AN INTRODUCTION TO
ASTRONOMICAL DATA REDUCTION
THE FIRST DETECTORS: THE HUMAN EYE

William Herschell's Drawing of the Milky Way Galaxy
PHOTOGRAPHIC PLATES

- First objective, permanent record of astronomical phenomena
- Long exposure times possible; faint objects can be detected
- Large format: many surveys carried out as recently as the 1990s
- BUT quantum efficiency is very low, around 1% (most of the photons lost)!
- Photographic plates do not respond linearly to the amount of incident light.

Quantum efficiency (QE): incident photon to converted electron (IPCE) ratio
PHOTOELECTRIC DETECTORS
PHOTOMULTIPLIER TUBES (1950 – 1990s)

- More sensitive (30% quantum efficiency)
- Extremely stable broad band devices which made possible the first highly accurate photometric systems
- Flux calibration of photomultiplier based systems still more accurate than modern array imaging detectors!
- BUT only a single element detector! (leading to applications specialized in the study of single objects, variable stars...)


CHARGED COUPLE DEVICES
CHARGED COUPLE DEVICES

- Since the late 1980s CCD arrays have been the detector of choice for astronomy.
- Deep counts, gravitational arcs.
- Incoming photons produce free electrons in semiconducting material.
- This charge is then ‘shuffled’ to the output electrodes by applying electrical pulses.
- The output signal is converted from electron units to digital units by amplifiers with a conversion factor or gain, “g” (measured in electrons per analogue-to-digital unit).
CHARGED COUPLE DEVICES

• Gain is simply a multiplicative factor.
• Pixel ADU value $\times$ gain = number of electrons.

$$\sigma(x) = \sigma(y)/g^2$$
HOW CCDS WORK

Figure 5: Silicon Photodiode Anatomy

Figure 2b: The basic layout of a three-phase two-dimensional CCD. The sequence 1, 2, 3 on each set of electrodes indicates the normal direction of charge transfer in the parallel and serial registers.
HOW CCDS WORK

[Diagram of Bucket Brigade CCD Analogy]

Integration of Photon-Induced Charge

Raindrops

Parallel Bucket Array

Serial Bucket Array

Parallel Register Shift (1 Row)

Serial Register Shift to Output

Conveyor Belt

Calibrated Measuring Container

Figure 6
THE GREAT ADVANTAGE OF CCD CAMERAS...

- Extremely good **quantum efficiency** at redder wavebands -- almost 100%!
- Limiting magnitudes increased by four to five magnitudes!
- Sensitivity of telescopes is now limited by light collecting area and not the detector area.
- Ease of use.

- Detectors are linear over a very wide flux range making calibrations easier
- Output is digital!
THE GREAT ADVANTAGE OF CCD CAMERAS...
SOME PROBLEMS WITH CCD DETECTORS

• Poor blue (4500A) response

**NOTE:** CCD chips do not have the same range of response to light as the human eye

• **Megacam** (Canada-France Hawaii Telescope(CFHT)) is one of the few wide-field detectors with good u* response
• Different CCD manufacturing techniques can help solve this problem
• It’s hard to make large **monolithic** CCD detectors (but this can be solved by making mosaics from ‘buttable’ detectors).
SOME PROBLEMS WITH CCD DETECTORS

• **Charge transfer efficiency** between wells is not 100% and degrades over time especially in space-based detectors; this can cause problems when attempting to measure shapes to galaxies (see Massey et al. COSMOS papers).

• Cryogenic cooling (normally liquid nitrogen is used) is necessary to reduce dark current

**NOTE:** Dark current is the constant response exhibited by a receptor of radiation during periods when it is not actively being exposed to light.

• Reading out CCDs can take time!
WHAT CAN GO WRONG...

Diffraction spike/megaprime secondary mirror

Cosmic ray hits in megacam-u* image
WHAT CAN GO WRONG...

