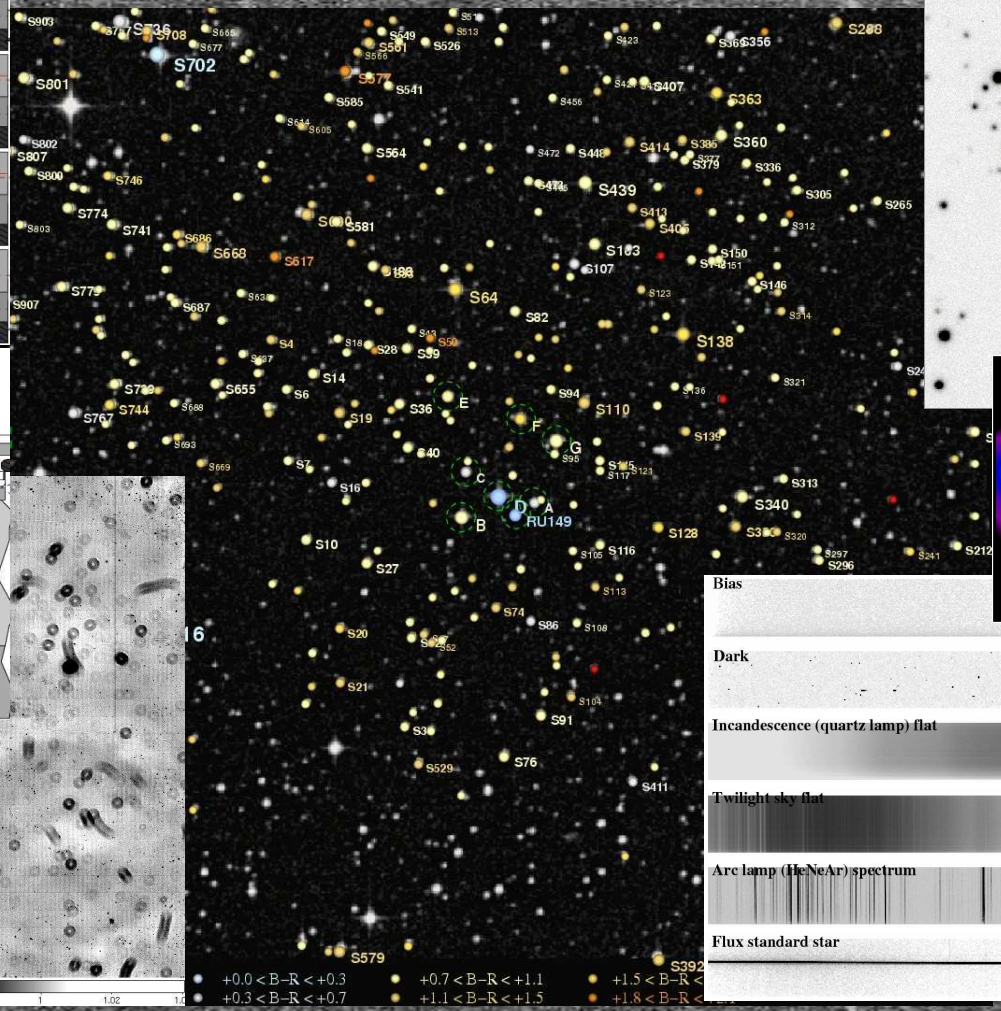
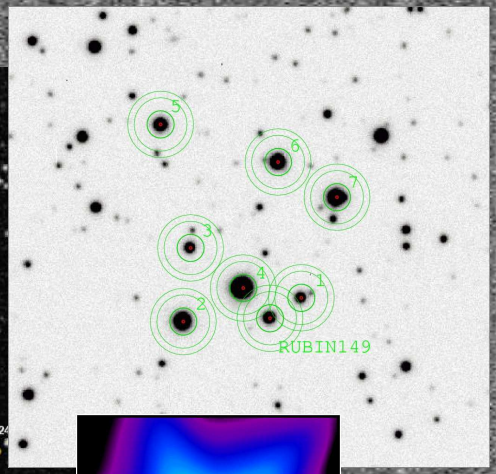
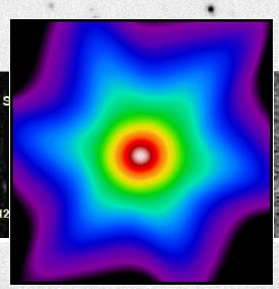
AST 598 'Instrumentation and Data Analysis' course, ASU, Fall 2013

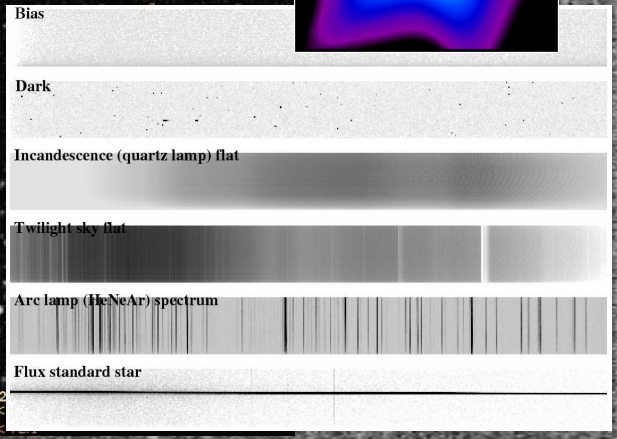
Astronomy with Charged Coupled Devices

R.A. Jansen









- **The e-Book companion to this section is available from:**
  http://www.public.asu.edu/~rjansen/ast598/2006ACCD.ebook...1J.pdf

- **Another useful companion book to this section is: Howell, S.B. 2006, *Handbook of CCD Astronomy*, 2nd (or 3rd) edition (Cambridge University Press, Cambridge UK)**

- ***An excellent review paper on the workings of a CCD is MacKay, C.D. 1986, ARA&A 24, 255.***

Introduction to Charge Coupled Devices

- ***Charge Coupled Devices*** **(CCDs) were originally developed at Bell Labs as a solid-state memory device (Amelio, Tompsett & Smith 1970; Boyle & Smith 1970).**

- **A CCD is essentially a regular array of large numbers of coupled *Metal-Oxide Semiconductor Field Effect Transistors* (MOSFETs).**

- **Sensitivity to light via the *photo-electric effect* and the ability to store charge in a regular array format was suitable for application as a 2-D detector.**

- **In a CCD, two or more electrodes (called *gates* in MOSFET terminology) per isolated *pict*ure *el*ement (*pixel*) create potential wells in the silicon.**

- **When a photon is absorbed in silicon, an *electron-hole pair* is created.**

  - **The generated photo-electrons migrate to and are confined within the potential well created by the gate kept at the largest positive Voltage.**

  - **The holes diffuse away into the bulk of the silicon, thus preventing recombination.**

Introduction to Charge Coupled Devices (cont'd)

- In a three-phase CCD (still the most common type in astronomical applications), three *gates* —each isolated from the silicon substrate and from one another— define a pixel *row*.

- Fig 1 illustrates the structure of a three-phase CCD, and Fig 2 shows a top view of a portion of an actual device.

- *Channel stops* (buried beneath and electrically isolated from the gates) of heavilly doped p-type Si define pixel *columns*, and prevent (due to their negative charge) electrons from migrating along the length of the gates.

- Every third gate is electrically connected in parallel and kept at *the same* voltage.

- The polysilicon gates partially overlap, but are electrically isolated from one another by a thin layer of SiO_2.

- The relatively thick ($\sim 1\mu$m) layer of n-type Si, below the thin ($\sim$80nm) insulating SiO_2 and Si_3N_4 layers, serves to reduce charge transfer losses: it keeps accumulated electrons away from the surface, where some might become trapped.

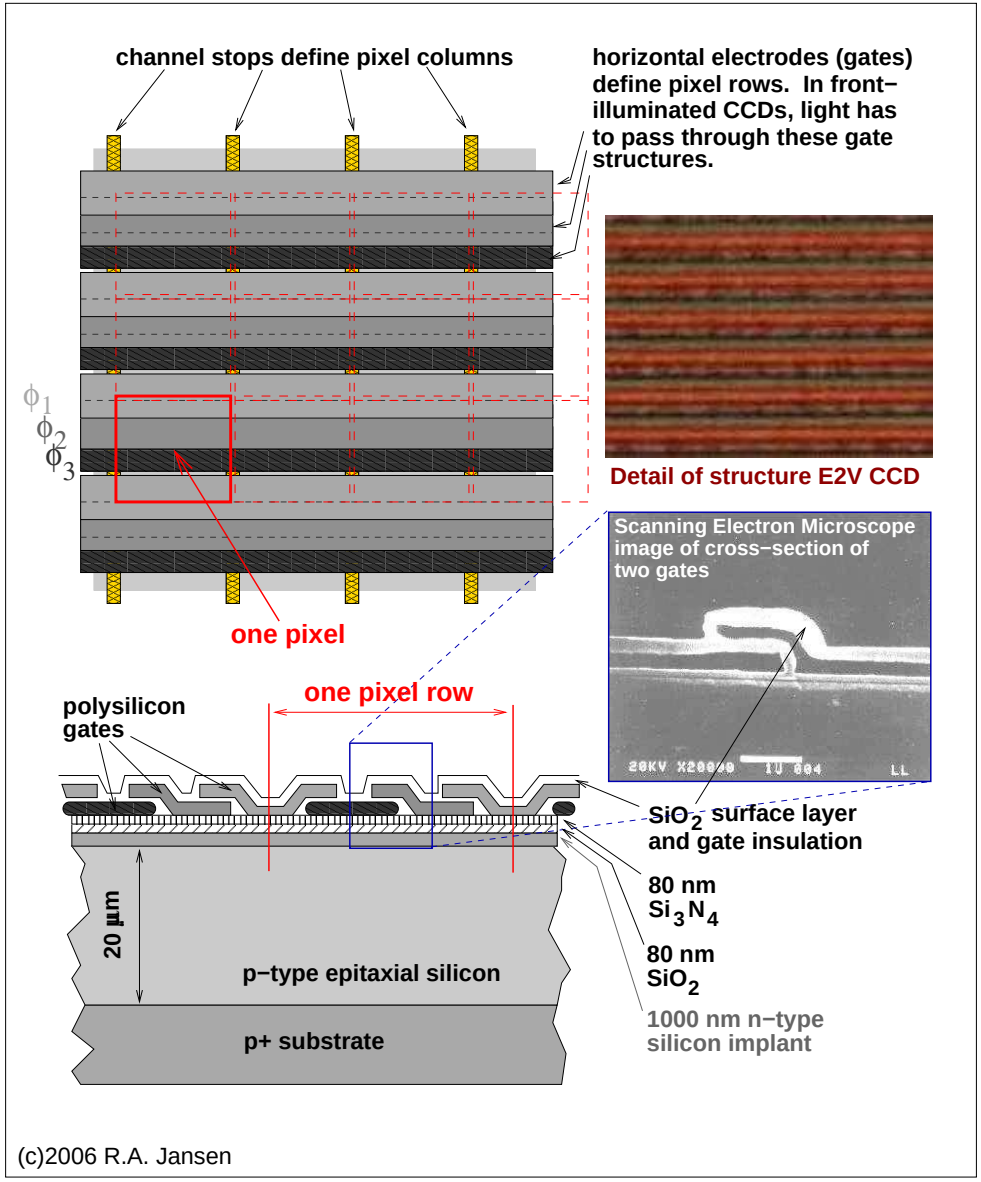


Figure 1: Schematic representation of the gate structures on a typical 3-phase CCD.

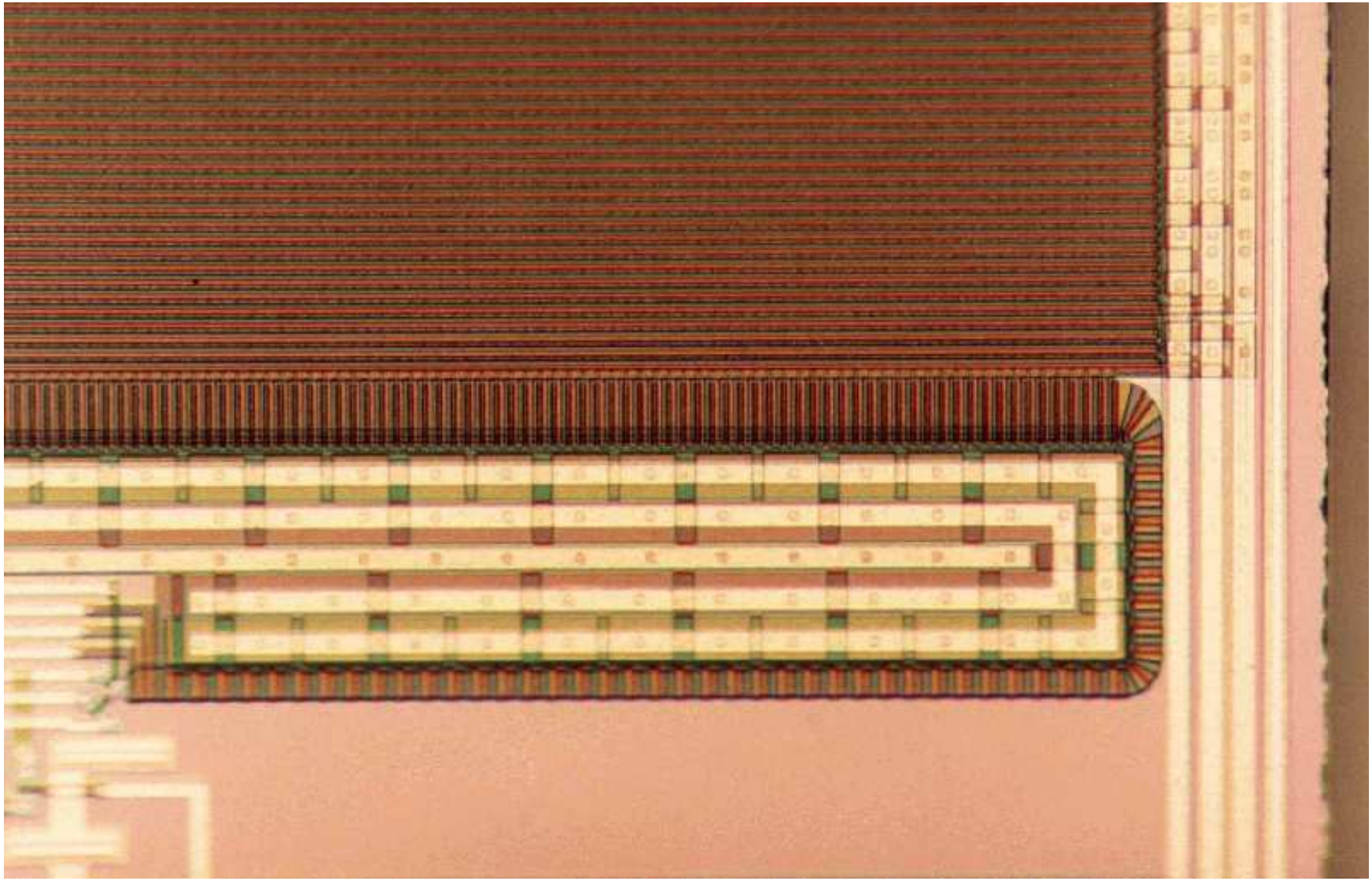


Figure 2: Photomicrograph of a corner of an E2V L3CCD.

Operation of Charge Coupled Devices

- **Read-out of the 2-D array of stored photo-electrons involves cycling the voltages on the electrodes of the array in such a way as to shepherd the stored charges along the *columns*, effectively passing them on from one pixel to the next.**

- **At the end of a column, charges are passed on to a special set of pixels, the *serial register* with gate structure that is orthogonal to that in the image area of the CCD.**

- **In this serial register, the charges collected in an entire pixel *row* are shifted one pixel at a time and passed on to on-chip electronics that *amplify* and *estimate* the stored charge.**

- **That estimated charge is subsequently converted by an *analog-to-digital* (A/D) circuit into a digital number, referred to interchangeably as either *Data Number* (DN), or *Analog-to-Digital Unit* (ADU), or simply *Count*.**

  - **Pixels in the *serial register* tend to be larger because they may need to hold more charge (if on-chip binning is required), and may be operated at higher voltages.**

  - **Device shown in Fig. 2 has a special extended register where the charge is amplified: clock voltages in that portion are sufficiently large to ensure a finite probability of dislodging additional electrons in each charge transfer from pixel to pixel.**

Operation of Charge Coupled Devices (cont'd)

- **Analogy for the process of reading out a CCD is that of a "Bucket Brigade", where the contents of buckets (each representing a pixel that can store charge during integration) on many *parallel* belts are emptied in a *serial* belt with one bucket per parallel belt.**

- **The buckets on the serial belt are then emptied one at a time into a measuring container.**

- **After the last bucket on the serial belt has been measured, the next row of buckets on the parallel belts is emptied into the buckets on the serial belt, and so on until all buckets have been measured.**

- **See: Fig. 3.**

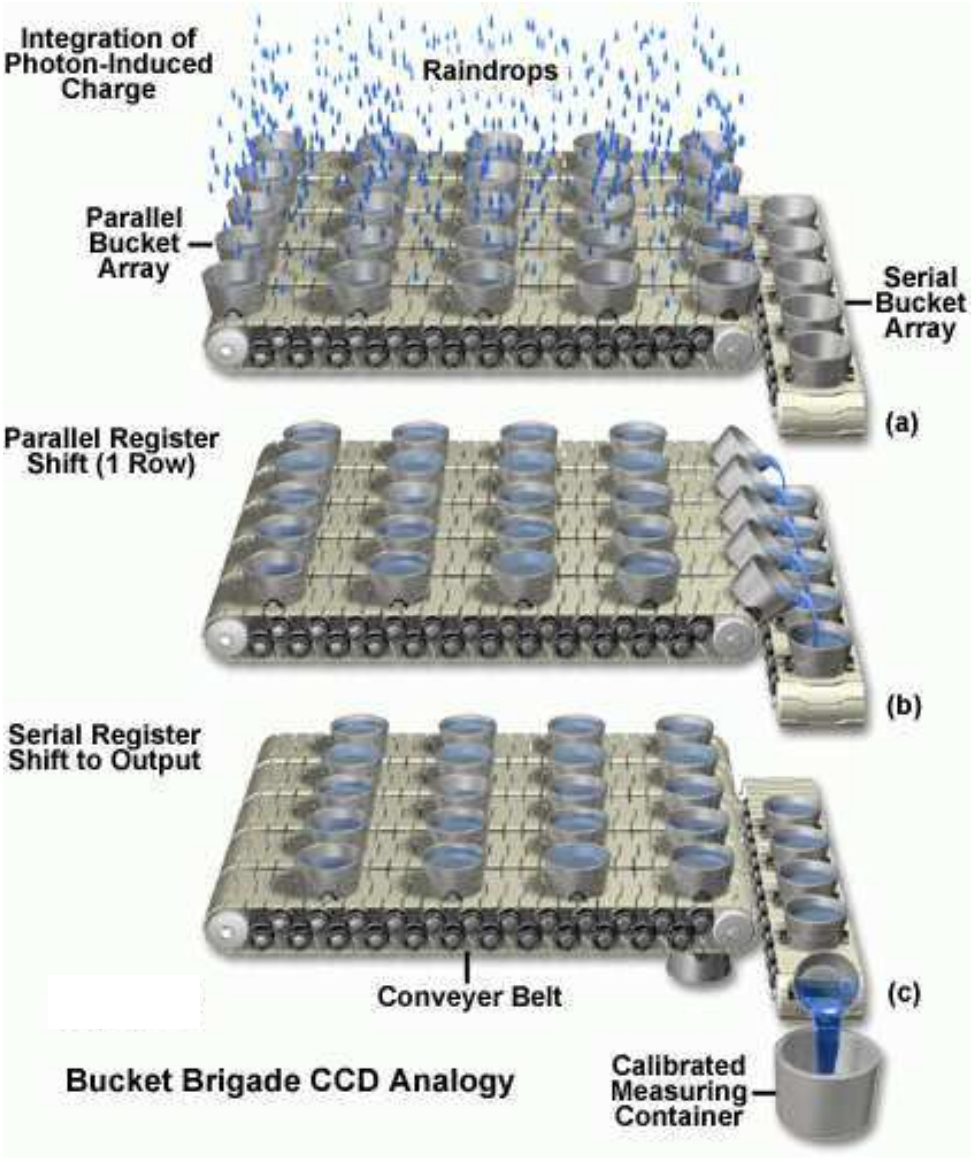


Figure 3: The "bucket brigade" analogy often used to visualize the operation of a CCD. *Courtesy: Nikon/K.R. Spring, T.J. Fellers & M.W. Davidson.*

Operation of Charge Coupled Devices (cont'd)

- **A more detailed illustration of how the charges are passed along from pixel to pixel by cycling the gate voltages is given in Fig. 4.**

- **During integration, the middle gate (ϕ_2) is kept at a large positive voltage: all photo-electrons generated under any of the three gates defining a pixel migrate to and remain confined in the deep potential well of this gate.**

  - **Note, that the effect of the portion of the middle gate that overlaps adjacent gate ϕ_3 is shielded.**

- **During read-out, the voltages on the gates are cycled in succession: the accumulated electrons follow the deepest potential well until, at the end of a single parallel *clocking cycle*, they have been transferred to the adjacent pixel along the column.**

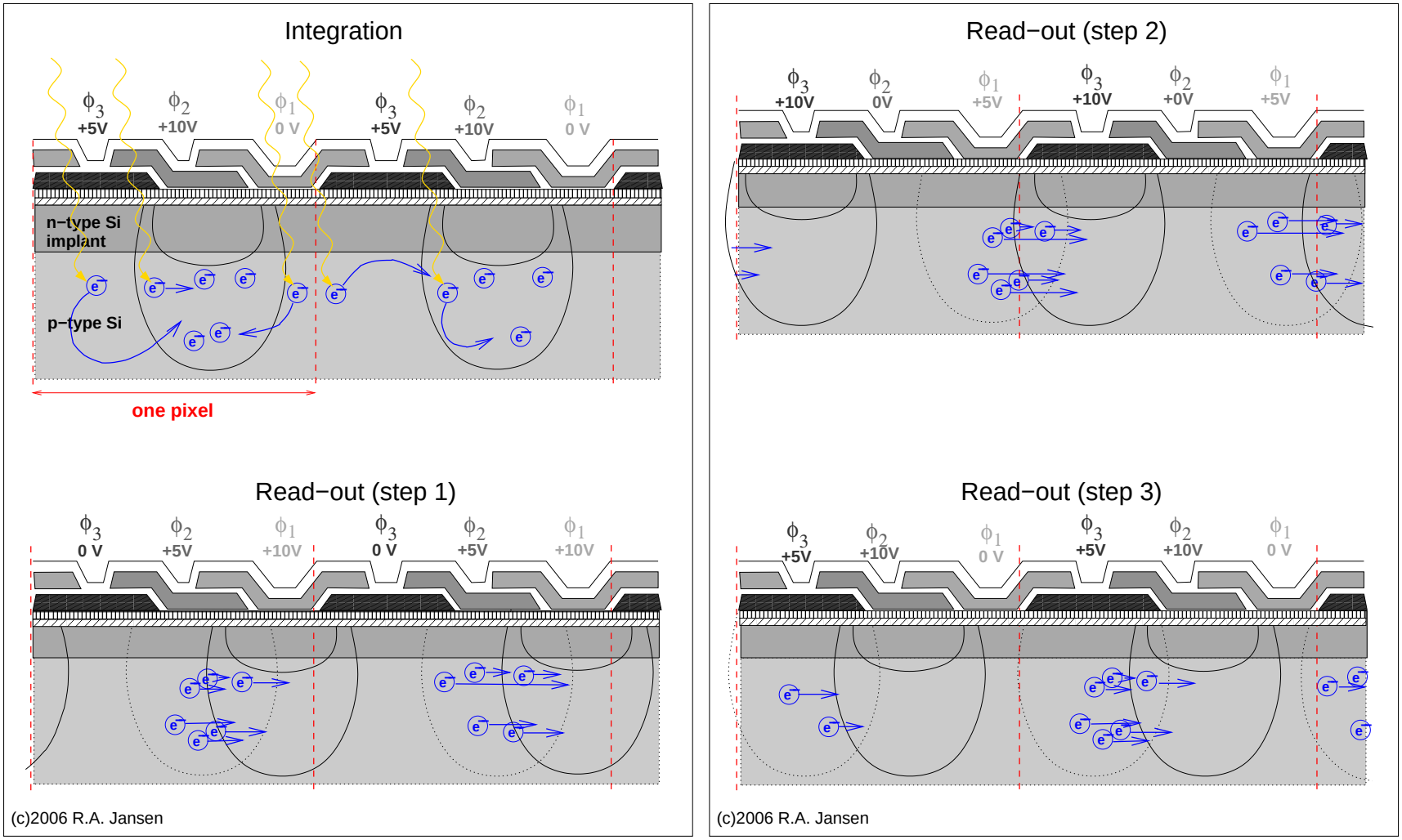


Figure 4: Schematic representation of integration and read-out of a CCD. The illustration follows the transfer of charge within a single pixel column.

