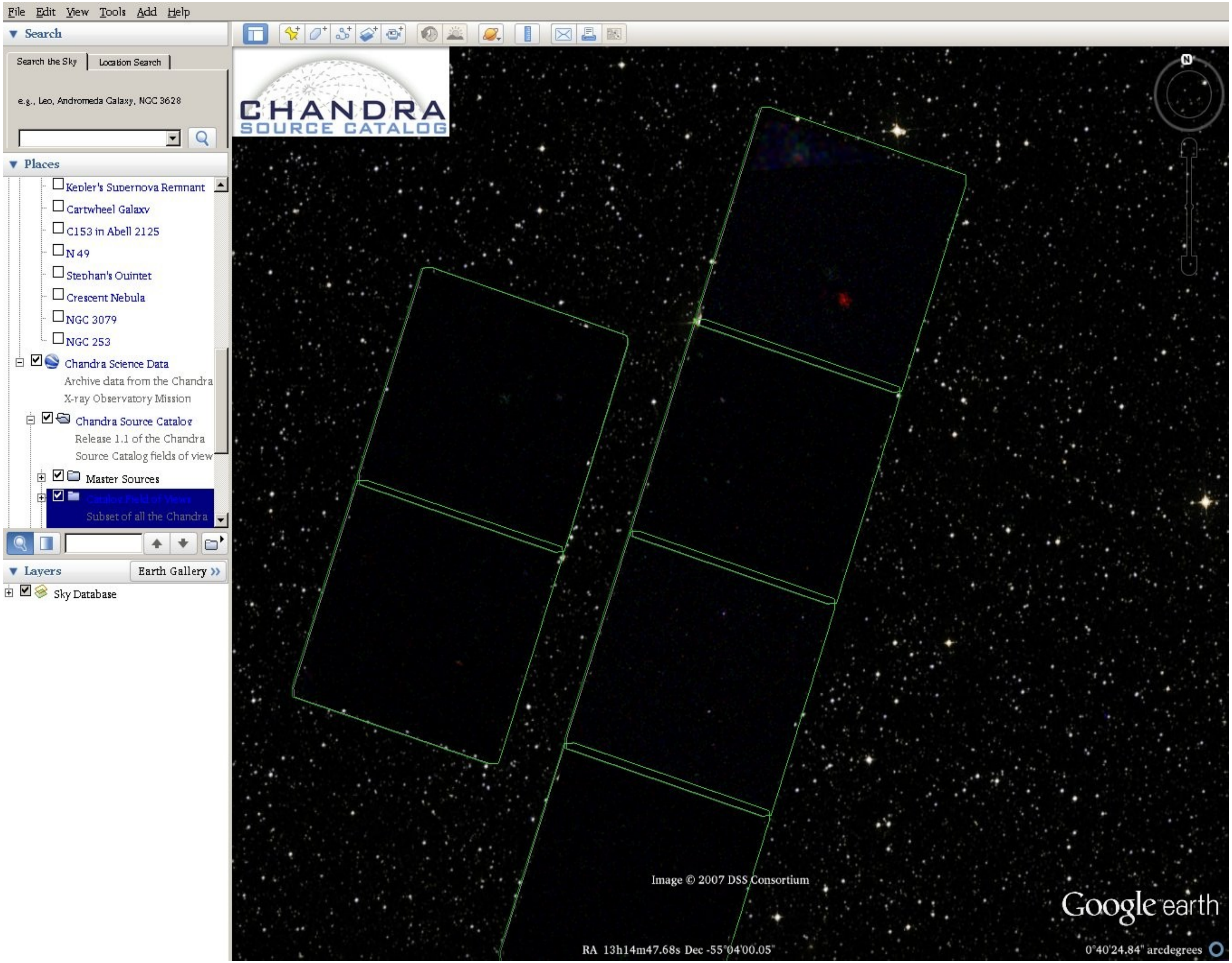


Outreach tools with real data: Sky in Google Earth



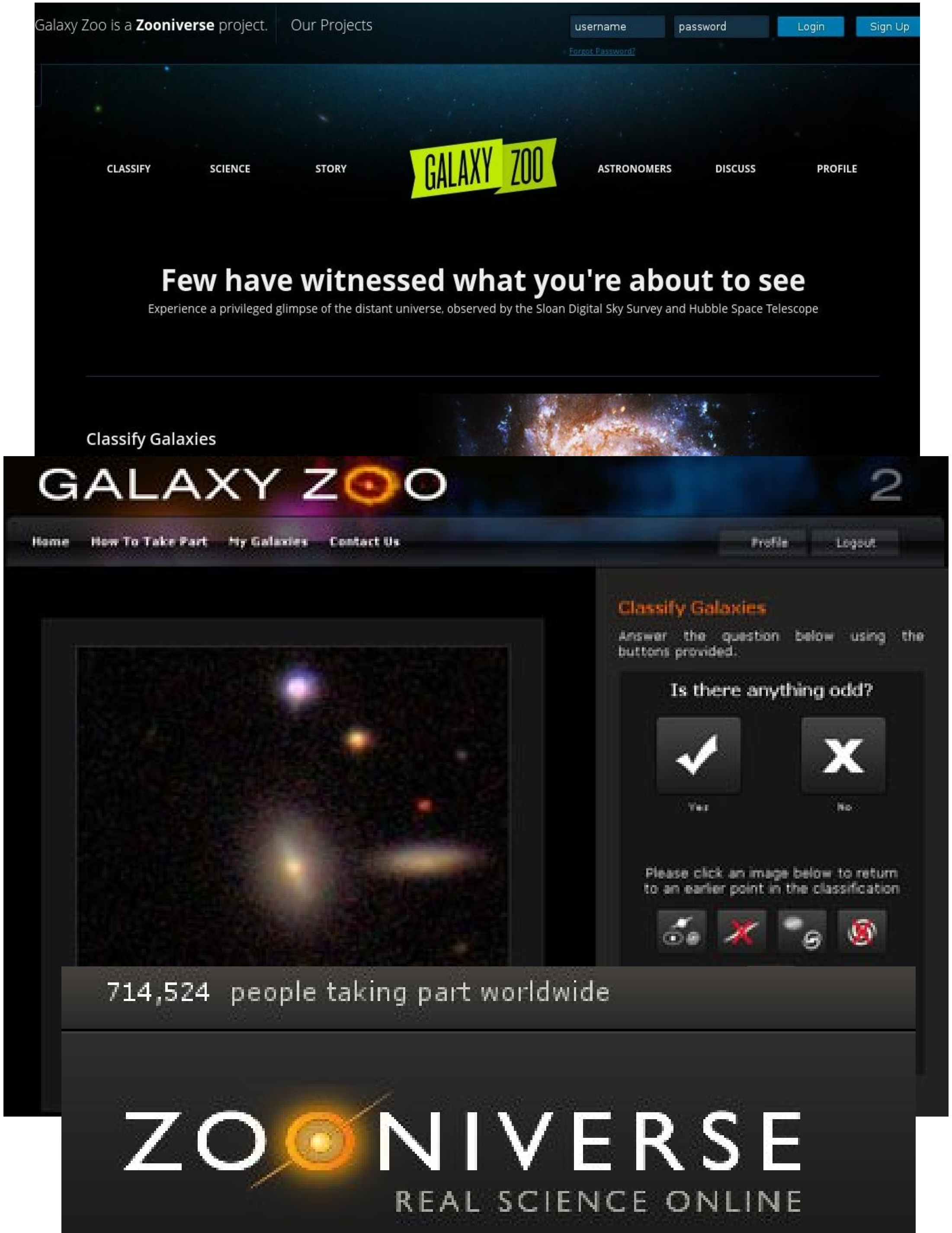
see also Microsoft's World Wide Telescope

Outreach tools with real data: Sky in Google Earth



see also Microsoft's World Wide Telescope

Citizen Science, Galaxy Zoo and the 'Zooniverse'



1999: SETI@home program borgs your unused computing resources

2007: Kevin Schawinski, Chris Lintott and others enlist the public to help classify galaxies on line – using spare human-brain compute power instead of spare digital CPUs: “Galaxy Zoo”

Is it a spiral or an elliptical?
Face or edge on?