Charge transfer bleeding in STIS (Space Telescope Imaging Spectrograph) detector
CHARACTERIZING CCD DETECTORS

- **Quantum efficiency (QE):** what percentage of incident photons produce a measurable signal. Can be as high as 90% in visible wavelengths for the best detectors today.
A graph showing variation of quantum efficiency with wavelength of a CCD chip in the Hubble Space Telescope's Wide Field and Planetary Camera 2.
CHARACTERIZING CCD DETECTORS

- **Dynamic range/full well capacity/saturation level.**
  Around 105e-. At the saturation level, object fluxes ‘peg’ or wrap around.
  - It’s important to choose an exposure time short enough so that objects of interest are not saturated! (But long enough so you are sky-noise limited).
CHARACTERIZING CCD DETECTORS

Saturated star from megacam-CFHT
CHARACTERIZING CCD DETECTORS

- **Resolution/pixel size**: for the images to be properly sampled, there must be at least twice the number of resolution elements as the size of the point spread function psf; otherwise the images are said to be *undersampled*.
CHARACTERIZING CCD DETECTORS

The point spread function of Hubble Space Telescope's WFPC camera before corrections were applied to its optical system.

Hubble Space Telescope stellar image showing light spread over a wide area instead of concentrated on a few pixels as planned.
SOURCES OF NOISE IN CCD OBSERVATIONS

- **Dark current**
  - Even in the absence of illumination, CCDs produced a residual or ‘dark’ current. Typical values of around $10/e$ hr

- **Readout noise**
  - This is electronic amplifier noise, usually quoted in electrons. Does not depend on exposure time

- **Sky noise**
  - The night sky is not dark. Sky background increases at longer and longer wavelengths
In order to avoid being dominated by read noise, CCD exposures should be long enough to accumulate enough counts in the sky background -> sky noise dominated.
DATA REDUCTION
WHAT IS DATA REDUCTION?

Transformation of raw data into a more applicable form
WHAT IS DATA REDUCTION
AIMS OF DATA REDUCTION

- Removal of instrumental signatures, like dark current and field curvature.
- Masking of unwanted signals, like cosmic rays, stellar halos and satellite tracks.
- Photometric and Astrometric Calibration.
- Coaddition of individual frames.
DATA REDUCTION PACKAGES

• A number of program packages exist to perform data reduction steps.
• The biggest and most prominent ones are IRAF and MIDAS.
• These are very powerful packages but technologically they are dinosaurs in many aspects.
• A number of standalone programs for specific tasks exists as well.
• New instruments now often come with automatic data reduction pipelines.
IRAF

• IRAF is an acronym for Image Reduction and Analysis Facility.
• IRAF is a collection of software written at the National Optical Astronomy Observatory (NOAO) geared towards the reduction of astronomical images in pixel array form. This is primarily data taken from imaging array detectors such as CCDs.
IRAF

- It is available for all major operating systems for mainframes and desktop computers. Although written for UNIX-like operating systems, use on Microsoft Windows is made possible by Cygwin. It is primarily used on Linux distributions, with a growing share of Mac OS X users.
- IRAF commands (known as tasks) are organized into package structures.
Additional packages may be added to IRAF. Packages may contain other packages. There are many packages available by NOAO and external developers often focusing on a particular branch of research or facility. Of particular note are the STSDAS and TABLES packages by the STScI.
Welcome to IRAF. To list the available commands, type '?' or '??'. To get detailed information about a command, type 'help <command>'. To run a command or load a package, type its name. Type 'bye' to exit a package, or 'logout' to get out of the CL. Type 'news' to find out what is new in the version of the system you are using.

Visit http://iraf.net if you have questions or to report problems.