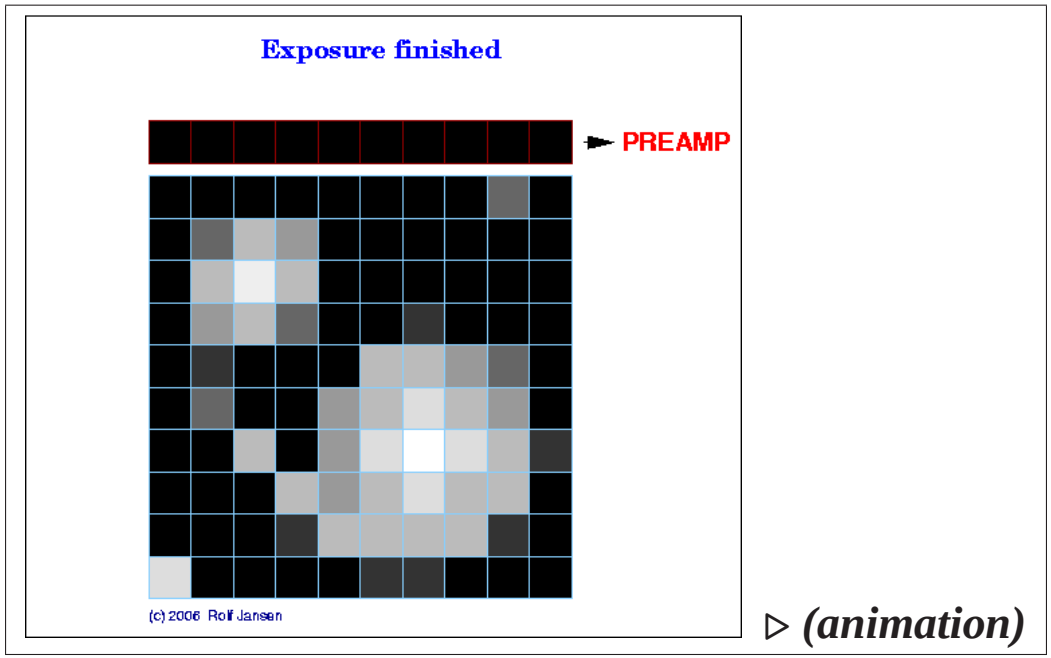
Operation of Charge Coupled Devices (cont'd)

- **Since the voltages on the gates in each pixel row along a column are connected in parallel, a single parallel clocking cycle will transfer the accumulated charge in the pixels of a row to the corresponding pixels of the serial output register (not shown).**

- **The charge is then transferred out of the serial register, one pixel at a time, to be amplified and sampled by the *A/D converter*. This is repeated until the entire array has been read out.**

Operation of Charge Coupled Devices (cont'd)

- Animation of the operation of a CCD, consisting of:
  - preparing the CCD: clearing any residual accumulated charge,
  - exposing the CCD to light, and
  - read-out of the CCD.



▷ *(animation)*

Analog to Digital Conversion

- **A charge accumulated in a single pixel and passed along columns and then along the serial register eventually makes it to the on-chip amplifier (*pre-amp*) circuitry, and out to the off-chip *CCD head electronics*.**

- **There, through a technique called *Correlated Double Sampling* (CDS), the pixel charge is sensed (as an analog voltage) as the difference between an arbitrary reset voltage V_0 and that arbitrary reset voltage plus the charge packet from the pixel.**

- **The CDS sampling time is finite, typically 10–20 μsec for both $-V_0$ and $V_0 + V_{\rm pix}$.**

  - **faster $\longrightarrow$ less accurate (higher *read-noise*); slower $\longrightarrow$ more accurate (lower *read-noise*)**

- **That analog voltage is subsequently represented as a digital number (*ADU*, *DN*, or *Count*) by the Analog-to-Digital converter.**

Analog to Digital Conversion (cont'd)

▷ ***Gain*** **— the average ratio of the analog charge measurement and the assigned data number. The gain is expressed in units of e^-/ADU.**

- **The *gain* value is usually ≥ 1, so the assigned data number is *smaller* than the number of photo-electrons (i.e., a "reduction" rather than a "gain"). That is why in the literature one sometimes encounters the term *inverse gain* for this parameter.**

- **The minimum statistical error in any pixel value is 0.5 gain. Thermal noise within the on-chip (extended register +) pre-amp or in the CCD head-electronics can only increase this error.**

- **To keep amplifier and A/D electronics simple, and to avoid wasting an additional sign-bit for storing the digital data, the CCD electronics add a fixed offset or *bias* voltage.**

  - **Even when *no* photons were detected in a given pixel, the A/D converter will be presented with a positive measurement.**

Analog to Digital Conversion (cont'd)

- **The maximum number that can be stored within an n-bit un-signed digital number is $2^n - 1$ (assuming we start counting at 0).**

  - **For a typical 16-*bit per pixel*, the maximum number is $2^{16} - 1 = 65535$; had we used a sign-bit, that maximum number would have been only $2^{n-1} - 1 = 32767$.**

- **With modern, non-truncating A/D converters, the average value in ADU returned for a signal of $S\ e^-$ is:**

$$\mathrm{int}(S/\mathrm{gain} + 0.5).$$

  - **Older truncating A/D converters would have returned: $\mathrm{int}(S/\mathrm{gain})$.**

- **A/D conversion is supposed to be highly linear over the entire operational range, from $N_{e^-} = 0$ to full-well capacity.**

  - **As electronics becomes smaller and cheaper, while overall noise is reduced, thermally induced non-linearities may become noticeable (due to slight warming and cooling of the amplifier, depending on how much power it draws, i.e., on how much charge it saw per unit time).**

Operation of Charge Coupled Devices (revisited)

▷ ***Full-well capacity*** **— the maximum number of photo-electrons that can be stored within a given pixel, typically $\sim 10^5 e^-$.**

- **When more charge accumulates than can be confined by the gate potential (*full-well saturation*), the excess charge may bleed into adjacent pixels along a column.**

- **When the measured charge is too large to be represented as a digital number in the available number of bits, we speak of *A/D saturation*.**

▷ ***Dynamic range*** **— the ratio of maximum to minimum measurable signal, typically $\sim 10^4$.**

- **To prevent thermal electrons from the bulk silicon to also be dislodged and collected in the pixel potential wells (*dark current*), and to minimize the thermal noise added by the read-out electronics (both on-chip pre-amp and in the adjacent "head electronics"), CCDs must be operated at cryogenic temperatures.**

  - **Typically, astronomical CCDs are cooled using liquid N_2 to ($-100°$C – $-120°$C, or 170–150 K). Thermo-electric cooling is also possible (common in high-end CCDs for (amateur)astrophotography).**

Operation of Charge Coupled Devices (revisited)

- **The *temporal resolution* is set by the maximum rate at which the array can be read out, typically $\gtrsim$0.01 s, which is set by the size of the array and the minimum sampling time per pixel to accurately estimate the accumulated charge.**
  - **In most integrating applications, the temporal resolution is *much* lower ($\sim$30 s–1800 s).**
  - **Effective read-out times have been reduced with CCD designs that allow simultaneous read-out of four quadrants of the CCD through serial registers and amplifiers in each corner of the CCD, and through development of faster, lower-noise read-out amplifiers.**

Characterization of CCDs

- **The active photosensitive material of CCDs is pure silicon. Silicon therefore largely determines the response of the detector to light of various wavelengths.**

▷ **The *absorption length* of photons in a material is that distance after which a fraction of $1/e$ remains (or, equivalent, that distance after which 63% of the photons are absorbed).**

○ **Fig. 5 shows the dependence of the absorption length on wavelength from X-rays ($\sim$10Å) to the near-infrared 1.1 μm cut-off (corresponding to the Si band-gap energy of $\sim$1.4 eV).**

- **Shortward of $\sim$400 nm, photons are absorbed very near the surface of the CCDs.**

  ○ **In *front-illuminated* CCDs, this absorption is likely to occur already within the layers of the gate structures and/or the Si_3N_4 and SiO_2 isolation layers;**

  ○ **In thinned, *back-illuminated* CCDs, this is much less of a problem.**

- **Shortward of $\sim$250 nm, the penetration depth of photons increases again, but such energetic photons can generate multiple electron-hole pairs within the silicon. At very short wavelengths (X-ray and shorter) they can even cause damage to the CCD itself.**

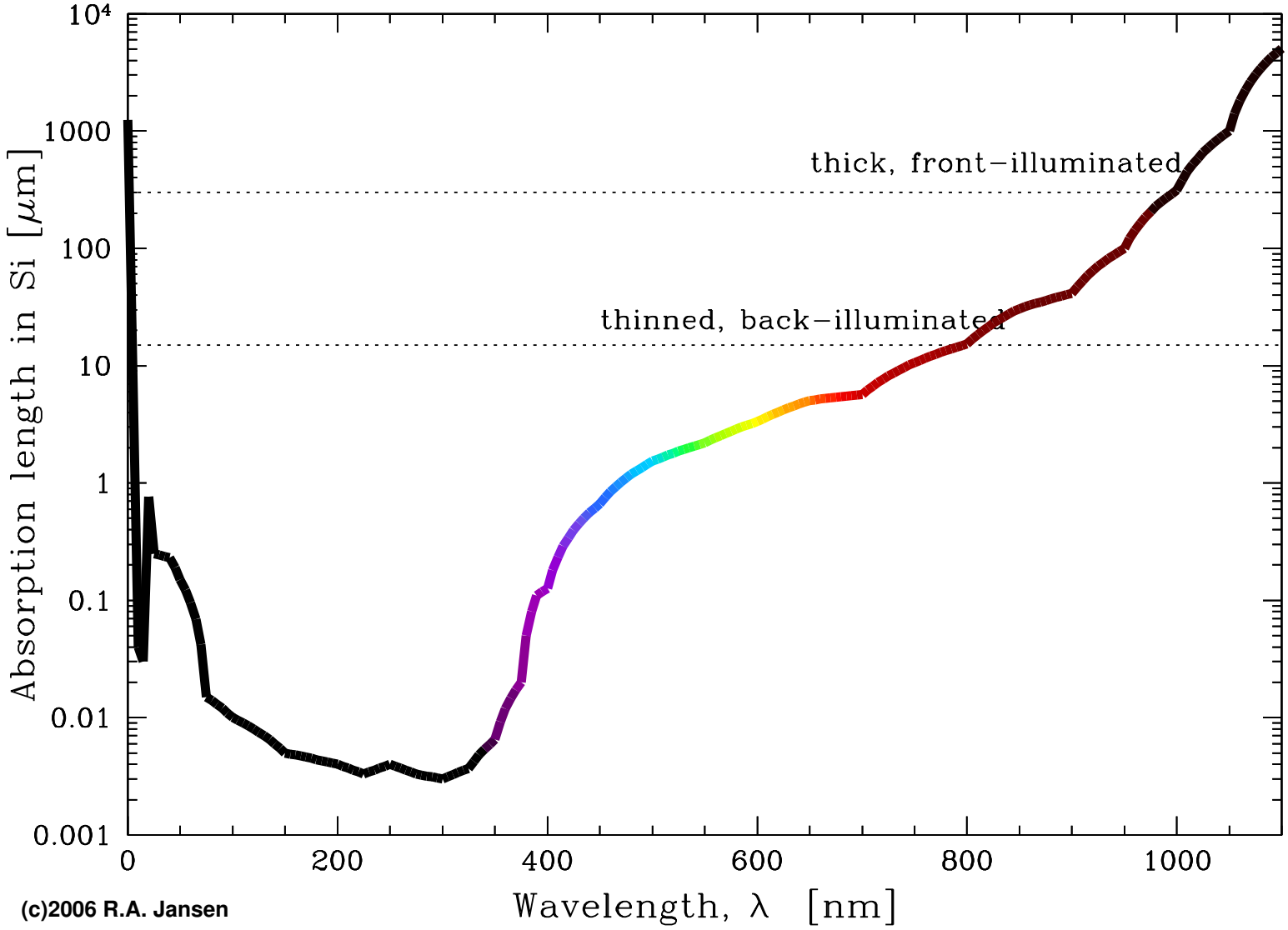


Figure 5: Photon absorption length in silicon (in μm) as a function of wavelength (in nm). Dotted lines indicate the typical thickness of the silicon substrate in front- and back-illuminated CCDs. Not included in this curve are photons that never made it into the Si, but were reflected off of the surface.

Characterization of CCDs (cont'd)

- **Longward of ∼800 nm, thinned CCDs become transparent to incident photons, because the absorption length exceeds the typical thickness of the silicon substrate.**

- **At redder wavelengths, the absorption may occur deep in the bulk substrate.**

  - **In *front-illuminated* CCDs, the absorption occurs where the pixel gate potential well is relatively shallow, rather than in the *depletion region*;**

  - **The resulting photo-electrons may migrate to a neighboring pixel along a column, or underneath the channel stops to a neighboring pixel along a row;**

  - **In *back-illuminated* CCDs, the *bluest* photons tend to get absorbed furthest from the gates and *depletion region*.**

▷ ***Charge diffusion*** **— migration of photo-electrons created deep within the bulk substrate and/or well away from the center of the pixel volume to neighboring pixels along a column or row.**

- **Because the depth of the potential well is not uniform within the pixel volume, charge diffusion is often ascribed to *CCD sub-pixel variations*. The net effect of charge diffusion is image blurring.**

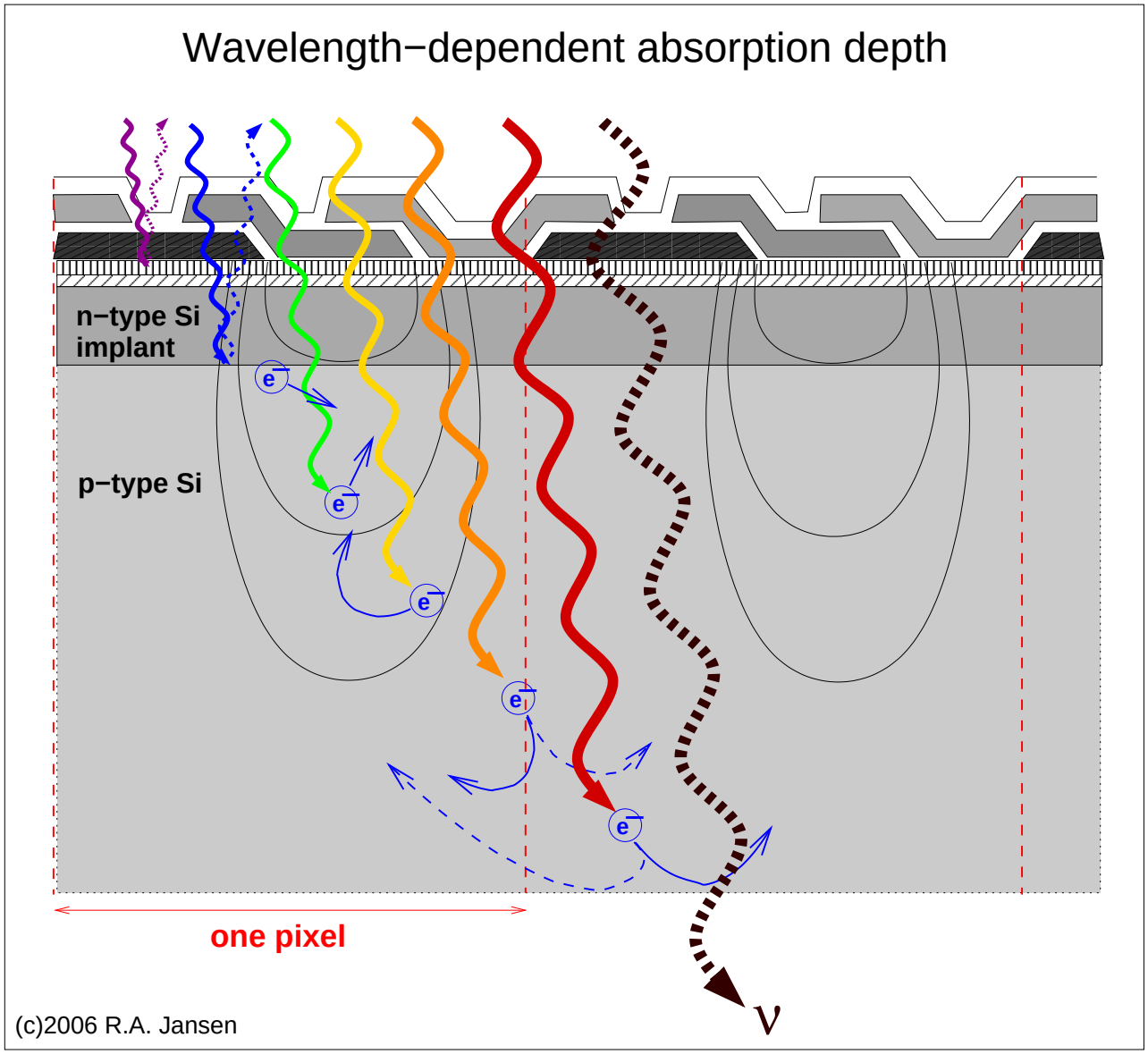


Figure 6: Implications of the wavelength-dependence of the photon absorption depth for a front-illuminated CCD.