Crowd-sourced science resulting in many published papers; on-line forum (Alice Sheppard et al) builds community

2012: Multiple projects including searching for extrasolar planets in Kepler light curves (“Planet Hunters”) and classifying cancer cells (Cell Slider)

How it all works: a space telescope team

Idea for experiment - astronomer anywhere in the world
Proposal planning - astronomer uses our software to see which camera is best for the experiment, how long they'd have to observe for

Proposal selection - 'peer review' by panel of experienced astronomers
Observation scheduling and planning - when is it best done? Which guide stars to use? What camera configuration?
Making the observation - flight control team sends up commands, retrieves the data
Making the calibration - Chandra's calibration team makes its own special observations throughout the year, updates our knowledge of the telescope

Data processing - go from raw data to science data with times, locations, brightness, color
Archiving - add lots of 'curation' information and stuff the data in the archive
Data analysis - astronomer gets the data from the archive, uses our software to turn the data into 'science'
Publication - astronomer writes up a paper and gets it peer-reviewed for the Astrophysical Journal, Monthly Notices of the Royal Astronomical Society, or another major journal

Astrometry: Where precisely is it on the sky?

Question: how can I find this object again?
What is its 'latitude and longitude'?

You can't just use Earth longitude – a star or galaxy rises and sets as the Earth spins. So we use a 'celestial longitude' (called 'right ascension' for historical reasons) that doesn't turn with the Earth.

Normal latitude is fine as a celestial latitude (but we call it declination also for historical reasons)

So we embed some data in the picture:

- what latitude and longitude is the middle of the picture?
- how many degrees across is the picture?
- more subtly, how do you map the sphere of the sky onto the flat picture? You probably know the 'Mercator projection' from atlases – we use a variety of different projections and we have to make a note of which one we used this time

Photometry: How bright is it?

Question: how bright is this object?
That's really two questions...

How bright does it appear to be to us? Is it blinding like the Sun, or super faint and needs a big telescope to see it?

How bright is it 'really'? That depends on how far way it is... the stars you see at night don't look as bright as the Moon but they are much, much further away, they only look faint because of their distance.

I'm going to ignore this second, harder question – once you know how bright it looks, you can figure out how bright it really is if you know the distance, but how we measure the distance to things in space is a tough problem and a whole other talk!

Even capturing exactly how bright each star and galaxy looks isn't easy. This is a process known as 'photometry'.

The multicolor universe: What kind of light?

The human eye sees different colors than a camera
Different filters can pick out different colors
An infrared or x-ray camera sees entirely different colors invisible to the human eye

A red-colored star may be brighter than a blue one when measured in a red filter, but the other way around in a blue filter

The nebula from an exploding star may be incredibly faint seen in any visible light colors, but really bright seen with an X-ray telescope.

It's all in the timing: when did I take this image?

In ancient times we thought the skies were unchanging, but actually things change on all timescales from microseconds to gigayears.

We obviously don't want to label our data in Eastern Daylight Time – astronomy's an international subject, someone in Japan might be studying the same object. So you'd think we'd put things in Greenwich Mean Time.

We sort of do, but... astronomers are the most persnickity people when it comes to calendars and watches. It turns out that GMT (or Universal Time) is tied to the spin of the Earth, which slows up and down as the continents slide around the core.

So instead we sometimes use Terrestrial Time, which is basically atomic clock time and is about a minute off from GMT

But wait! Einstein discovered that time isn't absolute – it runs at a different rate when you move fast (like the Earth does around the Sun) and when you go uphill in a gravity field (like the Earth does around the Sun!) So sometimes we use a Barycentric Time that corrects by a few milliseconds to account for Earth's motion around the sun.

So not only do we label our data with the time, but we must be careful to say what kind of time we used and where we measured it!