The following commands or packages are currently defined:

color. dimsum. imcnv. nmisc. rvx. tables.
cpio. fitsutil. language. noao. softools. utilities.
dataio. gmisc. lists. obsolete. spectime. vol.
dbms. guiapps. mscred. plot. stdas.
deitab. images. ndtasks. proto. system.

ecl>
MIDAS

- The ESO-MIDAS system provides general tools for image processing and data reduction with emphasis on astronomical applications including imaging and special reduction packages for ESO instrumentation at La Silla and the VLT at Paranal.
- In addition it contains applications packages for stellar and surface photometry, image sharpening and decomposition, statistics and various others.
MIDAS

- The official name, ESO-MIDAS, is a registered trademark. ESO-MIDAS is available under the GNU General Public License (GPL), and can be implemented on UNIX/Linux and Mac OSX systems.
MIDAS

![Graph and images related to MIDAS]
MIDAS
PIPELINE REDUCTION – PROS AND CONS

Diagram:
- Science data
- Calibration data
- Pipeline
- Reduced and calibrated data
PIPELINE REDUCTION – PROS AND CONS

- Allow for a fully automatic data reduction, saving a lot of cumbersome work.
- Can handle large data volumes.
- Can deal with complexities prohibitive for manual reduction (often true for IFUs).
- Are often black boxes. Do we really understand the data products of a pipeline reduction?
- Are a “one size fits all”. Never the ideal way to reduce data.
PIPELINE REDUCTION – PROS AND CONS

• Make assumptions about the data that may not even be true.
HISTORY OF THE THELI PIPELINE

- The new name THELI is either
  - a mystical dragon encompassing the whole Universe
  - Transforming HEavenly Light into Images
  - just a prettier name
- Written by a collaboration of scientists from Bonn, Bochum, and Garching.
- Publically available ftp://ftp.ing.iac.es/mischa/THELI/
HISTORY OF THE THELI PIPELINE

- THELI was initially developed for WFI@ESO/MPG2.2m.
- It is readily adaptable to other single and multichip cameras and to date has been used with about a dozen optical imagers.
- More recently it has been expanded to include reduction algorithms for near IR imagers (done by Mischa Schirmer).
- THELI was developed with weak lensing. It assumes empty fields without very extended objects.
HISTORY OF THE THELI PIPELINE

• More emphasis is put on precise astrometry, than precise photometry.
• It is also very modular and reduction steps can easily modified, skipped, or added.
• It makes heavy use of preexisting software. Together with our own software the pipeline is formed by a number of shell scripts.
• THELI can make use of multi-CPU architectures.
BACK TO DATA REDUCTION
NECESSARY FRAMES FOR DATA REDUCTION
SCIENCE FRAME
HOW WE CAN OBTAIN A BIAS FRAME

- The bias is measured by taking zero second exposures with the shutter closed.
- CCD electronics add “overscan regions” at the edges, i.e., a few virtual columns. These can also be used to measure the bias.
HOW WE CAN OBTAIN A DARK FRAME

- To measure the dark current a long exposure, 10 min or more, is taken with shutter closed. The only signal, except for bias of course, will be from thermal electrons.

NOTE: Most modern cameras are cooled with liquid N2 to 170 K. The dark current is negligible then.
FLATFIELD FRAME
HOW WE CAN OBTAIN A FLATFIELD FRAME

- Flatfields can be obtained by taking pictures of a uniformly illuminated screen inside the telescope dome (domeflats), observing a blank field during twilight (sky flat), or using deep science exposures using the night sky directly (night sky flats).
HOW WE CAN OBTAIN A FLATFIELD FRAME

• Domeflats can be taken during the day without hurry. But *uniform* illumination is difficult. Also, the spectral energy distribution of lamps is different from the night sky.