Characterization of CCDs (cont'd)

- **Pure silicon is shiny like a metal and a fairly good *reflector* of light, particularly at bluer colors (see Fig. 7). In addition to the small absorption length, bluer photons are more likely to be reflected off of the CCD surface.**
  - ***Anti-reflection coatings*** **such as MgF_2 can greatly reduce these losses.**

Figure 7: The lower of the above CCDs has had a coating of MgF_2 applied, while the upper chip is thinned but still without any coating. *Courtesy: M. Lesser/UA/ITL.*

Characterization of CCDs (cont'd)

▷ ***Quantum efficiency*** **(QE) — the (wavelength-dependent) ratio of incoming to actually detected photons, usually expressed as a percentage.**

- **The QE of a CCD is better at higher operating temperatures, particularly for $\lambda \gtrsim 800$ nm.**
  - **However, since raising the temperature means increasing the thermal dark current and read-noise (see also Fig. 10), there is very little net gain in doing so.**
- **With the purity of the Si used to fabricate a CCD, its resistivity goes up, allowing higher gate voltages, and hence larger electron storage capacity.**
- **Since higher voltages also mean *deeper* potential wells at larger distances from the gates, the probability of confining photo-electrons produced by near-IR (front-) and blue (back-illuminated CCDs) photons increases: i.e., better QE at the reddest(bluest) wavelengths.**
  - **The price to pay is, indeed, the much higher price of fabrication of such a pure device.**

▷ **Tuning the parameters of a CCD *always* means finding a compromise between conflicting constraints.**

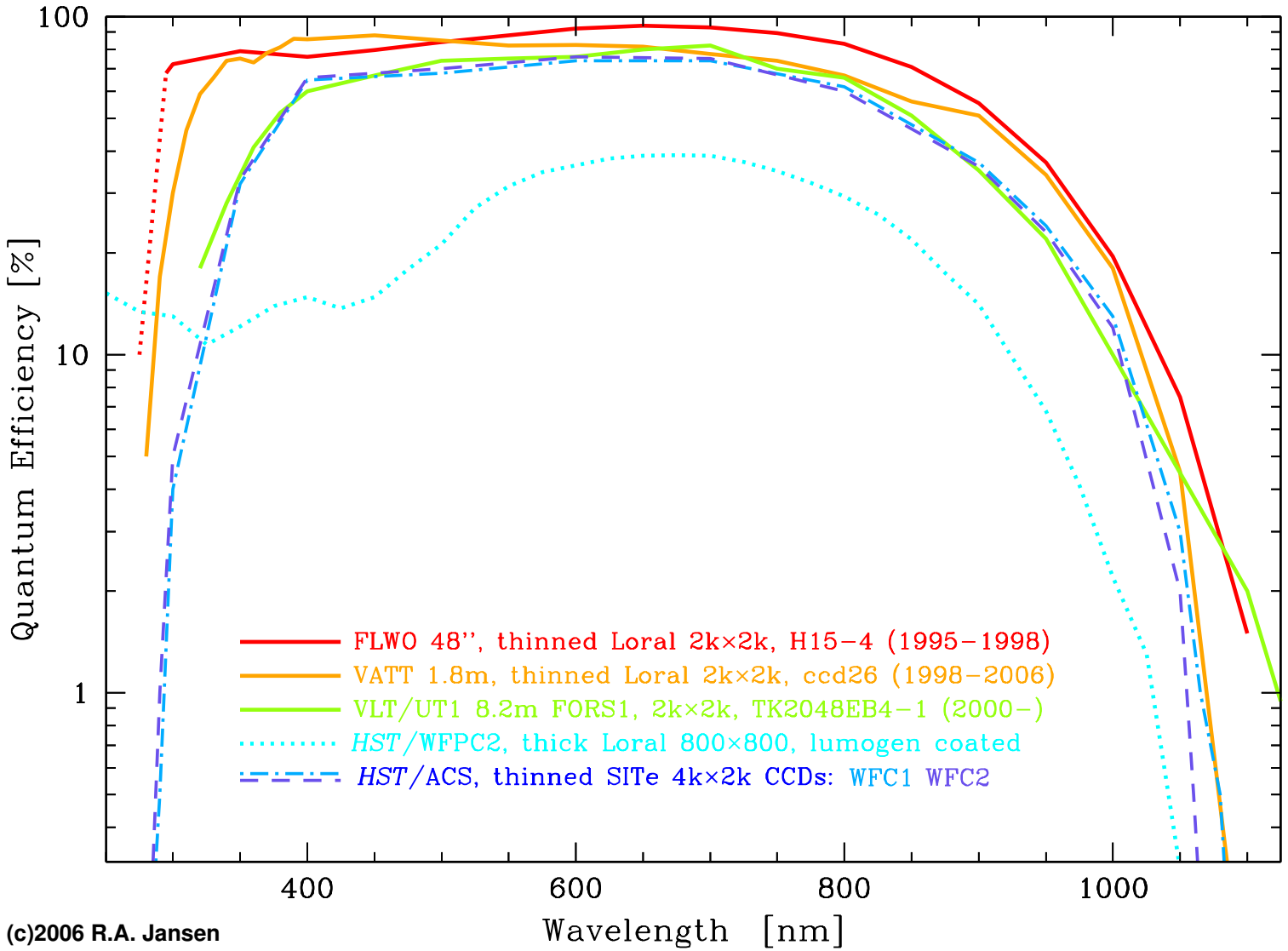


Figure 8: Comparison of QE curves for several CCDs. Most of these detectors are of the same generation. The 800×800 pixel front-illuminated *HST*/WFPC2 CCDs are older; their high UV response is due entirely to a lumogen coating, which converts UV–blue photons to green ones to which these CCDs are sensitive.

Characterization of CCDs (cont'd)

- **Measured QE curves (e.g., Fig. 8) are assumed to be representative *of each and every pixel* of the CCD.**

  - **This is not quite so, although often a reasonable approximation. One always has to correct for pixel-to-pixel variations in the reponse.**

- **The QE curve delivered with a particular CCD may only be representative of *a typical device of that kind*, but not be a measurement of the wavelength dependent response of that device itself.**

▷ **Hence, to calibrate observations of an astronomical object of interest through a given filter, one also needs to observe *photometric standard stars* (e.g., Landolt 1992; Stetson 2000), or *spectrophotometric standard stars* (e.g., Oke 1990; Hamuy et al. 1992,1994).**

  - **The apparent brightness of such standard stars — corrected for Earth's atmosphere — is accurately known either in several standard passbands (see: Bessel 1990), or as a function of wavelength in small wavelength intervals.**

Charge Transfer Efficiency and Deferred Charge

▷ ***Charge Transfer Efficiency*** **(CTE) — The fraction of the total charge accumulated in an arbitrary pixel that gets successfully passed on to the next pixel.**

▷ **The *Charge Transfer Inefficiency* (CTI), which quantifies the charge losses, is simply the complement of CTE, i.e., $1 -$ CTE.**

- **The *CTE* of modern CCDs is typically better than 0.99999, but never exactly 1; the charge losses accumulate the further a given pixel is removed from the read-out amplifier.**

  - **In a $n \times m$ pixel CCD, the charge of the last pixel to be read out, will be transferred along its column (parallel shifts) a total of m times, and along the serial output register (serial shifts) a total of n times. If n and m are large (as in most modern astronomical detectors), the CTE needs to be near-perfect.**

  - **The parallel CTE ($\mathcal{C}_p$) need not be identical to the serial CTE ($\mathcal{C}_s$), since the pixel sizes, gate structures, and voltages of the serial register differ from those in the imaging area.**

Charge Transfer Efficiency and Deferred Charge (cont'd)

If N_e- denotes the total number of electrons (total charge) stored in that final pixel, then the number of electrons that will actually arrive at the amplifier will be:

$$N'_e\text{-} = N_e\text{-} \cdot \mathcal{C}_p^m \cdot \mathcal{C}_s^n$$

Hence, the charge lost, L_e-, is given by:

$$L_e\text{-} = N_e\text{-} - N'_e\text{-} = N_e\text{-} \left(1 - \mathcal{C}_p^n \, \mathcal{C}_s^m\right)$$

△ ***Note that the expression given in Howell (2006) on page 44 (2nd edition) is incorrect!***

- ***Class problem*** **(put your smart phones, tablets, or laptops to work) — Assume that you have a modern 2048×2048 pixel CCD with a CTE of 0.999996, and that parallel and serial transfers *are* equally efficient. What fraction of the photo-electrons generated in the last pixel to be read-out reaches the pre-amplifier? (Or what fraction of electrons have you lost along the way?) And what if we had an older CCD with a CTE of only 0.99995 ?**

Charge Transfer Efficiency and Deferred Charge (cont'd)

- **Charge lost during a single transfer need not necessarilly be permanently lost. A photo-electron left behind, joins the charge that will be transferred from the next pixel in the same column, and is *indistinguishable* from that charge (electrons are all alike). Upon a subsequent clocking cycle, it again has a probability of value $\mathcal{C}_p$ to be transferred.**

▷ ***Deferred charge*** **— photo-electrons that fail to be transferred with the charge packet of the pixel in which they were generated, but that are transferred in subsequent transfers.**

- **Poor CTE in combination with deferred charge results in charge trails along columns behind bright sources.**

- **Prolonged exposure to ionizing radiation, and in particular bombardment with cosmic rays, results in permanent damage to the bulk silicon lattice within CCDs. Such damage results in charge traps and a general degradation of the CTE with time (see Fig. 9), and can also exacerbate charge diffusion.**

  ○ **CCDs in space age on time scales of only a few years. Also, coatings on the CCD may evaporate and degrade in the high vacuum of space.**

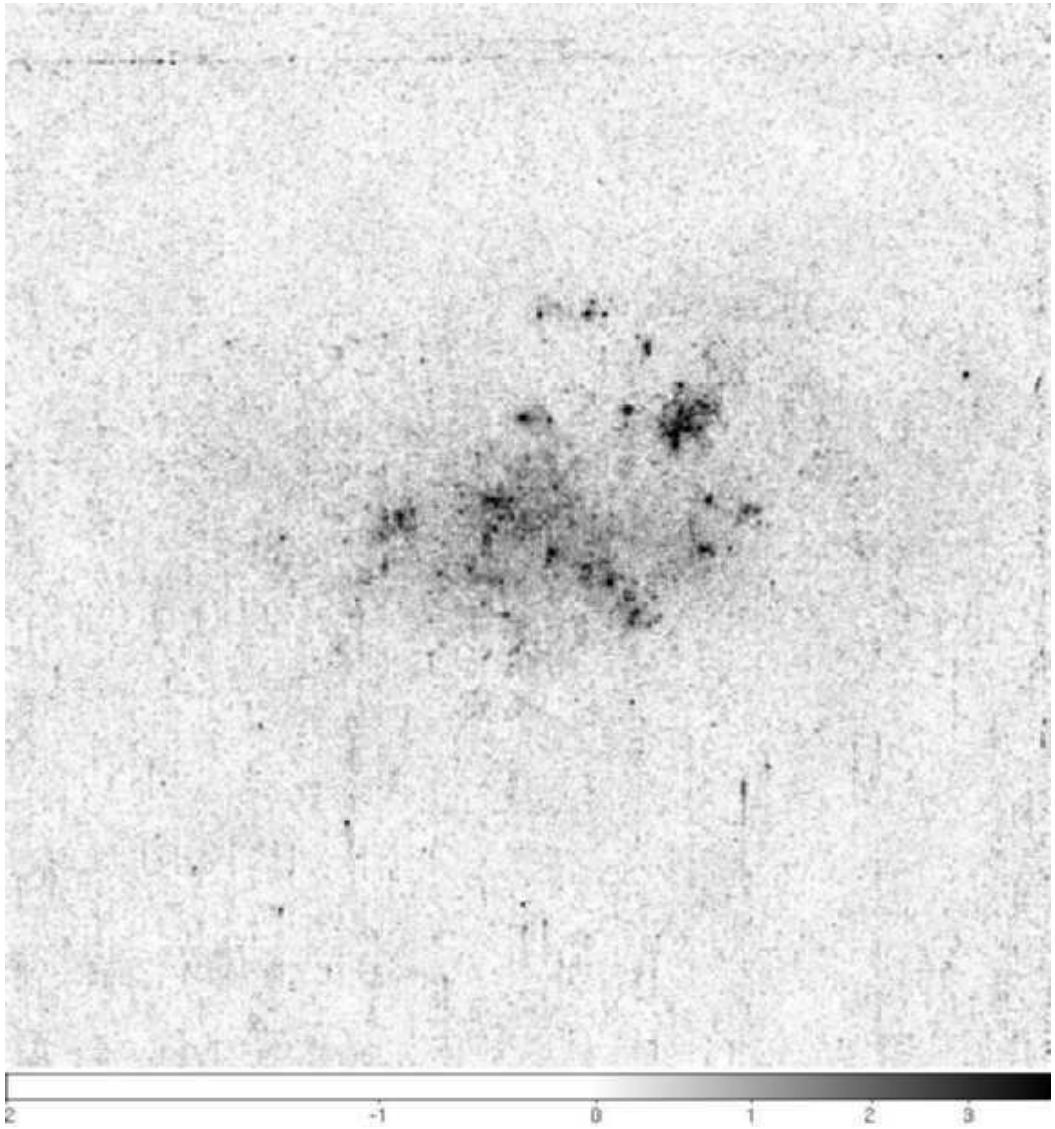


Figure 9: Example of the poor CTE and deferred charge trails behind bright sources that result from prolonged exposure to damaging particle radiation. The above near-UV image of NGC 14 was taken with *HST*/WFPC2 after its CCDs had been in space for over 7 years and had considerably degraded with respect to their performance just after launch.

Dark Current

- **Every material at T $\gg$ 0 K will be subject to thermal noise within. CCD *dark current* is due to electrons freed from the valence band and collected within the potential well of a pixel, where they become indistinguishable from the photo-electrons.**

- **The dark rate depends strongly on CCD temperature (see Fig. 10). For modern CCDs, operated at $T \sim$150–170 K, the dark rate is typically $<$10 e^-/pix/hr. The dark current would be lower if a CCD were to be operated at a lower temperature, but then the QE in the red and the CTE would also be lower.**

- **The dark signal also depends on the gate voltages and, for a given resistivity of the bulk Si substrate, will increase at higher voltages.**

- **The dark signal builds up *linearly* with exposure time. Hence, long exposures likely need to be corrected, while dark signal in short exposures may well be negligible if the dark rate is low.**

- **The distribution of the dark signal tends to satisfy Poissonian statistics.**

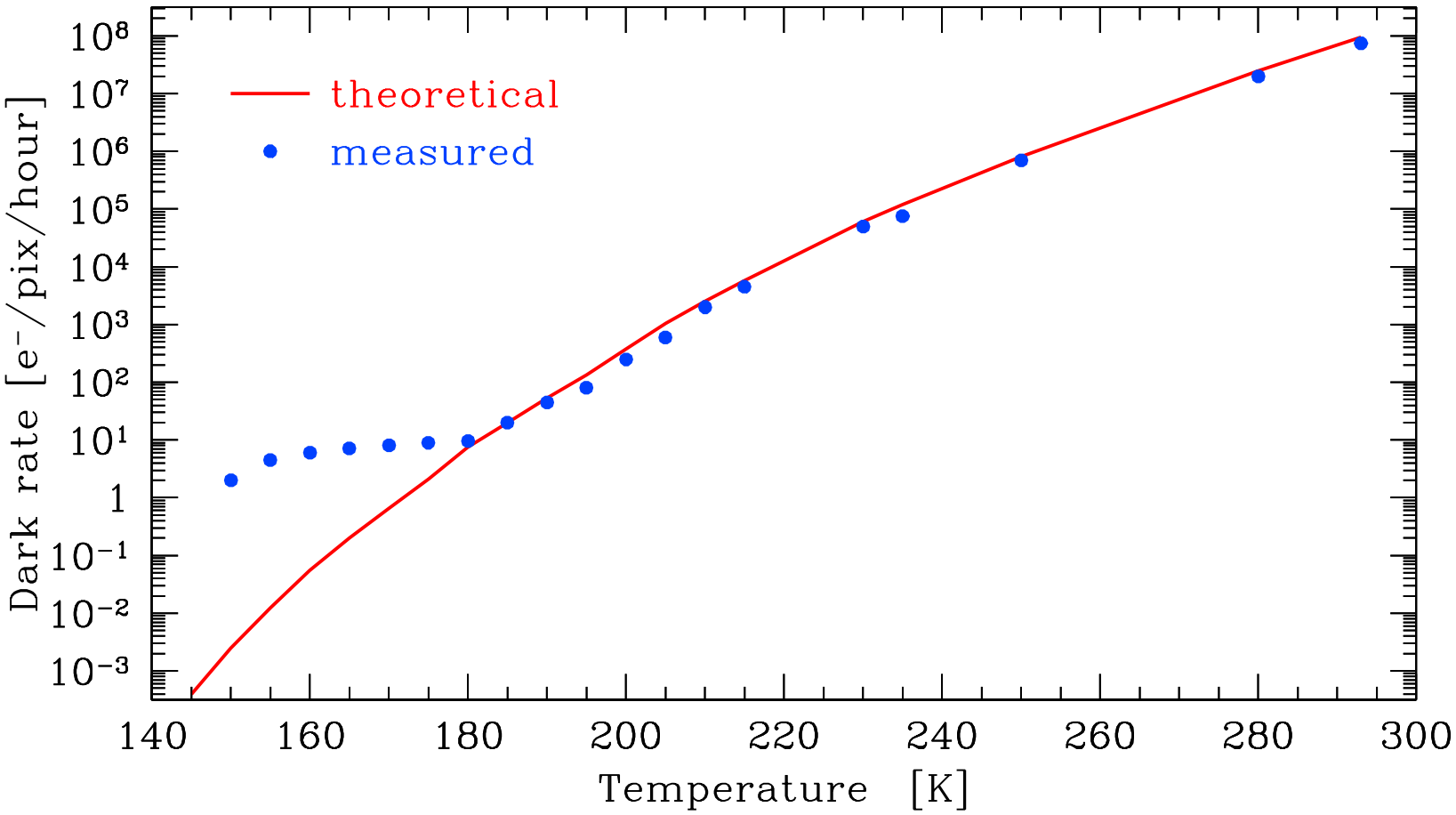


Figure 10: Measured and expected temperature dependence of the dark rate for a thinned back-illuminated 2k×4k EEV CCD delivered to ESO. Below ∼180 K, the measured dark rate does not diminish as expected due to other residual effects. Nonetheless, this particular device reaches a dark current of only $\sim\mathbf{2}\, e^-$/pix/hour when operated at 150 K.

Bias and Overscan

- **To ensure the A/D converter will never be presented with a negative number for the charge estimate — which could easily occur due to read-noise when the absorbed number of photons in a given pixel is small or zero — a sufficiently large *offset* or *bias* voltage is added at the amplifier stage.**

- **This *bias* is *independent* of the length of the integration of an exposure and is theoretically constant for all pixels.**

  - **In reality, there may be some (generally low-level) structure that varies from pixel to pixel in a repeatable fashion (Fig. 11).**

  - **On long time scales — particularly when cryogenic cooling is not maintained when an instrument is not in use — the bias pattern may show variations.**

  - **If patterns of sufficient strength are discernable and contribute measurably to the noise, then a sufficiently large number of *bias frames* — zero-length exposures with the shutter closed — should be taken such that: $\mathcal{R}/\sqrt{N_b} \ll \Delta\mathcal{S}_b$, with $\mathcal{R}$ the read-noise, N_b the number of bias frames, and $\Delta\mathcal{S}_b$ the peak-to-peak pattern strength.**

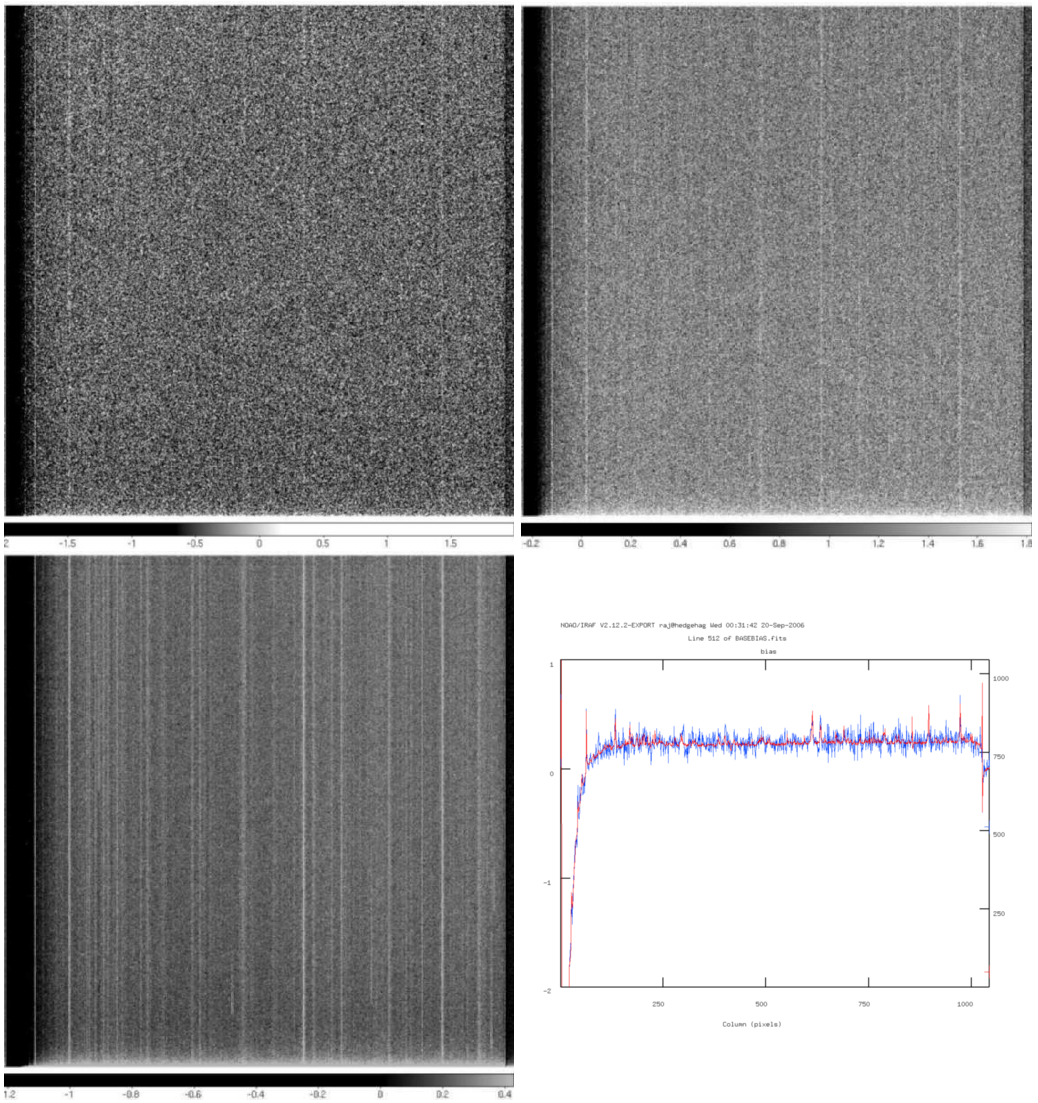


Figure 11: (a) Median average of 36, (b) of 128, (c) of 3500 bias frames obtained at the Vatican Advanced Technology Telescope (VATT) in Nov 2001. (d) A cross-cut along a single row (blue) and averaged over all rows (red) of the baseline bias reference frame of (c).

Bias and Overscan (cont'd)

- **Preceding read-out and after read-out of all pixels of the CCD has finished, the read-out electronics will perform several additional (usually 8–40) clocking cycles that do not correspond to any physical pixels.**
  - **These cycles will, however, exercise the amplifier and present a charge estimate corresponding to the instantaneous bias level as modulated by the read-noise to the A/D converter.**

▷ **These *virtual pixels* are referred to as the *overscan*, or *overscan strip(s)*. The digital numbers corresponding to these virtual pixel columns will be written to disk as if they were integral part of the CCD image.**

- **In the subsequent data reduction, the overscan strip(s) will be used to subtract the mean level and possibly any large-scale gradient in that level from each CCD frame.**

- **The bias level often displays small changes over time, e.g., as the temperature of the warm electronics changes during a night. These changes are tracked by the level measured in the overscan strip(s).**

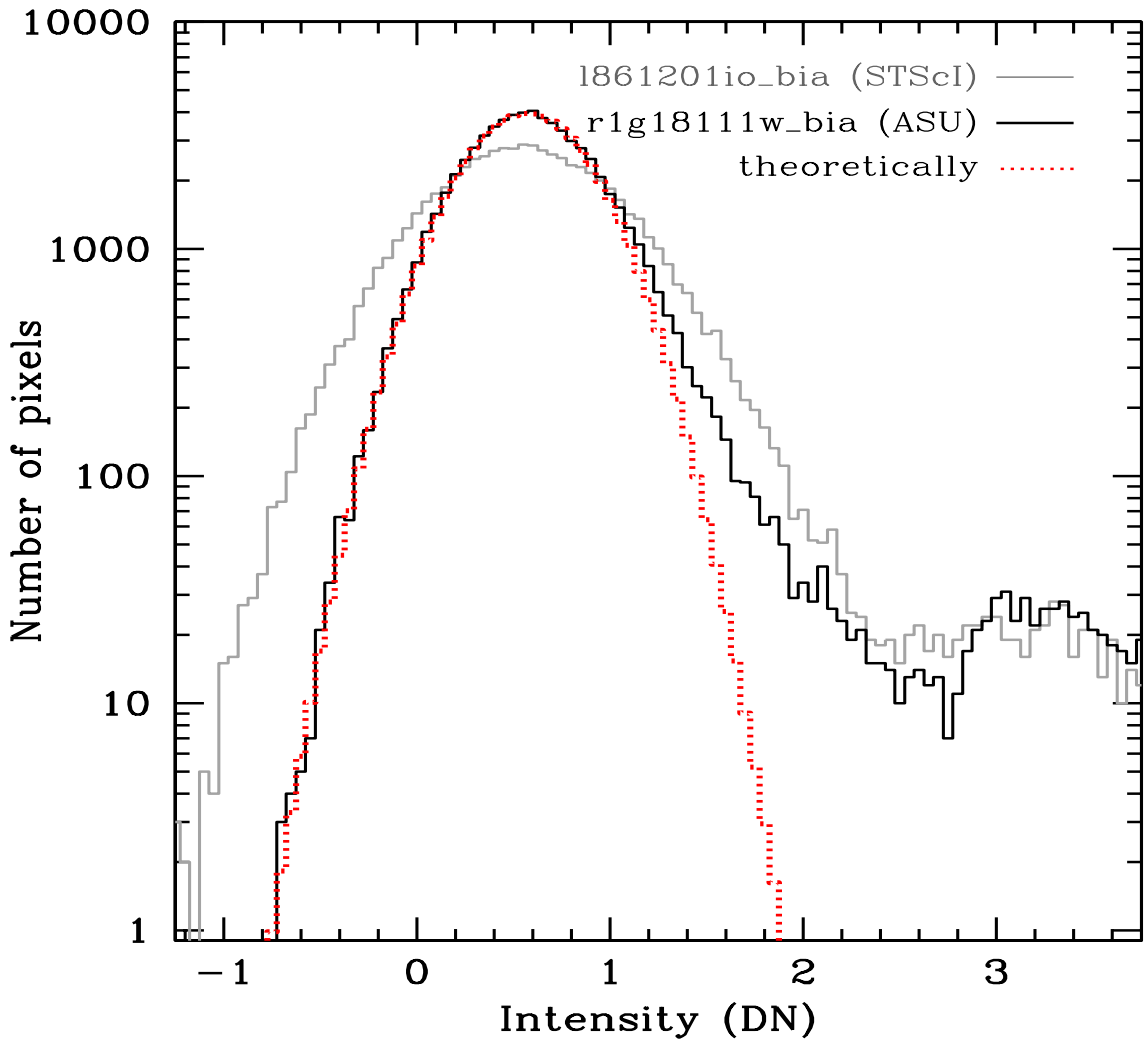


Figure 12: Pixel histogram of an *HST*/STIS bias reference frame. In theory, the pixel values should show a Gaussian distribution that reflects the read-noise (divided by $\sqrt{N_b}$). Columns with higher than average bias level, and warm and hot pixels, result in a tail towards higher pixel values (*right*), but the lower half (*left*) of the histogram should be unaffected. A re-calibration program (labeled *ASU*; Jansen et al. 2002) resulted in a marked improvement over the bias reference frame from the *HST* Archive (*STScI*).

Bias and Overscan (cont'd)

- **Some CCD designs also have some rows of physical pixels that are shielded from light. These are termed the *physical overscan*, and can be used to diagnose CTE and deferred charge, or other problems with the CCD electronics (e.g., reset voltages).**

Gain and Read-Noise

- **When Silicon (i.e., a CCD) is illuminated by the 5.9 keV Kα line of a radio-active ^{55}Fe source, the charge deposited by a single energetic X-ray photon in each pixel is 1620 e^- (an extreme case of the creation of multiple electron–hole pairs). This allows direct measurement of the gain factor, $\mathcal{G}$, of the A/D conversion to a precision of better than 1% as:**

$$\mathcal{G} = 1620/\mathrm{ADU} \quad [e^-/\mathrm{ADU}]$$

  - **In practice, this is rarely done this way.**

- **There is another method, *Janesick's method*, where one uses the property that the noise in an image in which the signal level is high should satisfy *Poissonian statistics*. In that case, *photon* or *shot noise* is much larger than the read-noise, and signal, noise and gain ought to satisfy (see Fig. 13):**

$$\mathcal{S}_{ADU} \equiv \frac{N_{e^-}}{\mathcal{G}} \quad (1) \qquad \text{and} \qquad \sigma_{\mathcal{S}_{ADU}} = \frac{\sqrt{N_{e^-}}}{\mathcal{G}} \longrightarrow \sigma^2_{\mathcal{S}_{ADU}} = \frac{N_{e^-}}{\mathcal{G}^2} \quad (2)$$

and, hence, combining (1) and (2): $$\frac{\sigma^2_{\mathcal{S}_{ADU}}}{\mathcal{S}_{ADU}} = \frac{1}{\mathcal{G}} \longrightarrow \mathcal{G} = \frac{\mathcal{S}_{ADU}}{\sigma^2_{\mathcal{S}_{ADU}}}$$

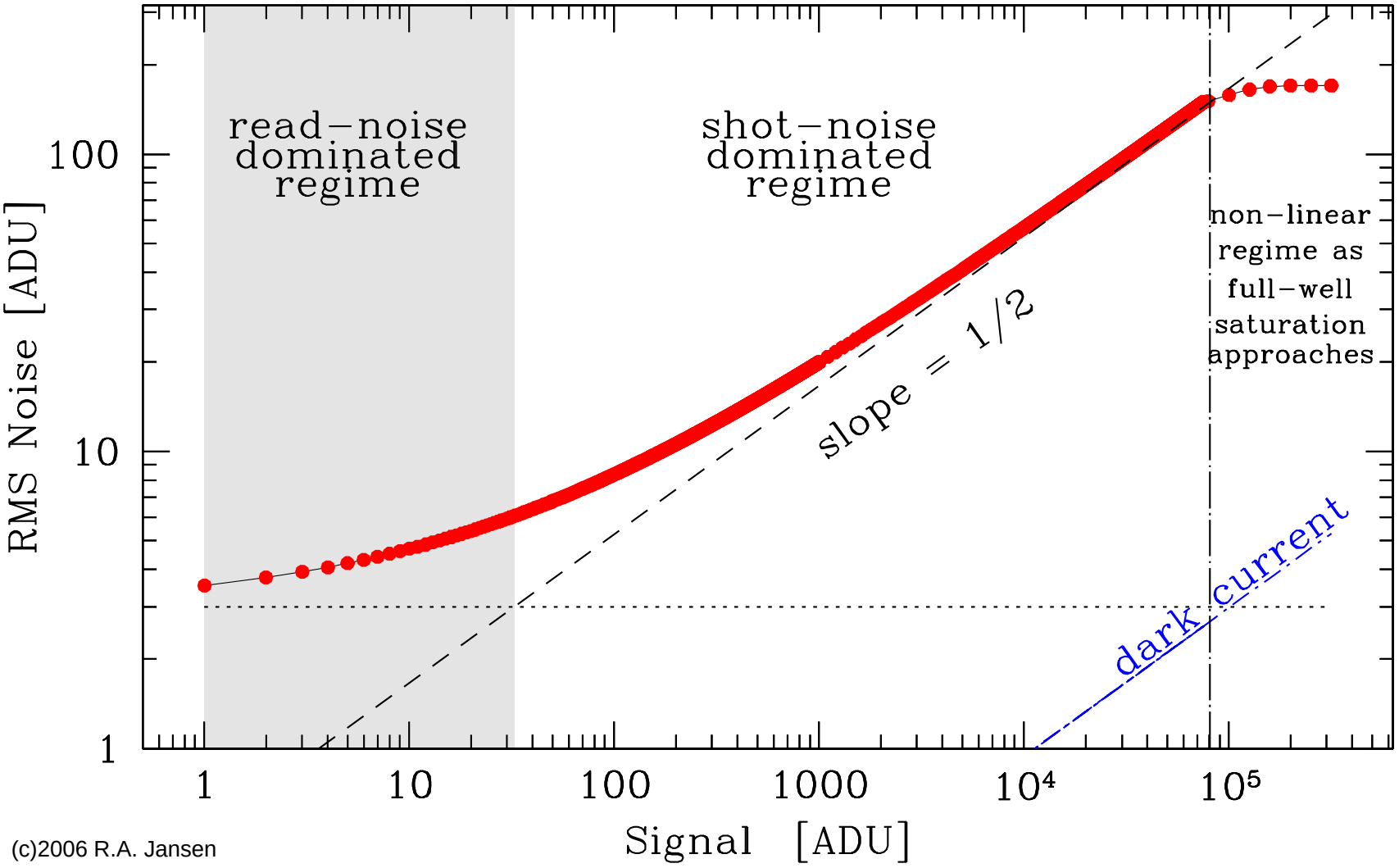


Figure 13: RMS noise as a function of signal level in a particular CCD. The read-noise sets a noise floor, to which photon noise (shot noise) increasingly adds. Once shot noise exceeds read-noise, the data are said to be shot noise limited. In the shot noise limited regime, the relation between noise and signal assymptotically approaches a slope of 1/2. If full-well capacity can be reached before A/D saturation occurs, then at very high signal levels the slope begins to deviate from 1/2.

Gain and Read-Noise (cont'd)

- Of course, the assumption that any noise contributed by the detector itself or by the read-out electronics is negligible, is only an approximation.

- We can take such noise explicitly into account by comparing a zero-length dark exposure (a *bias frame*), where we should only measure read-noise, with an exposure with a (fairly) uniformly high signal level, like a *dome*, *twilight* or *pupil flat* (see, e.g., Fig. 14). Since large-scale variations in signal level and pixel-to-pixel sensitivity variations would make our estimate uncertain, the trick is to use the differences of *pairs* of each, where such variations drop out.

  Janesick's method will yield estimates of both gain and read-noise simultaneously, as follows. If F_1 and F_2 represent a pair of (fairly) evenly illuminated frames with plenty of signal, and if B_1 and B_2 denote a pair of bias frames, then:

$$\Delta_F = F_1 - F_2 \qquad \text{and} \qquad \Delta_B = B_1 - B_2$$

$$\mathcal{G} = \frac{(\overline{F_1} + \overline{F_2}) - (\overline{B_1} + \overline{B_2})}{(\sigma^2_{\Delta_F} - \sigma^2_{\Delta_B})} \quad [e^-/\text{ADU}] \qquad \text{and} \qquad \mathcal{R} = \mathcal{G} \cdot \frac{\sigma_{\Delta_B}}{\sqrt{2}} \quad [e^-]$$

  Where $\mathcal{G}$ and $\mathcal{R}$ are the gain and read-noise, averaged over all pixels of the CCD, and $\overline{X}$ denotes the mean level of frame X.

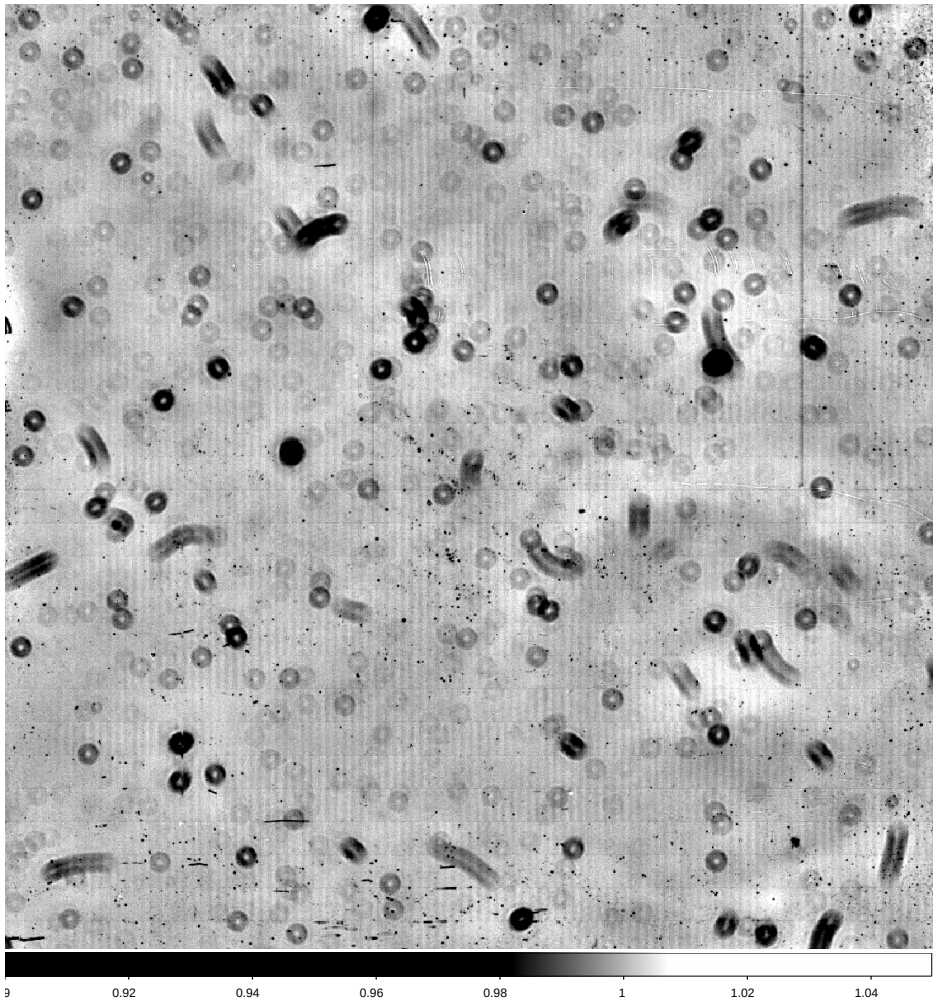


Figure 14: Example of a normalized median average of 60 bias-subtracted *R*-filter twilight flat frames, from the VATT CCD camera. The range between black and white corresponds to only a $\sim$3% difference in effective response (QE). Visible are both pixel-to-pixel variations intrinsic to the CCD, and variations extrinsic to it. To the first category belong the pixel-to-pixel variations in the response that form the regular vertical and horizontal bands ($\pm$1%), pixels and small clusters of pixels with much lower than average response ($\lesssim$**90**%), and a couple of bad partial columns ($\lesssim$**50**%). To the second category belong the larger-scale variations due to dust particles that have settled onto the CCD dewar window, or on the filter (at the $\sim$50–99.9% level). If dust particles remain stationary, division of all bias-subtracted science exposures by this normalized flat frame will correct for the variations in effective QE.

Saturation and Linearity

▷ ***Saturation*** **can set in two different ways:**

- **Pixel *full-well saturation* occurs when the pixel potential well can no longer confine photo-electrons to the pixel where the photon was absorbed.**

  ○ **Shortly before full-well saturation sets in, the electrons already stored in the pixel lower the potential well to the point where additional photo-electrons may migrate deep into the substrate or even out of the pixel, resulting in a non-linear response of that particular pixel (see Fig. 15).**

- ***A/D saturation* occurs abruptly when the charge estimate presented for conversion exceeds the value that can be represented with the available number of bits per pixel.**

  ○ **That level for an n-bit un-signed integer is $2^n - 1$ (assuming we start counting at 0). Hence, for a 12-bit A/D converter (e.g., *HST*/WFPC2) the maximum pixel value is $2^{12} - 1 = 4095$, for typical modern 15 or 16-bit A/D converters it would be 32767 or 65535, respectively.**

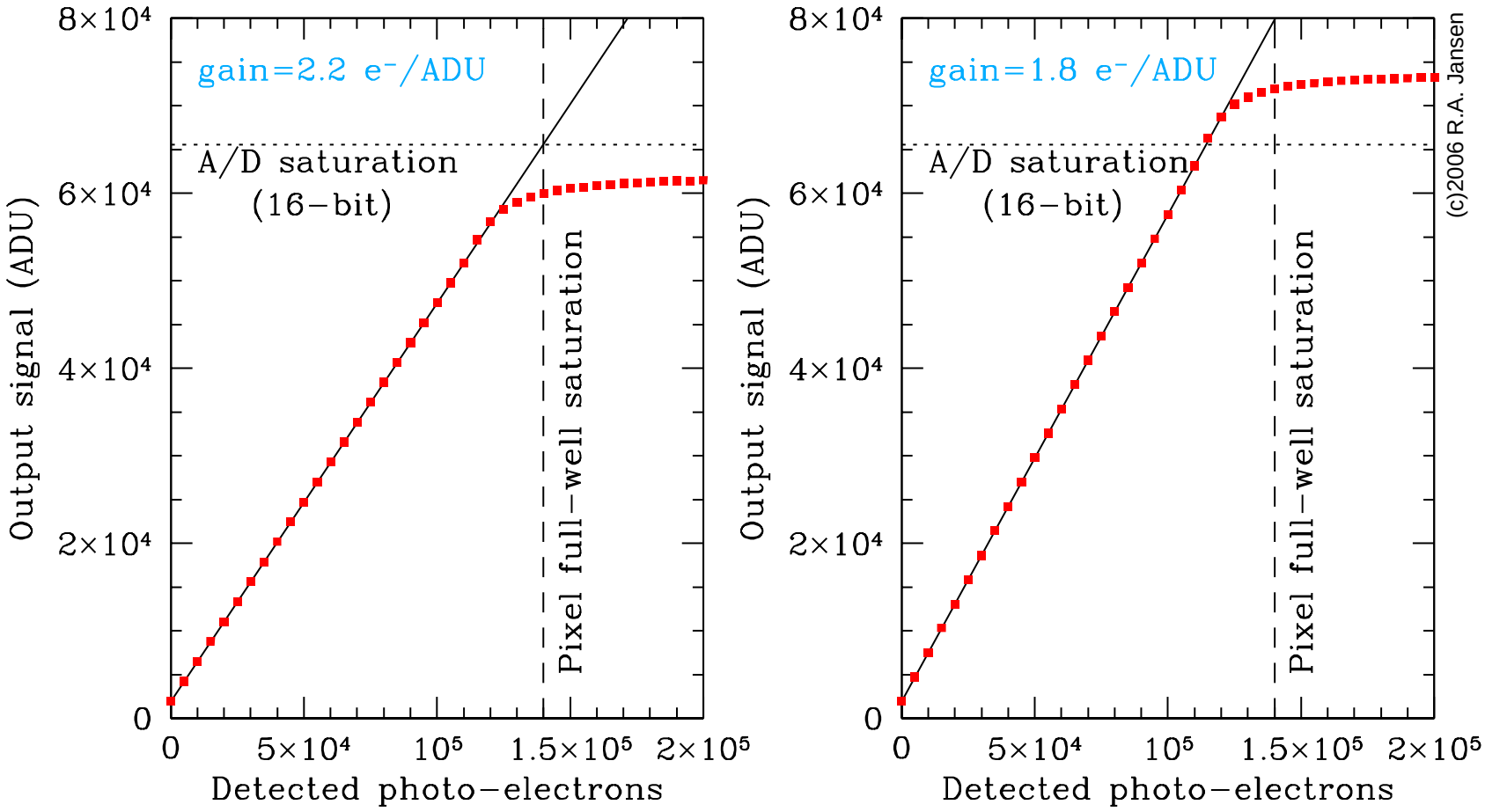


Figure 15: CCD linearity curves of a typical three-phase CCD for two different gain values. The device is linear over most of the output range from the bias level (the offset of 2000 ADU for 0 detected photo-electrons) to (*left*) ∼58,000 ADU (∼127,500 e^-), where the CCD response becomes non-linear as the pixel full-well capacity of 140,000 e^- is approached, and (*right*) 65,535 ADU, set by the saturation of the 16-bit A/D converter. In both cases, the slope of the linear part of the response curve is equal to $1/\mathcal{G}$. The gain is usually hard-wired into the CCD electronics.