• The twilight sky spectrum is much closer to the night sky spectrum than lamps and it is very uniform over the field-of-view of cameras. But suitable twilight is very short, especially if one has to take flatfields for many filters.
HOW WE CAN OBTAIN A FLATFIELD FRAME

- Night sky flats (superflats) are ideal but they require many deep observations at different positions on empty fields to sample the night sky several times at every pixel.
STANDARD FRAME
HOW WE CAN OBTAIN AN STANDARD FRAME

- From astrometric standard catalogs, like the GSC2.2, USNOA2, USNOB1, UCAC2, . . .
PRE-REDUCTION
PRE-REDUCTION

• PreReduction (or Run processing) steps which take care of instrumental contaminations are done on each CCD chip independently, except for skybackground equalization.
BIAS FRAME
THE BIAS CORRECTION

- The bias a base level of charge always placed by the electronics on the CCD. It must be subtracted from all frames.
- Some CCDs show different horizontal patterns in every readout. Then the overscan must be used to subtract this pattern.
- Some CCDs show large scale gradients or regular patterns in bias frames. Then the full bias frame must be used.
THE BIAS CORRECTION

• How stable is the bias with time? Temperature changes in the electronics may cause drifts of the bias level.
THE DARK CURRENT CORRECTION

- Most modern cameras are cooled with liquid N2 to 170 K. The dark current is negligible then.
- If we need to correct for dark current, it is an additive effect that scales with exposure time.
- Dark frames can still be useful to identify chip defects that only show up when charge is gathered in neighboring pixels, e.g., next to hot columns or pixels.
FLATFIELD FRAME
THE FLAT FIELD CORRECTION

- Sensitivity variations across the CCD must be corrected for. These sensitivity variations can be measured by uniformly illuminating the CCD. This is a flat field exposure.
- The sensitivity variation is a multiplicative effect. To correct for it one needs to divide by the flat field.
Fringing is an additive – Newton-ring like – effect in red pass-bands caused by interference in the CCD substrate.
FRINGING PATTERN
Often the fringing pattern is spatially stable. Then the following method can be used to subtract the fringe pattern:

- The night sky flat contains the fringe pattern as an additive component.
- Smoothing the superflat gives an ‘illumination’ frame. The small scale fringe pattern is smoothed out. A frame with only the fringes is obtained by \( \text{FRINGE} = \text{SUPERFLAT} - \text{ILLUMINATION} \).
FRINGING PATTERN

- But the intensity of the fringe pattern depends on the sky brightness. Before subtracting the FRINGE frame from the flatfielded science frame, it must be rescaled according to the sky brightness.
FRINGING PATTERN
MANUAL EYEBALLING, MASK CREATION

Bright star reflection

Satellite track
REDUCTION
AFTER PRE-REDUCTION

Now have a set of frames free of instrumental signatures. We still have to

• photometrically calibrate them;
• astrometrically calibrate them;
• (optionally) subtract the sky background;
• combine them.
CREATING WEIGHT FRAMES
**WEIGHT IMAGES**

- The master flat image contains information on the relative sensitivity and hence noise variation in an image.
- The master flat rescaled to mode 1 provides the starting point for the weight map creation.
- Bad pixels are detected by thresholding the master dark, sky flat, and super flat.
- Bad pixels are set to zero in the weight map.
- THELI uses EyE to detect cosmic ray hits in the individual images. Their position is set to zero as well.
WEIGHT IMAGES

- Region masked by hand to eliminate, e.g., satellite tracks and ghosts are also set to zero in the weight map.
ADVANTAGES OF WEIGHT MAPS

• A weighted coaddition assumes Gaussian sky noise. Non-Gaussian pixels like hot/cold pixels violate this assumption. Hence, they must be known before coaddition.

• Rejecting bad pixels or cosmics based on a median or sigmaclip rejection require at 5–10 frames for robust statistics. Smaller number of images can be coadded cleanly with prior knowledge.