Saturation and Linearity (cont'd)

- **Even when the CCD response to photons is perfectly linear at the stage of photon absorption, *and* when storing and passing along photo-electrons, the conversion to a digital data number will result in a non-linear behavior if the gain is (significantly) larger than 1. This *differential non-linearity*, better known as *digitization noise*, is illustrated in Fig. 16*a*.**

  - **The main gain value of *HST*/WFPC2 was 7 e^-/ADU, which is *larger* than the read-noise of $\sim$5.3 e^- of the CCDs. Combined with the low sky background levels from space, digitization is a much more important source of noise for WFPC2 images than common in ground-based CCD images: data values below $\sim$10 ADU tend to be affected, and also result in pixel distributions that satisfy neither Gaussian nor Poissonian statistics.**

    △ ***WFPC2 sky background values computed in UV–blue or narrow-band filter images by taking the median pixel value will be systematically in error. Median filtering to reject outlying pixels values will also be affected.***

  - **As read-noise of CCDs has come down over the years (from $>$300 e^- to typically less than 10 e^- today), gain values had to come down as well, with photon-counting set-ups ($\mathcal{G} = 1$) far more common today. Digitization noise, therefore, tends to be less of a problem.**

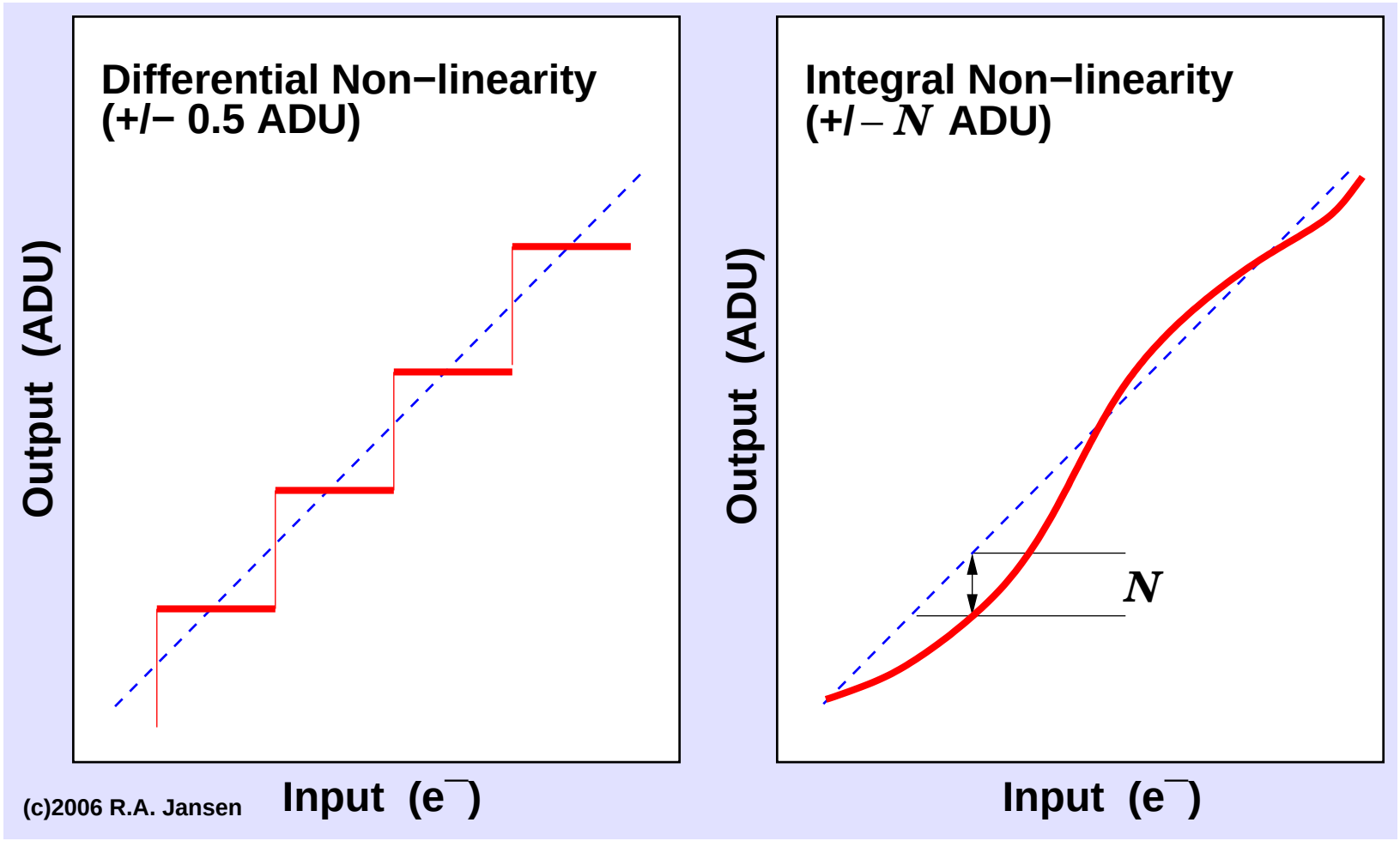


Figure 16: Illustration of the two types of CCD non-linearity. Differential non-linearity (*left*) results from the finite steps (when $\mathcal{G} > 1$) in the A/D conversion. The amplitude of this type of non-linearity is ± 0.5 ADU. Integral non-linearity (*right*) works over larger than single A/D step ranges, and can be quite complex. The magnitude of this type of non-linearity is given as the *maximum* deviation, N, with respect to a perfectly linear response.

Saturation and Linearity (cont'd)

- **A/D converters are not perfect. Some values are reported more (or less) frequently than they would be by a perfect device. This results in the second type of non-linearity, *integral non-linearity*, illustrated in Fig. 16*b*.**

  - **Although the "true" data number can never be recovered, one can *statistically* correct for this systematic behavior. Fortunately, in most modern A/D converters, the maximum deviation, N, from the true linear response is small.**

  - **The *HST*/WFPC2 A/D converters are known to be relatively well-behaved with $N \simeq 1.8$–2.0 ADU ($\sim$13–14 e^-) for bit 12, i.e., the one set for the 2048–4095 ADU range (Baggett et al. 2002), or an integral non-linearity of $\lesssim 0.2\%$.**

    △ **Note that the non-linearity described on page 58 and Fig. 3.9 of Howell (2006; 2nd Ed.) is an *A/D non-linearity* encountered at high signal levels in the particular device discussed there, and *not* the intrinsic non-linearity that might occur when the charge in a pixel approaches full-well capacity, shown in Fig. 15 !**

- **Some of the charge that is lost to a given pixel that is near full-well capacity may diffuse to neighboring pixels and be recovered for analysis when these pixels are read out. Only the fraction of photo-electrons that manage to recombine with holes deep within the substrate of a pixel are forever lost.**

Transient Effects: Cosmic Rays, Cross-talk

- ***Cosmic rays*** **can dump a significant amount of charge within one or more adjacent pixels, possibly saturating some, and often render the affected pixel(s) useless for subsequent scientific analysis.**

  - **Thick, front-illuminated CCDs tend to be more susceptible to cosmic rays than thinned ones. Although there are several methods to detect and remove the induced signal (e.g., `la_cosmic`, by P. van Dokkum 2001), and one may sometimes decide to simply interpolate over affected pixels, cosmic rays generally necessitate splitting the desired exposure time over two or more exposures and limit any single exposure to at most $\sim$30 min (too avoid hits in a large fraction of pixels, and thereby reduce the chance that the same pixel gets hit in subsequent exposures).**

- **An example of *cross-talk* in the read-out electronics, when a CCD is read through more than one amplifier is visible in Fig 17, where a saturated star in one half of the CCD resulted in a ghost image at the mirrored location within the other half.**

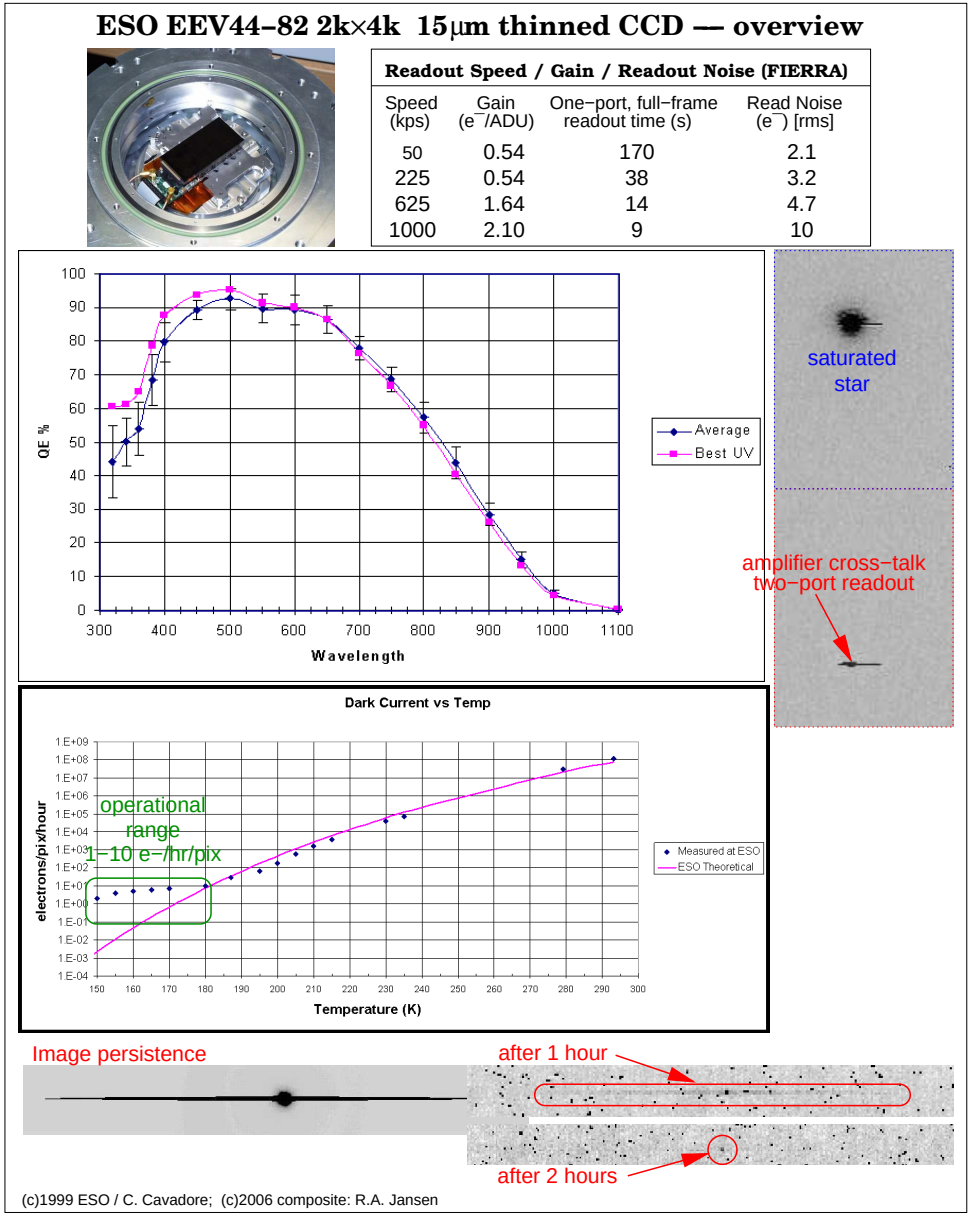


Figure 17: Performance overview of an EEV 44-82 2k×4k CCD delivered to the European Southern Observatory.

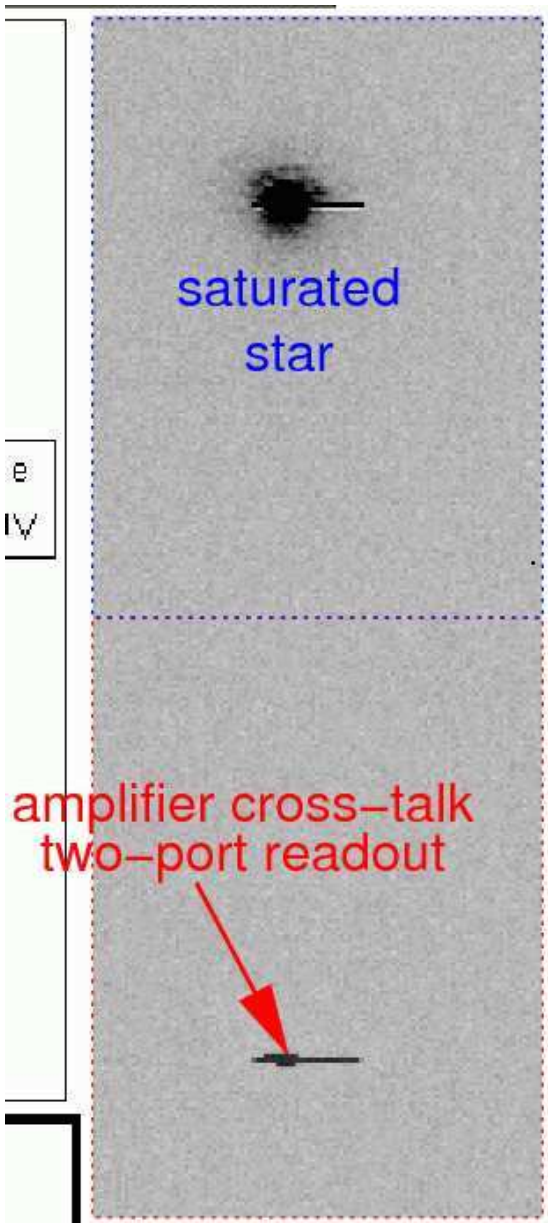


Figure 18: Example of crosstalk in an EEV 44-82 2k×4k CCD.

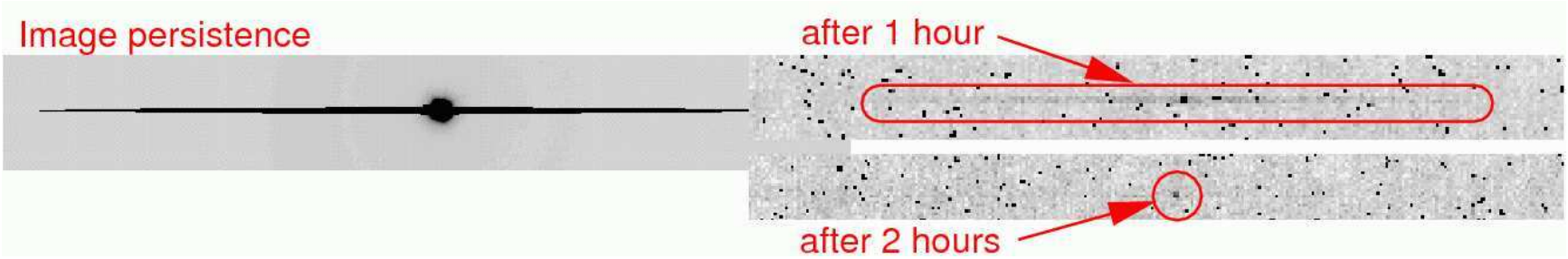


Figure 19: Example of persistence in an EEV 44-82 2k×4k CCD

Transient Effects: Persistence, Hot Pixels and Bad Columns

- **Overexposure of bright sources can result in electrons that persist deep within the bulk substrate. Over time they may slowly migrate back up, mimicking a strongly localized increase in the dark rate, or after-images of the bright source. The time constant for this *persistence* can be several hours (see Fig 17), and depends on the details of the CCD design and on how severe the overexposure was.**

- **As a result of defects introduced during manufacturing, or als a result of particle and radiation damage over time, some pixels will show a much larger dark current than the one characteristic for a particular CCD. Such pixels (or even entire pixel columns) are known as *warm* or *hot pixels*, depending on the severity. By warming up the device to room temperature (a process called *annealing*), a certain fraction of warm/hot pixels may return to normal.**

- **Unresponsive (dead) pixels and bad/dead columns typically also result at the manufacturing stage. Bad pixels can act as charge traps or shorts that drain charge away into the substrate or that prevent the transfer of charge from any pixels along a column located further away from the serial output register. As a result a (portion of a) column in the output image will appear dark. Modern science-grade CCDs show only few of these cosmetic defects.**

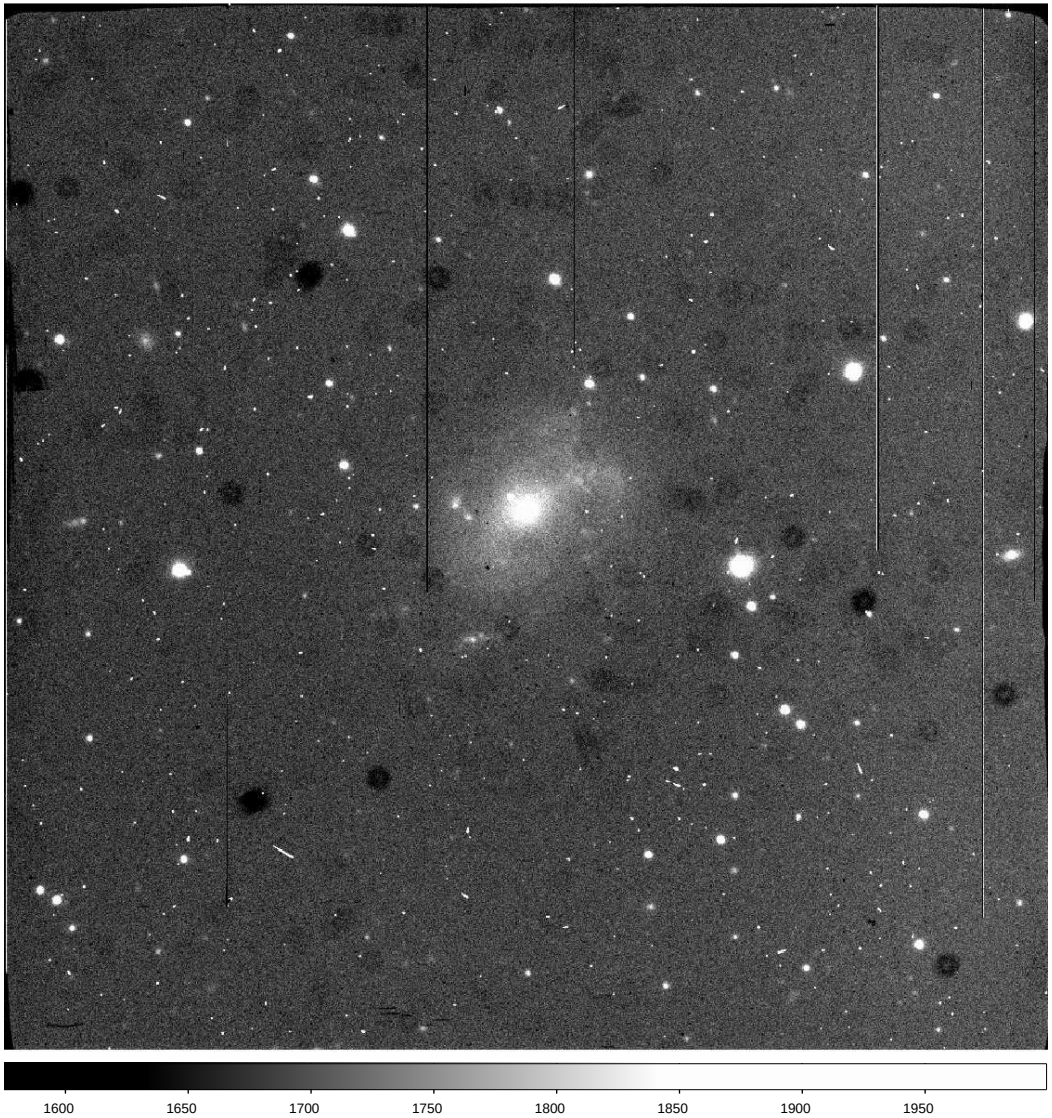


Figure 20: Raw 1200 sec Hα image of UGC 10445, obtained at the VATT in Sep 2003 (Jansen & Tamura), that features several of the transient effects and defects discussed. Several bad columns are seen (some hot, some dead), and the image is riddled with cosmic ray hits. Since cosmic ray hits do not bear the imprint of the PSF, they tend to be sharper than genuine point sources (e.g., stars) whenever the PSF is properly sampled. This example also demonstrates quite effectively why it is important to flat-field the data.

Astronomical Signals and Noise: Poisson vs. Gauss

- **Read-noise closely follows a *Gaussian* or *normal distribution* associated with sampling measurements in a near-equilibrium electronic circuit. *Shot noise*, i.e., the noise associated with the stochastic arrival of discrete energy or charge packets instead follows a *Poisson distribution*.**

  An important property of the Poisson distribution is, that its *standard deviation* equals the square root of the average measured number (i.e, detected photo-electrons in our case). For an average score of m, the probability of measuring n when the values of n are distributed according to a Poisson distribution is given by:

$$P_n = \frac{m^n \cdot e^{-m}}{n!} \qquad \text{, with } m = \text{constant}$$

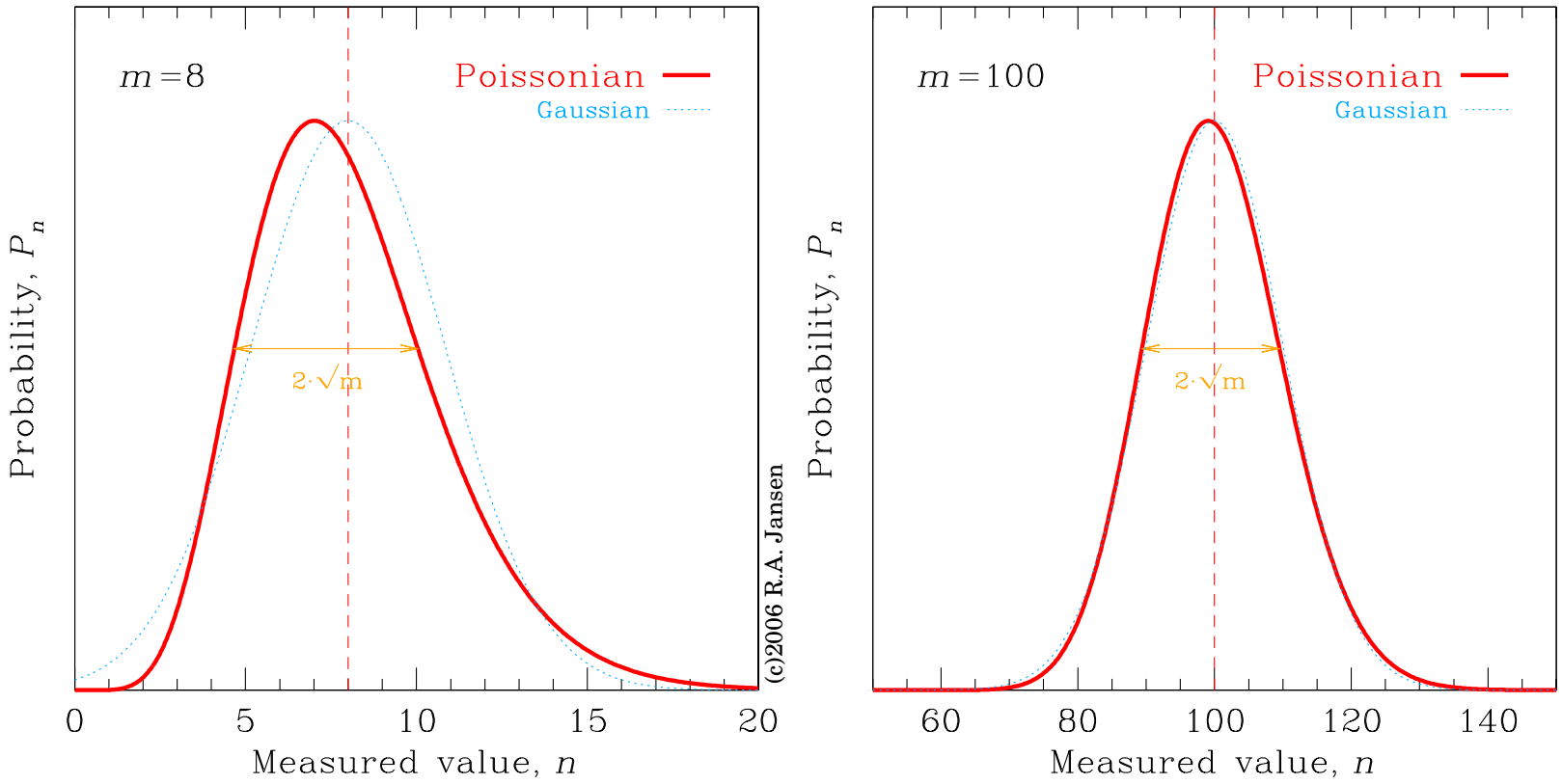


Figure 21: Examples of Poisson distributions. (*left*) For $m = 8$, the distribution is markedly skewed compared to a Gaussian distribution with the same average score m. An important property of the Poisson distribution is that the interval between $-1\,\sigma$ and $+1\,\sigma$ is exactly equal to twice the $\sqrt{m}$. (*right*) For large m, such as $m = 100$ in this example, the Poisson distribution becomes indistinguishable from the normal distribution. *Note: for visualization purposes we here plot the mathematical extension of the discrete Poisson statistics (integer n) to continuous functions!*

First, let us verify that the sum of the probabilities over all n is indeed equal to 1, and that the average score, $\bar{n}$, is indeed equal to m:

$$\sum_n P_n \equiv \sum_n \frac{m^n \cdot e^{-m}}{n!} = e^{-m} \cdot \sum_n \frac{m^n}{n!} = e^{-m} \cdot e^m = 1$$

The average of n scores is given by:

$$\bar{n} \equiv \frac{\sum_n P_n \cdot n}{\sum_n P_n} = \sum_n \frac{m^n \cdot e^{-m}}{n!} \cdot n = m \cdot \sum_n \frac{m^{n-1} \cdot e^{-m}}{(n-1)!} = m$$

Lastly, the variance can be computed as:

$$\begin{aligned}
\sigma^2 \equiv \frac{\sum_n \left(P_n \cdot (n-\bar{n})^2\right)}{\sum_n P_n} &= \sum_n P_n \cdot (n^2 - 2nm + m^2) \\
&= \sum_n P_n \cdot n^2 - 2m \cdot \sum_n P_n \cdot n + m^2 = -m^2 + \sum_n \frac{m^n \cdot e^{-m}}{n!} \cdot n^2 \\
&= -m^2 + m\, e^{-m} \cdot \sum_n \frac{m^{n-1}}{(n-1)!} \cdot n = -m^2 + m\, e^{-m} \cdot \sum_l \frac{m^l}{l!} \cdot (l+1) \\
&= -m^2 + m \cdot \left[\left(\sum_l \frac{m^l \cdot e^{-m}}{l!} \cdot l\right) + \left(\sum_l \frac{m^l \cdot e^{-m}}{l!}\right)\right] \\
&= -m^2 + m^2 + m = m
\end{aligned}$$

And, hence, we find that for the Poisson distribution the standard deviation is equal to the square root of the average score: $\sigma = \sqrt{m}$.

Astronomical Signals and Noise: Poisson vs. Gauss

- **When we talk about an $x\,\sigma$ deviation or an $y\,\sigma$ result, we mean a deviation of x times the standard deviation σ (i.e, $x\sqrt{m}$), or a result that is significant at the probability of 1 minus that of obtaining a y times σ deviation purely by chance:**

$$1 - P\left(n < y\sqrt{m}\right) \qquad \textbf{or} \qquad 1 - P\left(|n| < y\sqrt{m}\right)$$

Signal and Noise: The CCD Equation

- **In science, publishing a measured value is *meaningless* unless the associated *measurement uncertainty* (also commonly referred to as the *error* on that measurement) is given as well.**
  - **To avoid *round-off errors* in subsequent computations, one may want to specify one digit more than strictly merited.**

- **In order to obtain new observations with a telescope, you have to write an observing proposal. A proposal without a thorough discussion of (a) the minimum *signal-to-noise ratio* (S/N) required for a definite result, and (b) the S/N that a particular configuration of telescope + instrument will give you in a given amount of time, is rarely awarded time.**

- **It should, therefore, be clear that an evaluation of the measured (or expected) S/N is an important topic to address. Also, by comparing the theoretically expected S/N with that actually obtained (e.g., Fig 22), one may catch problems (i.e, user errors) in the data processing.**

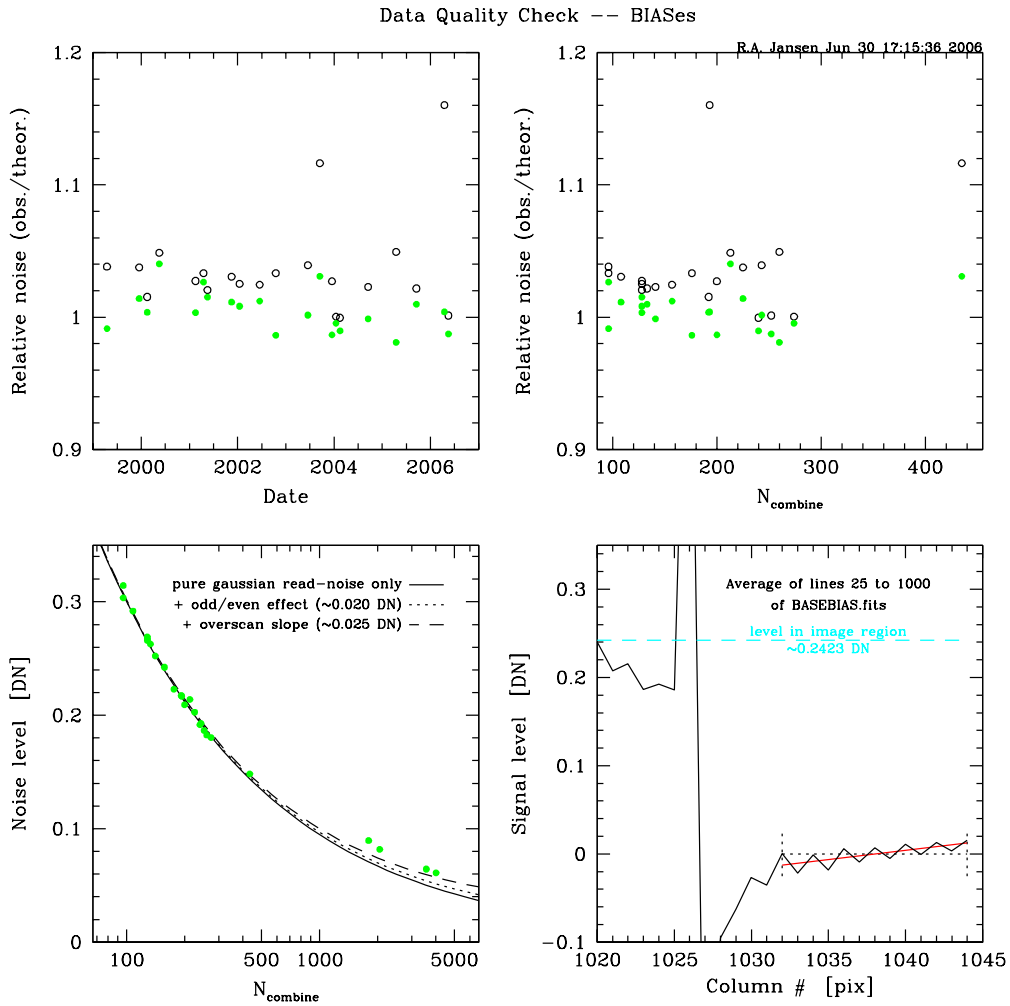


Figure 22: Comparison of theoretically expected and actually observed S/N for the simplest case: the combination of many bias frames obtained at the VATT, Mt. Graham, between 1999 and 2006. Note also in the lower right panel, that the level in the overscan region (which formally starts at column #1025) is not flat: only pixels in columns 1032–1044 are reliable for measuring the overscan level. Furthermore, there is a small but distinct offset of $\sim$0.24 ADU between the bias level measured in the overscan region (virtual pixels) and that measured in the image region (actual, illuminated pixels). Such offsets are common in ageing CCD devices and need to be taken into account during the data processing.

Signal and Noise: The CCD Equation (cont'd)

- **So, which noise terms would contribute to the uncertainty in the measured number of electrons in a single pixel?**

  First, when exposing a CCD, we accumulate signal due to arriving photons:

  - **If we denote the portion of the signal that originated from the astronomical source of interest by $S_\star$ (expressed in e^-), then the Poisson statistics give the 1 σ uncertainty, or noise, on that signal as $\sqrt{S_\star}$.**
  - **Usually, there is another signal present due to photons: the sky background. If we denote that portion of the signal by S_S, then the associated noise will be $\sqrt{S_S}$.**

  During an integration, we also collect signal that is *not* associated with arriving photons:

  - **The dark signal, S_D, due to thermal electrons from the bulk Si, builds up linearly with the integration time t (*whether the CCD is exposed to light or not*), i.e., $S_D = t \cdot d$. The *dark current* or *dark rate* d is usually expressed in units of e^-/pix/hour (where the "/pix" is a dummy unit, to remind us that, on average, every pixel will collect the same amount of dark signal). Here, if t is expressed in seconds, d should be in e^-/sec. The noise associated with the discrete (stochastic) dark signal also satisfies Poissonian statistics and will be $\sqrt{S_D} = \sqrt{t \cdot d}$.**

Signal and Noise: The CCD Equation (cont'd)

And finally we have a noise term that has no signal associated with it at all:

- **The *read-noise*, $\mathcal{R}$ (in e^-), adds directly to the total noise in the pixel under consideration.**

- **Upon read-out of the charge in the CCD pixel, we have no way to distinguish electrons originating from the various sources. We just measure the total signal. Hence, the associated noise is also grouped together as:**

$$N = \sqrt{S_\star + S_S + t \cdot d + \mathcal{R}^2}$$

Note, that the read-noise — since it should add linearly to each pixel — enters as the read-noise *squared* underneath the square root sign.

- **The S/N ratio for this single-pixel measurement can then be expressed as:**

$$\frac{S}{N} = \frac{S_\star}{\sqrt{S_\star + S_S + t \cdot d + \mathcal{R}^2}} \tag{1}$$

Signal and Noise: The CCD Equation (cont'd)

- **We can also express the signal-to-noise in terms of measured signals in ADU (but with the d and $\mathcal{R}$ still expressed in their usual units, since they are not measured from the pixel under consideration) by explicitly including the CCD *gain factor*, $\mathcal{G}$:**

$$\frac{S}{N} = \frac{S_\star^{ADU} \cdot \mathcal{G}}{\sqrt{S_\star^{ADU} \cdot \mathcal{G} + S_S^{ADU} \cdot \mathcal{G} + t \cdot d + \mathcal{R}^2}}$$

- **In most instances, the astronomical signal of interest is collected within $n_{\rm pix}$ pixels, rather than in one pixel. Eq. (1) then becomes:**

$$\frac{S}{N} = \frac{S_\star}{\sqrt{S_\star + n_{\rm pix} \cdot (S_S + t \cdot d + \mathcal{R}^2)}} \qquad (2)$$

Equation (2) is informally referred to as *"the CCD Equation"*.

Note, that if the number of pixels containing the signal is large, the contribution of the read-noise may grow large because it enters linearly, while the other noise terms enter only as the square root of their signals.

Signal and Noise: The CCD Equation (cont'd)

- **For Eq. (2) to hold, however, one would have to gain knowledge of the amount of signal in the sky background through some magical means.**

  If we explicitly include the uncertainty in determining the sky background level by taking the mean over $n_{\rm sky}$ background pixels, the equation is modified as follows:

$$\frac{S}{N} = \frac{S_\star}{\sqrt{S_\star + n_{\rm pix}\cdot\left(1+\frac{n_{\rm pix}}{n_{\rm sky}}\right)\cdot(S_S + t\cdot d + \mathcal{R}^2)}} \tag{3}$$

  Note, that $n_{\rm pix}/n_{\rm sky}$ is large when $n_{\rm sky}$ is small, while for $n_{\rm sky} \gg n_{\rm pix}$ term $(1 + n_{\rm pix}/n_{\rm sky})$ approaches 1. This properly reflects the notion that one, generally, needs to average of many background pixels to obtain an accurate estimate of the sky level.

Signal and Noise: The CCD Equation (cont'd)

▷ **If we, lastly, also include the effects of *digitization noise* as a result of the A/D conversion process, then we arrive at *the complete CCD Equation*:**

$$\frac{S}{N} = \frac{S_\star}{\sqrt{S_\star + n_{\rm pix}\cdot\left(1+\frac{n_{\rm pix}}{n_{\rm sky}}\right)\cdot(S_S + t\cdot d + \mathcal{R}^2 + \mathcal{G}^2\sigma_f^2)}} \tag{4}$$

Here, σ_f is the $1\,\sigma$ error introduced by the digitization process. Although somewhat dependent on the details of the design of the A/D electronics, a likely value of σ_f may be $\sim$0.289 (Merline & Howell 1995). If the gain is large, this term may become significant. For photon-counting operation ($\mathcal{G} = 1$), it is likely very much smaller than the read-noise term.

Signal and Noise: The CCD Equation (cont'd)

Problem: Compute the S/N obtained for an astronomical source in a CCD image with an exposure time $t = 300\,$sec. The CCD has a gain of $2\,e^-$/ADU, a read-noise of $6\,e^-$, and a dark rate of $10\,e^-$/pix/hr. The sky background level is measured to be $604\,$ADU, by averaging the pixel values in an annulus of blank sky around the object of interest with an inner radius of 20 pixels and outer radius of 30 pixels. The signal due to the source, measured in an aperture of radius 5 pixels, is found to be $S_{\star+S} - S_S = 6487\,$ADU.

Signal and Noise: The CCD Equation (cont'd)

Solution: Compute the S/N obtained for an astronomical source in a CCD image with an exposure time $t = 300\,$sec. The CCD has a gain of $2\,e^-$/ADU, a read-noise of $6\,e^-$, and a dark rate of $10\,e^-$/pix/hr. The sky background level is measured to be $604\,$ADU, by averaging the pixel values in an annulus of blank sky around the object of interest with an inner radius of 20 pixels and outer radius of 30 pixels. The signal due to the source, measured in an aperture of radius 5 pixels, is found to be $S_{\star+S} - S_S = 6487\,$ADU.

$$\frac{n_{\rm pix}}{n_{\rm sky}} = \frac{\pi \cdot 5^2}{\pi \cdot (30^2 - 20^2)} = 0.05 \quad \longrightarrow \quad n_{\rm pix} \cdot \left(1 + \frac{n_{\rm pix}}{n_{\rm sky}}\right) \simeq n_{\rm pix} = \pi \cdot 5^2$$

$$d = 10\,e^-/\text{hr} = \frac{10}{3600}\,e^-/\text{sec}$$

The gain is sufficiently close to the photon counting case, that we do not need to worry about the digitization noise $\mathcal{G}^2\sigma_f^2$. We then have:

$$\frac{S}{N} = \frac{6487 \cdot 2}{\sqrt{(6487 \cdot 2) + \pi \cdot 5^2 \left(604 \cdot 2 + 300 * \frac{10}{3600} + 6^2\right)}} \simeq 39$$

Note, that this is quite a bit smaller than $S_\star/\sqrt{S_\star} \simeq 114$. Had we omitted the additional noise terms in the CCD Equation, we would have grossly overestimated the S/N ratio.

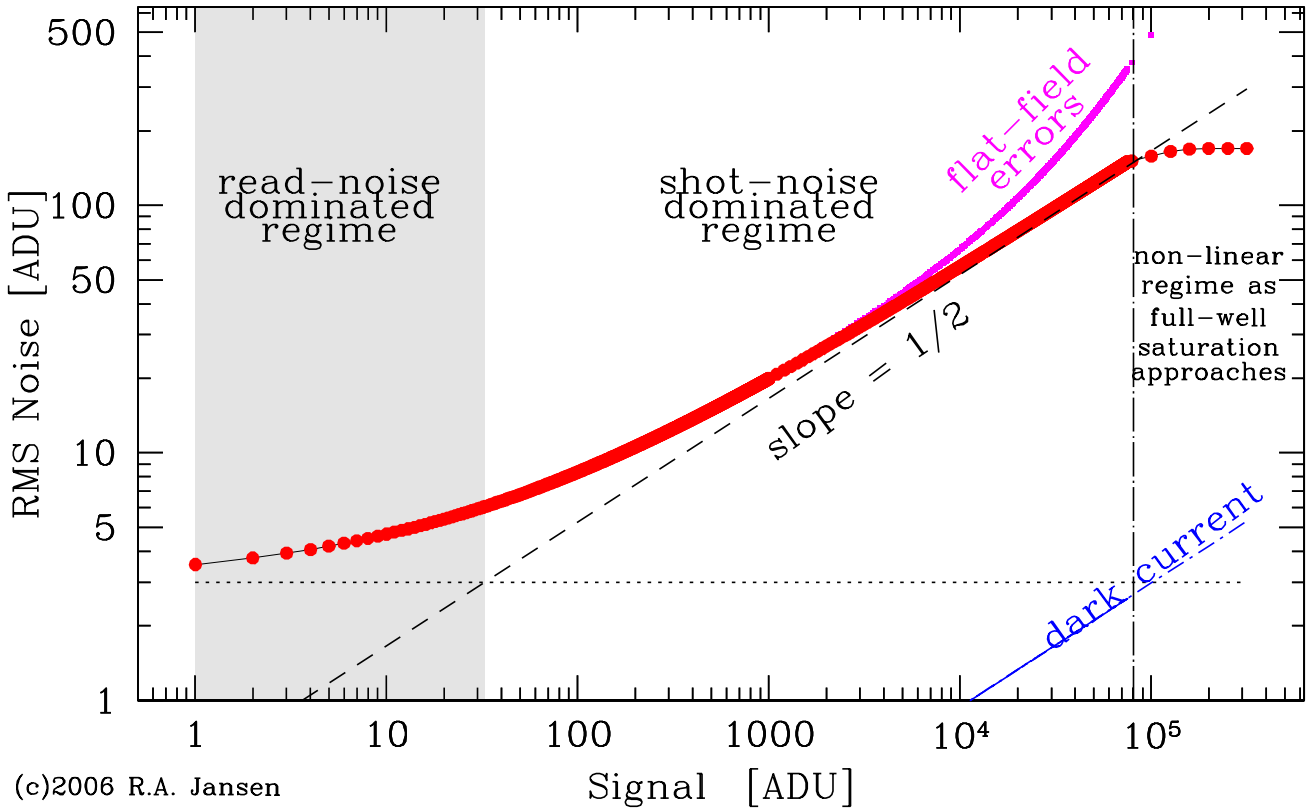


Figure 23: RMS noise as a function of signal level in a particular CCD. The read-noise sets a noise floor, to which shot noise increasingly adds. Once shot noise (photon noise) exceeds read-noise, the data are said to be shot noise limited. In the shot noise limited regime, the relation between noise and signal assymptotically approaches a slope of 1/2. If full-well capacity can be reached before A/D saturation occurs, then at very high signal levels the slope begins to deviate from 1/2. The additional noise due to dark signal is likely negligible compared to the shot noise for modern CCDs and observations through broad-band filters. It may not necessarilly be negligible compared to the read-noise, so for very faint sources in the near-UV or when observing through narrow-band filters (i.e., when the sky background is very low), the shot noise limited regime may not be reached. When the signal is shot-noise limited due to the signal from the astronomical source itself, the CCD Equation reduces to $S/N \simeq S_\star/\sqrt{S_\star} = \sqrt{S_\star}$. In this case, we speak of a "bright" object, in all other cases of a "faint" object. When the signal is shot-noise limited due to the sky, the noise scales as $\sqrt{S_\star + n_{\rm pix} \cdot S_S}$. The worst case is where the noise is dominated by the read-noise: $S/N \simeq S_\star/\sqrt{n_{\rm pix} \cdot \mathcal{R}^2} = S_\star/(\mathcal{R}\sqrt{n_{\rm pix}})$.

Magnitudes and Errors Thereon

▷ ***Magnitude*** **— a logarithmic scale for measuring flux ratios. The magnitude *difference* between to stars is related to the *ratio* of their fluxes by:**

$$m_1 - m_2 \equiv -2.5\,{}^{10}\log\frac{f_1}{f_2} = -2.5\,\log f_1 + 2.5\,\log f_2 \qquad (5)$$

where the factor -2.5 serves to approximate the ancient Greek magnitudes and is called the *Pogson scale*. Note, that a flux ratio of 100 corresponds to a difference of exactly 5 magnitudes.

- **While *magnitudes* are relative brightness units, the apparent magnitude of the Sun in the V filter is understood to be -26.76 ± 0.02 mag, as if it is a physical unit on some absolute scale.**

  So how did we get from what is inherently an ambiguous relative unit to a unit with an absolute physical meaning?

Magnitudes and Errors Thereon (cont'd)

If m_2 were the magnitude that corresponds to a ***known*** flux density f_2 in physical units of ergs s^{-1} cm^{-2} Å^{-1} (in the case of f_λ) or of ergs s^{-1} cm^{-2} Hz^{-1}(for f_ν) — in particular if $m_2 \equiv 0$ for that known flux density —, then the term $2.5 \log f_2$ becomes equivalent to an absolute ***zeropoint*** (zp) for the magnitude scale:

$$m_1 = -2.5 \log f_1 + zp \tag{6}$$

Vega magnitudes

The ultimate standard and reference for all classical broad-band photometry is the star α Lyrae (i.e., Vega). By substituting the flux density in a given filter of Vega for f_2 in Eq. (5) and dropping the subscripts "1", we obtain:

$$m - m_{\rm Vega} = -2.5 \log f_\lambda \; + \; 2.5 \log f_{\lambda,\rm Vega}$$

However, this would still leave us with $m_{\rm Vega}$. The final step, then, to yield a usable physical scale is to set:

$m_{\rm Vega} \equiv 0$ **in all filters** ***per definition*** $\Rightarrow m - m_{\rm Vega} \equiv m$

Vega magnitudes (cont'd)

Hence:

$$m \;=\; -2.5\,\log f_\lambda \;+\; 2.5\,\log f_{\lambda,\mathrm{Vega}} \qquad (7)$$

where the last term is the zeropoint, $zp(\lambda)$, on the Vega magnitude system. For a flux density outside the Earth's atmosphere of $(3.59 \pm 0.08) \times 10^{-9}$ ergs s^{-1} cm^{-2} Å^{-1} at 5480Å, this zeropoint $zp(5480\text{Å}) = -21.112$ mag.