• Mosaic data is covered inhomogenously. The relative noise levels must be known for reliable object detection.
OBJECT DETECTION WITHOUT AND WITH WEIGHT IMAGE
PHOTOMETRY
WHAT AND WHY

• Many people are interested in astronomy because it is visually exciting.
• The many marvelous pictures of celestial objects taken using large telescopes on the ground or in space are certainly the most visible manifestation of modern research astronomy.
• However, to do real science, one needs far more than pictures. Pictures are needed as a first step in classifying objects based on their appearance (morphology).
WHAT AND WHY

• To proceed past this initial stage of investigation, we need quantitative information - i.e. measurements of the properties of the objects.
• Observational astronomy becomes science only when we can start to answer questions quantitatively: How far away is that object? How much energy does it emit? How hot is it?
WHAT AND WHY

• The most fundamental information we can measure about celestial objects past our solar system is the amount of energy, in the form of electromagnetic radiation, that we receive from that object.
• This quantity we will call the flux. The science of measuring the flux we receive from celestial objects is called photometry.
• Photometry usually refers to measurements of flux over broad wavelength bands of radiation.
Measurement of flux, when coupled with some estimate of the distance to an object, can give us information on the total energy output of the object (its luminosity), the object’s temperature, and the object’s size and other physical properties.

If we can measure the flux in small wavelength intervals, we start to see that the flux is often quite irregular on small wavelength scales. This is due to the interaction of light with the atoms and molecules in the object. These “bumps and wiggles” in the flux as a function of wavelength are like fingerprints.
WHAT AND WHY

- They can tell us lots about the object—what it is made of, how the object is moving and rotating, the pressure and ionization of the material in the object, etc.
- The observation of these bumps and wiggles is called spectroscopy.
- A combination of spectroscopy, meaning good wavelength resolution, and photometry, meaning good flux calibration, is called spectrophotometry.
- Obviously, there is more information in a spectrophotometric scan of an object compared with photometry spanning the same wavelength range.
WHAT AND WHY

• Why would one do low wavelength resolution photometry rather than higher resolution spectrophotometry or spectroscopy, given the fact that a spectrum gives much more information than photometry?
• It is much easier to make photometric observations of faint objects than it is to make spectroscopic observations of the same object. With any given telescope, one can always do photometry of much fainter objects than one can do spectroscopy of.
WHAT AND WHY

• On a practical note, the equipment required for CCD imaging photometry is much simpler and cheaper than that needed for spectroscopy.
• With low cost CCDs now readily available, even small telescopes can do useful photometric observations, particularly monitoring variable objects.
Astrometry determines how the pixel in an image are matched to positions on the sky. This is done using astrometric standard catalogs, like the GSC2.2, USNOA2, USNOB1, UCAC2, . . . These provide typically at least 30–40 stars per WFI chip with an accuracy of 300 mas. UCAC2 has an accuracy of 70 mas but much fewer stars. Other high precision astrometry catalogs exist but they only have a few thousand stars.
ASTROMETRY

- Object catalogs are created for each science image.
- The Astrometrix package is used to simultaneously match the frames with each other and with the external reference catalog.
- This is done by fitting a bicubic polynomial describing shift, rotation, and distortion for each chip.
- The external accuracy is limited by the accuracy of the reference catalog.
- The internal accuracy is typically 1/10th of a pixel (1 error). This is necessary to not distort the shapes of small objects.
Nonlinear component of astrometric solution for each WFI chip. A global solution for the entire mosaic cannot be used.
ACCURACY OF ASTROMETRIC SOLUTION

Comparison with USNOA2 (top left) and UCAC2 for three fields.
PREPARATION FOR COADDING

Sky background is calculated with SExtractor BACKGROUND check image for every large-object-subtracted image and is subtracted from all SCIENCE frames.
COADDICTION

THELI’s SWarp first undistorts and resamples (with LANCZOZ3 kernel) all input SCIENCE and WEIGHT frames according to astrometric solution. With all resampled input images belonging to a given output pixel present, SWarp calculates final results using weighted mean method.
EXAMPLE OF THE FINAL RESAULT
ՇՆՈՐՀԱԿԱԼՈՒԹՅՈՒՆ