▷ **As a corollary of the requirement that $m = 0$ at all wavelengths, the *color* of Vega in any pair of filters is 0, as well.**

- **Fig. 24(*a*) shows the spectrum of Vega in physical units as a function of wavelength.**
  - **Note, that the flux density of Vega varies greatly over the UV–near-IR regime, yet the magnitude corresponding to that flux density is equal to 0 at every wavelength.**

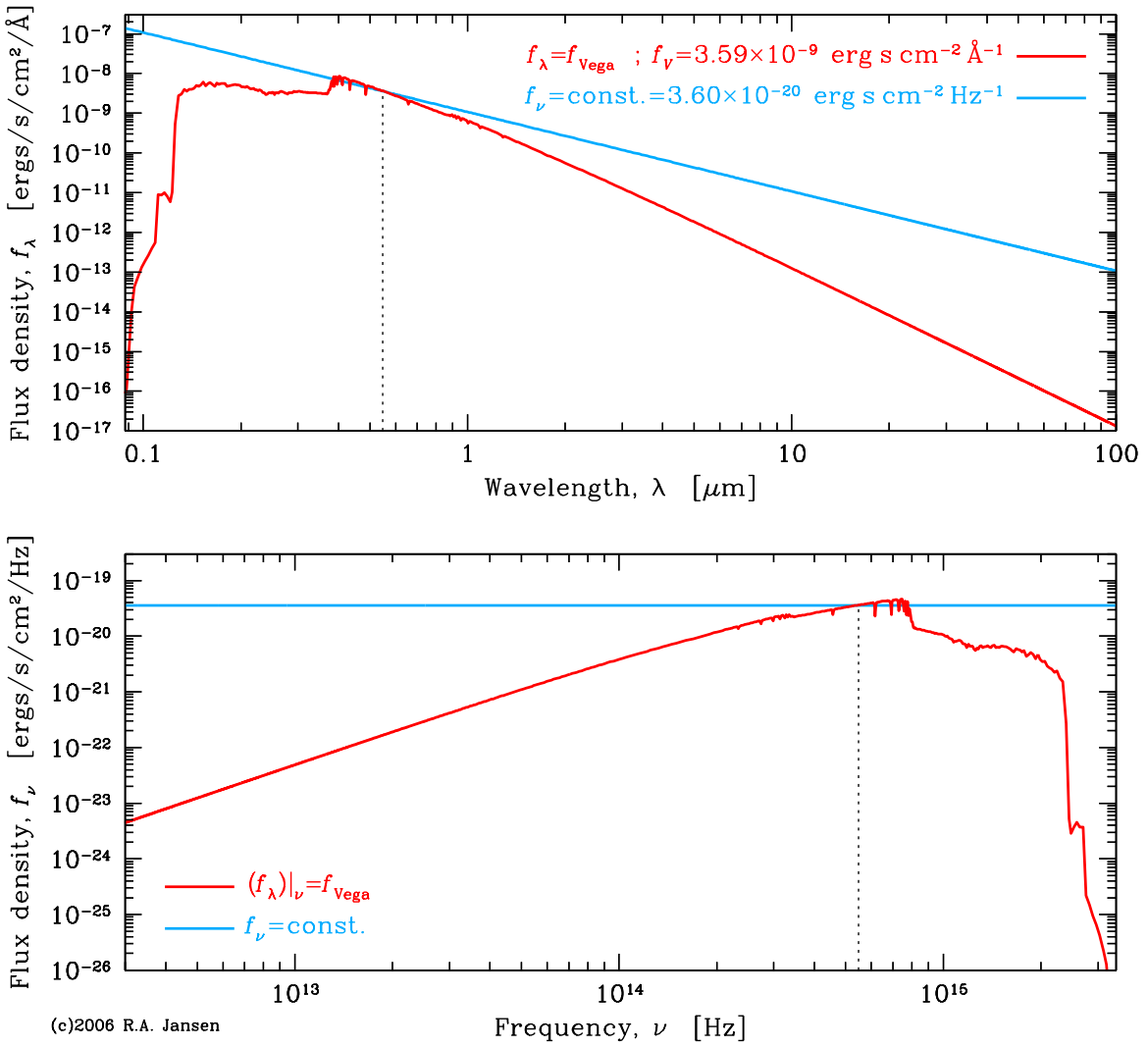


Figure 24: Comparison of the spectrum of α Lyrae (Vega) and a spectrum that is flat in f_ν. In the top panel, both are plotted as f_λ (in ergs s^{-1} cm^{-2} Å^{-1}) versus wavelength λ (in μm), while in the bottom panel they are presented as f_ν (in ergs s^{-1} cm^{-2} Hz^{-1}) versus frequency ν (in Hz). Both spectra are equal at the effective wavelength of the V filter, at 5480Å. The mid- to far-IR portion of the spectrum of Vega was replaced by the stellar atmospheric model of Kurucz (1979): an actual IR spectrum of Vega would show a significant additional, non-photospheric component due to its circum-stellar debris disk.

Vega magnitudes (cont'd)

- **Since it is impractical for every observer to observe Vega, a reference set of several dozens of, generally fainter, secondary standard stars was observed.**

  - **During WW II, Johnson made extensive photo-electric observations, spanning many years, through standard apertures in his UBV filter system. The UBV filter system was the first known standardized photometric system. The filter set was later extended toward the near- and mid-IR with $RIJK$ and L filters.**

  - **As progressively more precise magnitude and color measurements of an increasingly large number of stars, sampling a large range in colors, were placed onto the standard system of Johnson, slight inconsistencies in magnitude and colors became apparent. To remain as consistent as possible with earlier work, the V magnitude of Vega had to be adjusted slightly upward from exactly zero. The current best estimate is $m_V = +0.035 \pm 0.012$ mag for $f_{\lambda,V} = (3.593 \pm 0.084) \times 10^{-9}$ ergs s^{-1} cm^{-2} Å^{-1} (Colina & Bohlin 1994).**

    - **$m \equiv 0$ at all wavelengths for the *Vega magnitude system*, but the magnitude of Vega, *the star*, can differ slightly from 0.**

Vega magnitudes (cont'd)

- **More recently, Landolt (1992) published photo-electric $UBVR_cI_c$ photometry of a large number of equatorial (Dec $\sim$ 0°) fields in which stars of very different colors are relatively close together on the sky (within $\sim$5′–15′ on the sky). The transformations of his photometry onto the original system, again meant very slight changes to the original system. When using Landolt standard stars to photometrically calibrate you CCD images, you are actually calibrating onto Landolt's system, and not quite the original Vega system of Johnson.**

- **Even more recently, Stetson (2000) published a very large database of BVR_cI_c standard stellar magnitudes within the Landolt equatorial fields and in many other common fields across the sky based on (largely archival) CCD observations. His photometry is reduced onto the Landolt system, but his measurement method differed. Whereas Landolt used fixed circular apertures of 14″ with his photomultiplier tube, and hence often also measured the flux of adjacent fainters stars within that aperture, Stetson used the technique of PSF fitting, where each star is measured seperately.**

△ ***When calibrating CCD images, care needs to be taken to reproduce as close as possible the measurement method employed to obtain the original standard star photometry.***

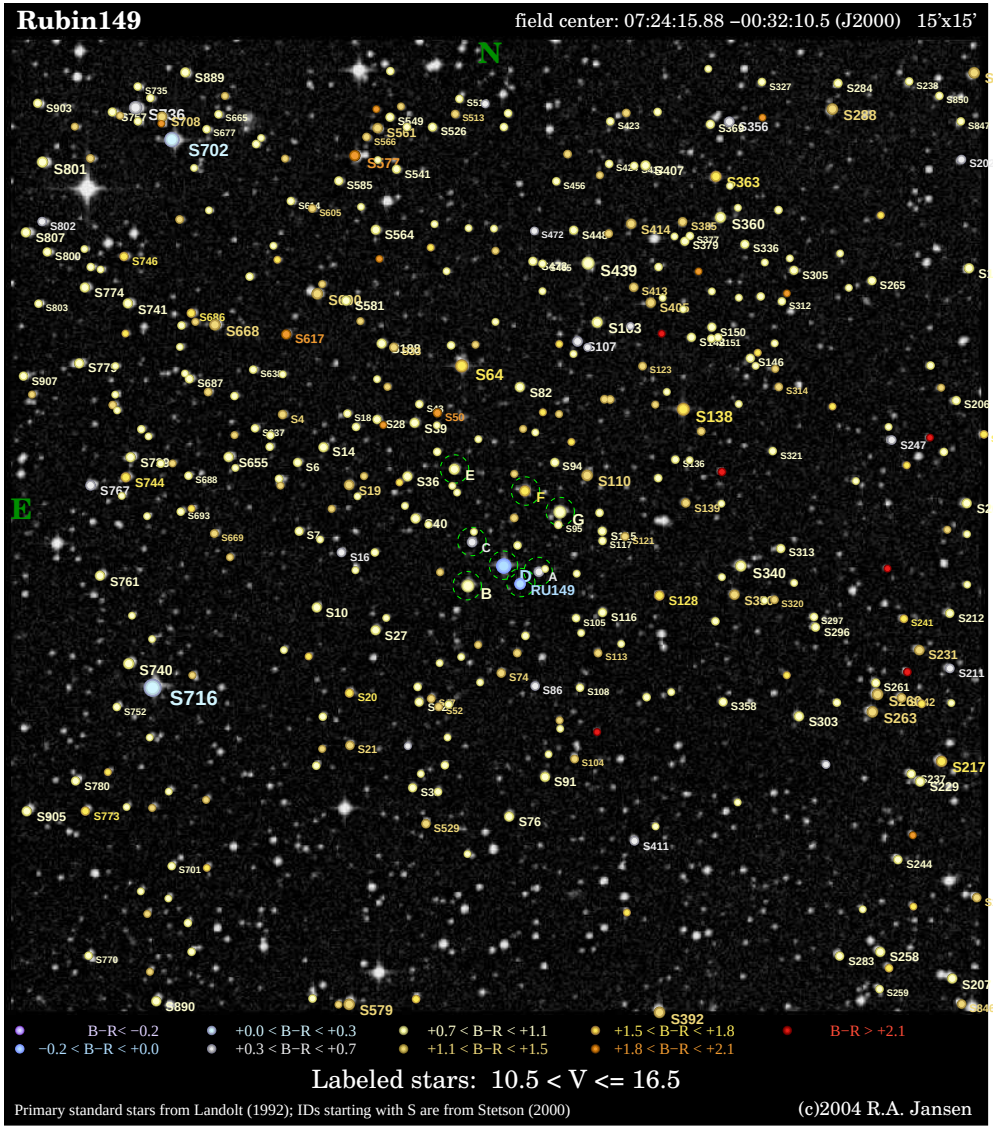


Figure 25: In standard field Rubin 149, Landolt (1992) published $UBVRI$ filter photo-electric photometry of the 8 stars labeled "Ru149" and A through G. Stetson (2000) published standard star photometry for all labeled stars in the larger general field of the Rubin 149 asterism.

AB magnitudes

- **To avoid the problems with Vega magnitudes — that the flux density that corresponds to $m = 0$ differs at every wavelength, and that the flux of Vega can become exceedingly small at wavelengths outside the UV–near-IR regime —, another magnitude system, the *AB* or *spectroscopic* or *natural magnitude system*, was deviced (Oke & Gunn 1983). In the AB magnitude system, the reference spectrum is *a flat spectrum in* f_ν:**

$$f_\nu \equiv \text{const.} \quad [\text{ergs s}^{-1}\,\text{cm}^{-2}\,\text{Hz}^{-1}].$$

That constant is per definition such that in the V filter: $m_V^{Vega} \equiv m_V^{AB} \equiv 0$ (or more accurately: $f_\nu\,\mathrm{d}\nu \equiv f_\lambda\,\mathrm{d}\lambda$ when averaged over the V filter, or at the effective wavelength of the V filter, $\lambda_{\rm eff} = 5480$Å.

△ ***Note that the AB magnitude system is expressed in f_ν rather than f_λ!***

AB magnitudes (cont'd)

- **The flux density in f_ν is related to the flux density in f_λ by:**

$$f_\nu \; [\mathrm{ergs\,s^{-1}\,cm^{-2}\,Hz^{-1}}] \;=\; \frac{\lambda^2}{c} \cdot 10^8 \cdot f_\lambda \;\; [\mathrm{ergs\,s^{-1}\,cm^{-2}\,\AA^{-1}}]$$

where we used $\lambda\,\nu \equiv c$ and transformed from $f_\lambda\,\mathrm{d}\lambda$ to $f_\nu\,\mathrm{d}\nu$. The factor 10^8 is included, because the natural units of wavelength are cm, not Å.

- **Converting the Vega magnitude zeropoint gives:**

$$m \;=\; -2.5\,\log f_\nu \;-\; (48.585 \pm 0.005) \qquad (8)$$

where the exact value of the zeropoint depends somewhat on the literature source (e.g., Hayes & Latham 1975; Bessell 1988,1990), based on a magnitude at 5556Å for Vega of 0.035–0.048 mag and a corresponding flux density of $(3.56\text{–}3.52)\times 10^{-20}$ ergs s^{-1} cm^{-2} Å^{-1}.

- **The AB magnitude system is also called the *spectroscopic magnitude system*, because with its constant zeropoint, it is useable at any wavelength in bandpasses of any width, and hence, also for narrow-band imaging and spectroscopy.**

- **In ground-based broad-band imaging, however, the Vega magnitude system is still a common system. Unless explicitly noted otherwise, one should assume Vega magnitudes.**

Computing errors on magnitudes

- **Using the identity $\sigma \equiv N/S$, we can express Eq. (4) (the complete CCD Equation) as the standard deviation on the measured magnitude:**

$$\sigma \equiv N/S \quad \longrightarrow \quad \sigma = \frac{\sqrt{S_\star + p}}{S_\star}$$

$$\sigma_{\text{mag}} \simeq \frac{1.0857\sqrt{S_\star + p}}{S_\star} = 1.0857\left(\frac{1}{S/N}\right) \qquad \text{[mag]} \tag{9}$$

where p represents the entire term under the square root sign in Eq. (4) other than $S_\star$. The scale factor $1.0857 = 2.5/\ln 10$ comes from the definition of magnitude (factor 2.5), and from computing the error on a decimal logarithm (factor $\ln 10$).

We used the fact that the probable error on a base-a logarithm $^a\log(x)$ is given by:

$$\frac{d}{dx}\{^a log(x)\} = \frac{d}{dx}\left\{\frac{ln(x)}{ln(a)}\right\} = \frac{1}{\ln(a)} \cdot \left(\frac{dx}{x}\right)$$

Computing errors on magnitudes (cont'd)

In our earlier example, $S/N = 39$ and so: $\sigma_{\text{mag}} \simeq 1.0857 \left(\frac{1}{39}\right) = 0.028$ **mag**

Above and in Eq. (9), we used an "approximately equal to" ($\simeq$) notation rather than an equals sign ($=$). Let's consider this for a bit.

Per definition: $m = -2.5 \cdot \log(S) + zp$

Therefore, errors of $+1\,\sigma$ and $-1\,\sigma$ in source signal S mean:

$$\begin{aligned}
\sigma_m^+ &= m^+ - m &&= -2.5 \cdot \log(S+N) + 2.5 \cdot \log(S) \\
&&&= -2.5 \cdot \log\left(\frac{S+N}{S}\right) = -2.5 \cdot \log\left(1 + \frac{1}{S/N}\right) \\
\sigma_m^- &= m^- - m &&= -2.5 \cdot \log(S-N) + 2.5 \cdot \log(S) \\
&&&= -2.5 \cdot \log\left(\frac{S-N}{S}\right) = -2.5 \cdot \log\left(1 - \frac{1}{S/N}\right)
\end{aligned}$$

So we find that $|\sigma_m^+| \neq |\sigma_m^-|$ and that "upper" errors (toward *brighter* fluxes or *smaller* magnitudes) are therefore smaller than the "lower" errors (toward *fainter* fluxes or *larger* magnitudes)! This behavior is clearly illustrated in Fig. 26(b).

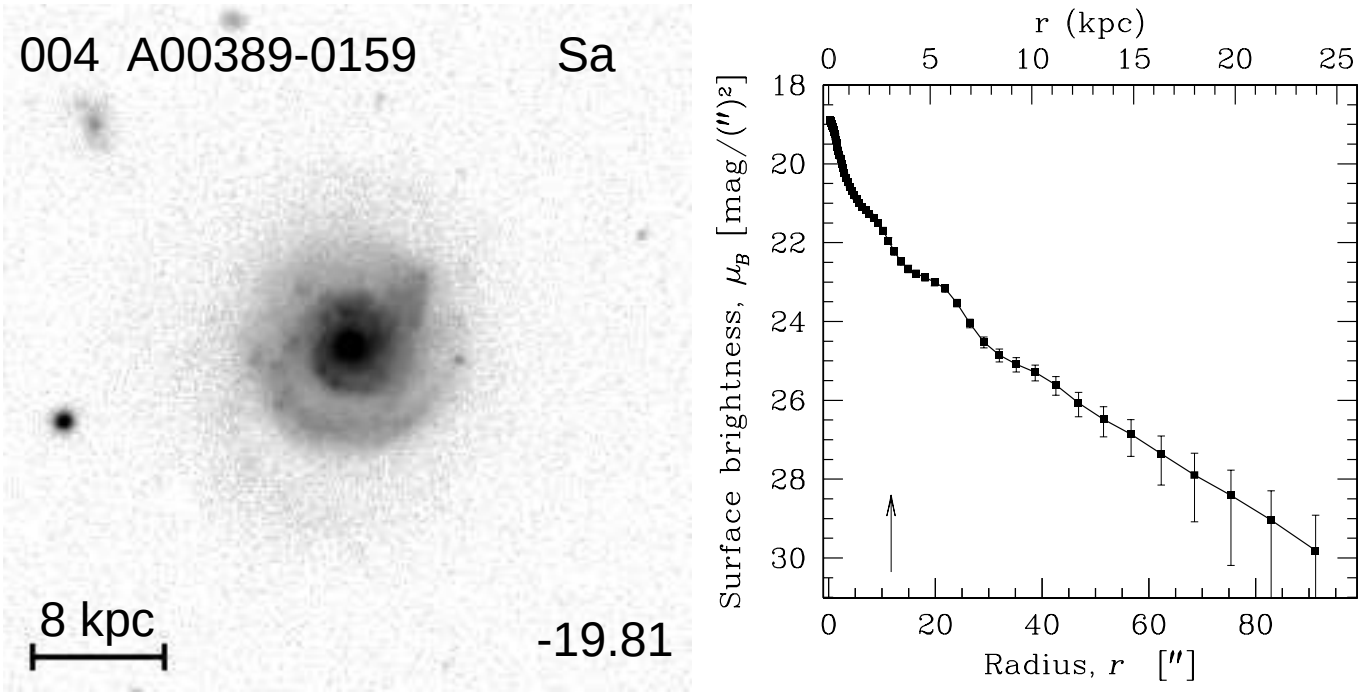


Figure 26: (*a*) *B* filter CCD image obtained at the SAO 48" telescope (Mt. Hopkins, AZ) of Sa galaxy A00389-0159 (UGC 439), observed as part of the Nearby Field Galaxy Survey (Jansen et al. 2000a,b). (*b*) Radial surface brightness profile measured from this image. Note that at low S/N (larger radii), the upper and lower magnitude errors on the profile become increasingly asymmetric.

- **For relatively large magnitude errors (low S/N), the upper and lower errors become progressively different in size: *magnitude errors are* not *symmetric*.**

Computing errors on magnitudes (cont'd)

A final important observation is that the signals $S_\star$ and S_S scale linearly with the exposure time t. The S/N ratio in the shot-noise dominated regime (see Fig. 23) will therefore scale as $t/\sqrt{t} = \sqrt{t}$.

- **This implies that, to obtain a S/N that is *twice* as good, one has to integrate *4 times* as long. In the read-noise dominated regime, the S/N will increase much slower — if $n_{\rm pix} \cdot \mathcal{R}^2 \gg (S_\star + n_{\rm pix} \cdot S_S)$, the S/N remains nearly constant and < 1, no matter how long you integrate!**

Filter systems galore

- **Over time, a large number of distinct *filter systems* have been developed, each with its strengths for particular applications.**

  - **Johnson-Morgan/"Johnson" — U, B, V, R, I broad-band filters; extended to the near-IR with J, K, and L (and M). The main disadvantage of the U filter was that its blue cut-off was mainly determined by the transparency of the atmosphere and telescope optical system (rather than by the glass of the filter).**
  - **Kron-Cousins-Glass/"Cousins" — R_c, I_c broad-band filters; better behaved in their red-tail, better positioning in wavelength with respect to UBV.**
  - **Bessell (1979,1990), Bessell & Brett (1988)/"Bessell" — better characterization and formalization of what the $UBVR_cRI_cIJHKLM$ broad-band bandpass curves (resulting from filter plus detector) *should* look like.**
  - **Strömgren (1966) filters — u, v, b and y; medium-band filters, specifically designed for stellar astrophysics (hot vs. cool stars; stellar composition): u and v straddle the Balmer break (actually the Ca II H+K break at $\sim$4000Å), and $u-v$ and $v-y$ colors provide the strength of that break and the continuum slope redward of the Balmer break.**

Filter systems galore (cont'd)

- **Washington filters (defined by G. Wallerstein, developed by Canterna (1976), calibrated on the Geisler (1996) system of CCD standards; see also Bessell (2001))/"Washington system" — C, M, T_1 and T_2, specifically designed for metallicity studies in old stellar populations.**
- **Straižys et al. (1966); Straižys & Zdanavičius (1970)/"Vilnius system" — U, P, X, Y, Z, V, and S. Mainly used for stellar classification.**
- **Gunn, Thuan & Gunn / "Gunn" — u, g, r (later also i, z), have filters with transmission curves with steeper cut-on and cutt-off; the precursor of the Sloan and various *HST* filters.**
- **2MASS filters; Jarret et al./"2MASS system" — J, H, K_s; by reducing the effective wavelength of the K filter from 2.2 to 2.15 μm and designing a steeper red cut-off, the sky background in K_s is significantly darker than in the Johnson K filter.**
- **Sloan Digital Sky Survey/"Sloan" — u', g', r', i', and z'. Square filter transmission curves with minimal overlap and minimal gaps between the filters. Optimal broadband filter set for photometric redshifts and quasar searches.**

Filter systems galore (cont'd)

- **Beijing-Arizona-Taiwan-Connecticut filter set; spectro-photometrically calibrated by Yan (2000)/"BATC" — System of 16 medium-band filters in near-UV through near-IR that avoid night-sky emission lines. Optimal medium-band filter set for photometric redshifts and (high-redshift) emission-line object searches.**
- **filters for various space-based observatories (notably *HST*'s four generations of camera's: WF/PC, WFPC2, ACS/WFC+HRC, WFC3); optimizations of and compromises between existing ground-based filters; extensions toward the UV and near-IR.**

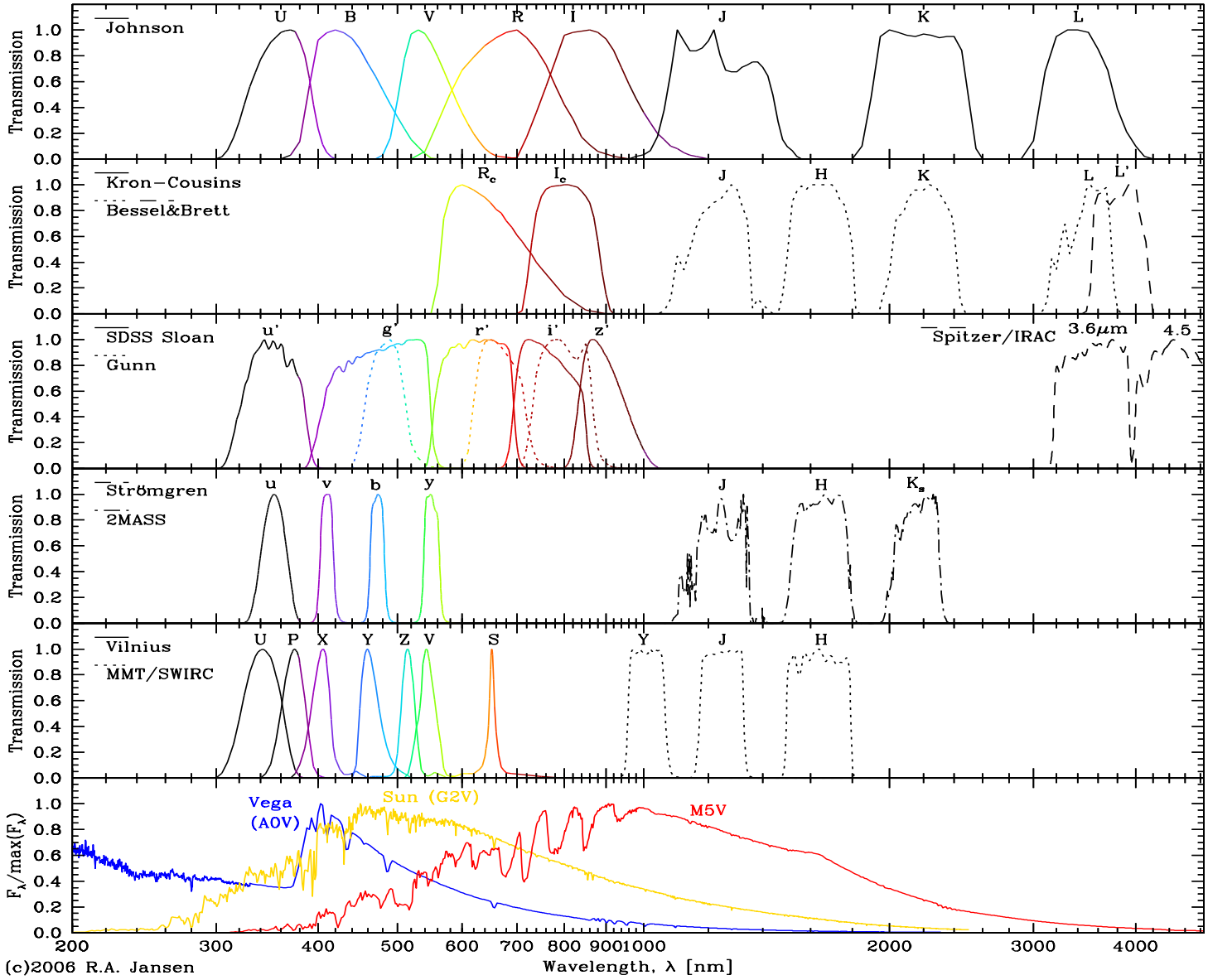


Figure 27: Overview of various filter sets (as labeled) and comparison with the spectra of an A0V (Vega), G2V (Sun) and M5V star.

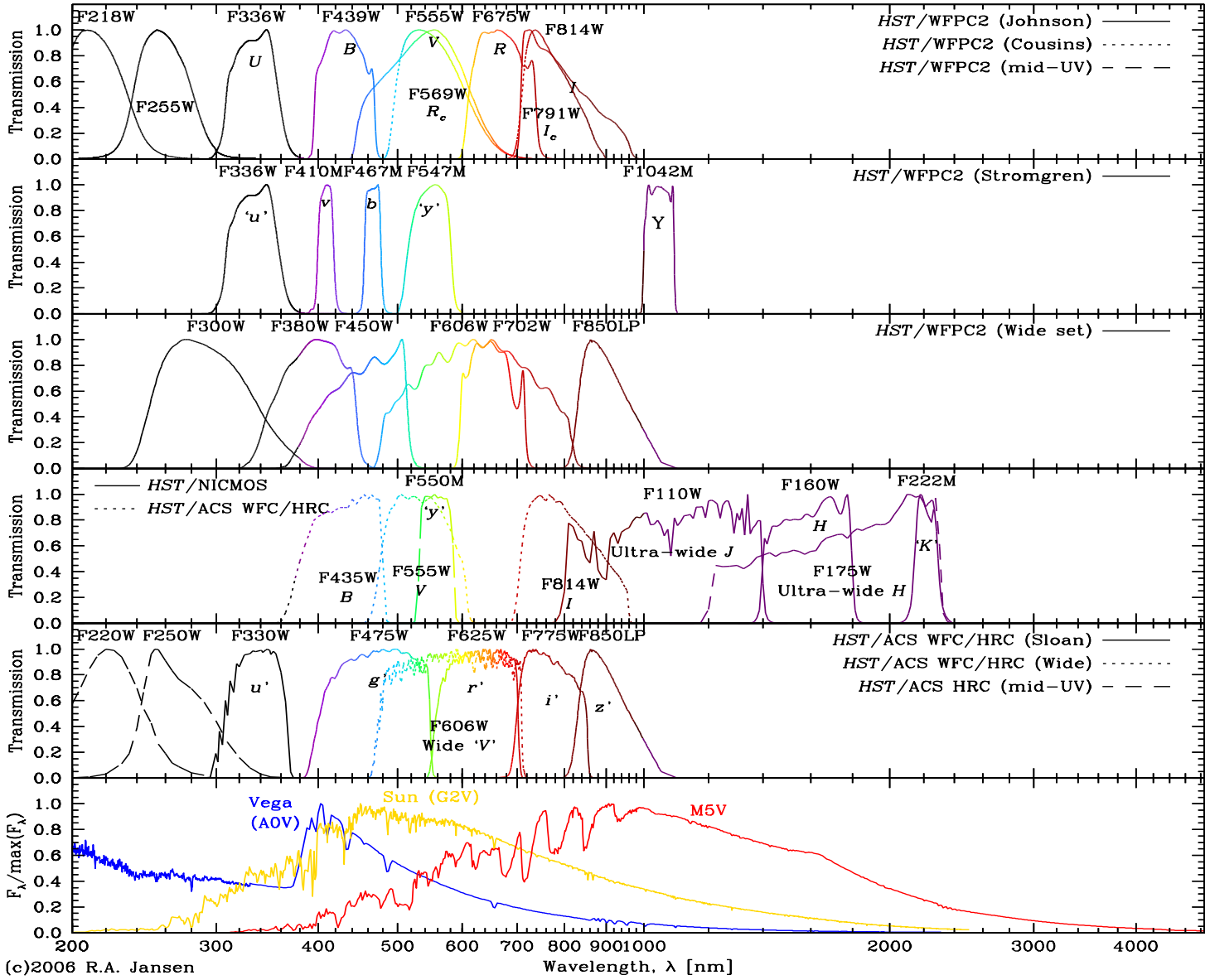


Figure 27: [*Continued*] Overview of various filter sets (as labeled).

Looking through the Earth's Atmosphere

Atmospheric Transmission, Emission, and Scattering

- **All observations using ground-based instruments suffer the effects of having to look through the Earth's atmosphere:**

  - **Turbulence and temperature variations in the atmosphere can blur any incident astronomical signal**

  - **The transparency of the atmosphere is a function of wavelength — in some wavelength regimes a strong one.**

    - **Below $\sim$310 nm, the atmosphere is essentially opaque, due to the combined effects of ozone (O_3) and Rayleigh scattering ($\propto \lambda^{-4}$).**

    - **Above $\sim$800 nm there are only discrete windows where observations are possible (see Fig. 28). Atoms and molecules in the atmosphere are also emitting light at specific wavelengths (see Fig. 29).**

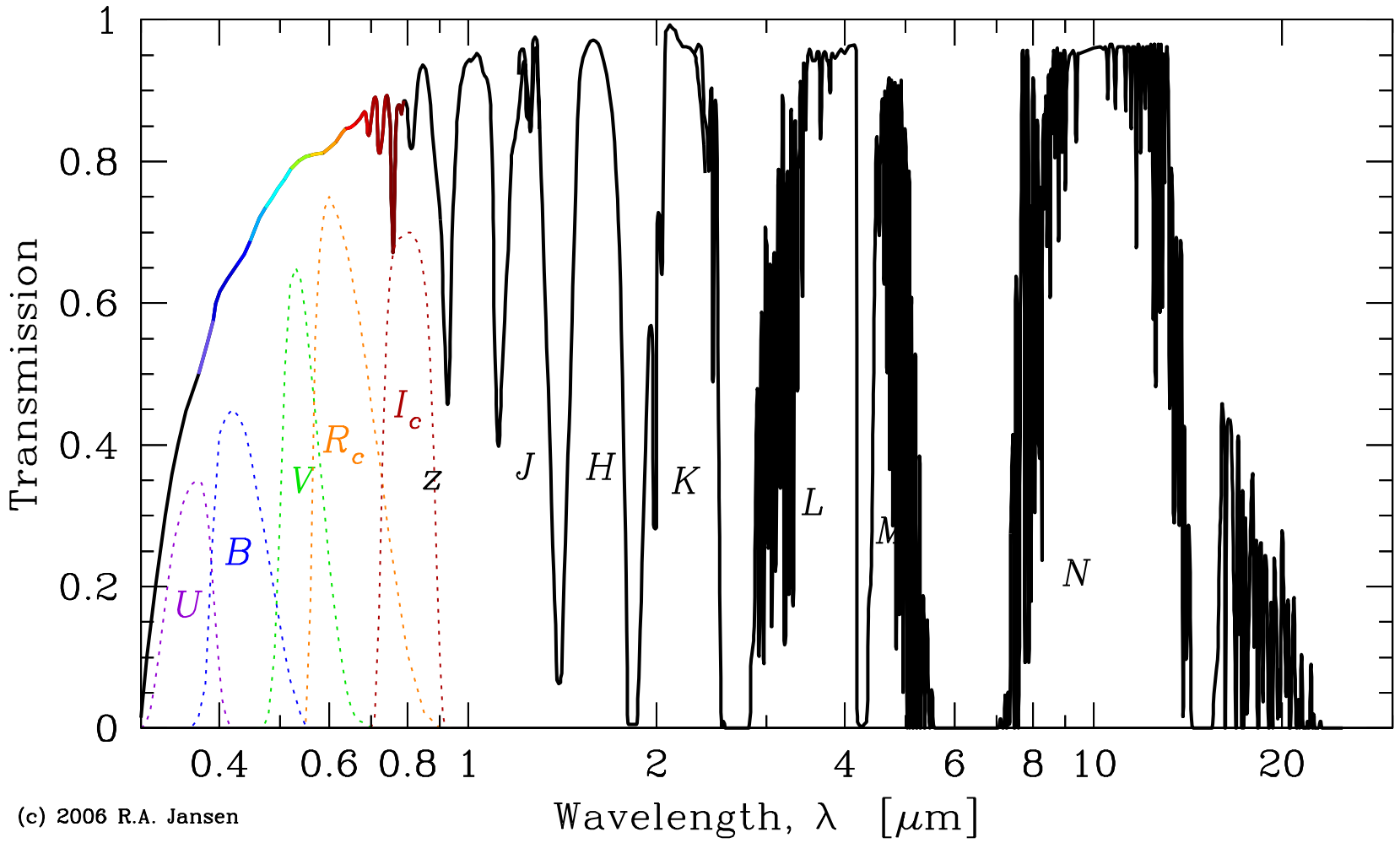


Figure 28: Atmospheric transmission from the atmospheric cut-off in the near-UV (∼310 nm) to the cut-off redward of the ***N*** window in the mid-IR (*After: RCA Electro-Optics Handbook*). The shape of several classical filter passbands (***U B V R$_c$ I$_c$***) is indicated, and the atmospheric windows in the near- and mid-IR are labeled. Most of the atmospheric absorption is due to molecules: O_3 in the near-UV and visible, and O_2, H_2O, CO_2, and N_2O in the infrared. The transmission of the atmosphere is strongly dependent on the amount of precipitable water vapor in the atmosphere above the telescope. The effects of atmospheric *emission*, which limits the usable wavelength intervals even further, is not included in this plot (but see Fig. 29).

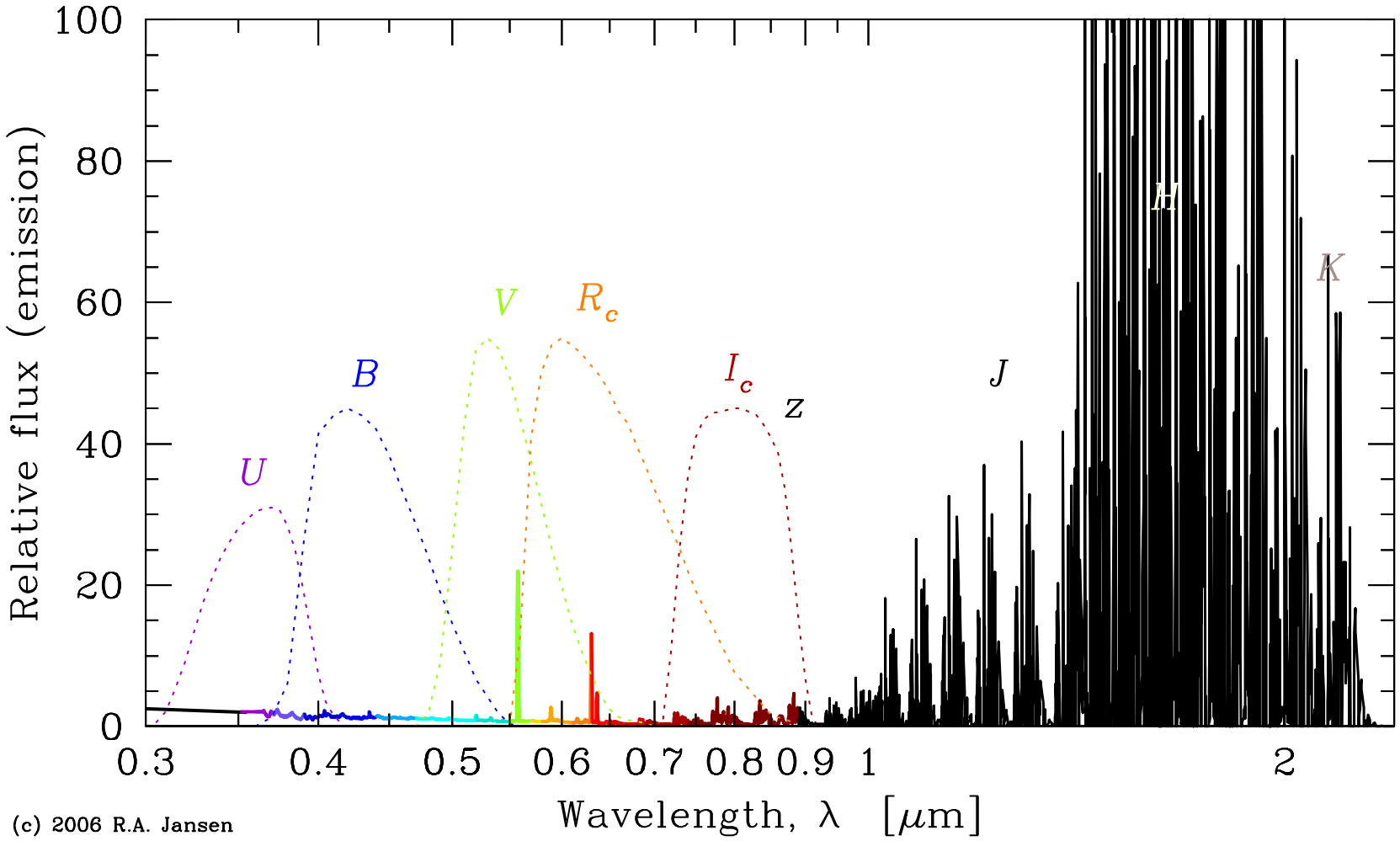


Figure 29: Atmospheric emission from the atmospheric cut-off in the near-UV through the near-IR ***K*** band (Source: *0.3–0.9 μm: based on VLT/UT3+FORS1 spectra obtained by R. Jansen & P. Jakobsen; 0.9–2.5 μm: theoretical OH night-sky lines from Rousselot et al. 2000*). Again, shapes of several classical filter passbands or their effective wavelengths are indicated. Although the optical night-sky spectrum — but for the very strong O I lines at 557.7, 630.0 and 636.3 nm — is relatively clean, the sky spectrum beyond $\sim$750 nm is dominated by strong series of OH lines. The strengths of night-sky emission lines tend to vary throughout a night and from night to night. Beyond $\sim$4 μm, the sky becomes exceedingly bright.

Atmospheric Transmission, Emission, and Scattering (cont'd)

- **And lastly we have to contend with light from our Moon that is scattered by the atmosphere.**

Table 1 tabulates the average surface brightness of the sky background and its dependence on lunar phase for five filters ($UBVRI$) and lunar ages of 0 (New Moon) through 14 (Full Moon). This dependence is due to scattering of moonlight in the Earth's atmosphere.

Table 1: Sky brightness for different lunar ages at CTIO

Lunar age (days)	Sky surface brightness (mag/arcsec^2)				
	U	B	V	R	I
0	22.0	22.7	21.8	20.9	19.9
3	21.5	22.4	21.7	20.8	19.9
7	19.9	21.6	21.4	20.6	19.7
10	18.5	20.7	20.7	20.3	19.5
14	17.0	19.5	20.0	19.9	19.2

Source: Alistair Walker, NOAO Newsletter #10

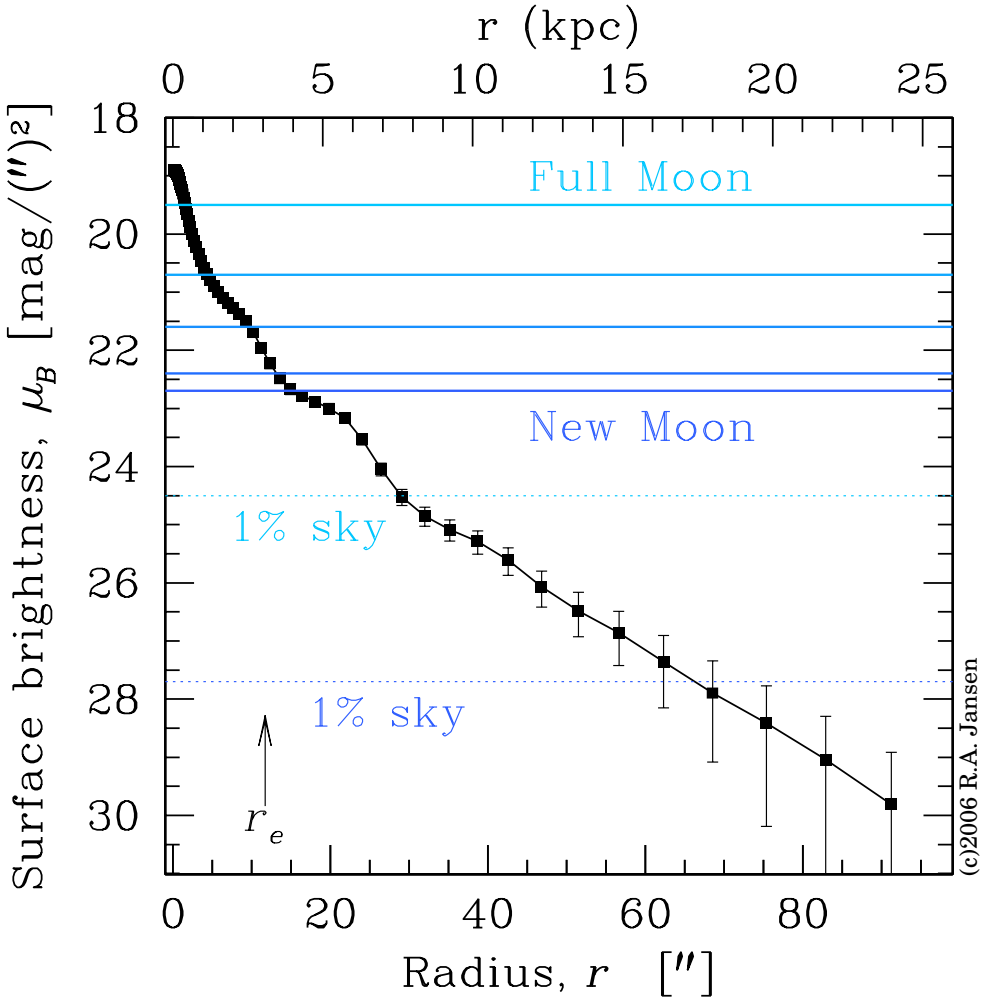


Figure 30: Comparison of the B-filter surface brightness profile of a nearby galaxy (A00389−0159; Jansen et al. 2000a) with the surface brightness of the sky, for the moon phases of Table 1. Note that, even at New Moon, most of the galaxy is fainter than the sky. With a central surface brightness, $\mu_0^B \simeq 19\,\mathrm{mag\,arcsec^{-2}}$ and a profile that declines exponentially with radius in the outer disk, this galaxy is quite typical of nearby spiral galaxies. Generally, with sufficiently long exposures, one can confidently detect signals to ∼1% of the sky background. The difference between observing during *dark* and *bright* sky conditions, corresponds to a factor of more than 2 in radius within most spiral galaxies.

Atmospheric Transmission, Emission, and Scattering (cont'd)

- **The actual brightness also depends on how close to the moon you look and, for partial phases, whether the Moon is above the horizon. (Remember, that for partial phases, the Moon is up only part of the night, while at Full Moon it is up *all* night). The main trends are, that:**
  - **The Moon's effect is strongest in the bluest bandpasses: in U, the sky is 5 mag (i.e., a factor 100!) brighter at full moon than at new Moon or when the Moon is well below the horizon. In B, the difference is $\sim$3.2 mag. In the redder I filter, the difference is reduced to a factor $\sim$2. This dependence on wavelength is partly because the moonless sky is fainter in the bluer bandpasses than it is in redder passbands — in the mid-IR the sky is so bright that there is little difference between day and night!**
  - **Taylor, Jansen & Windhorst (2004) show that the darkest skies at the summit of Mt. Graham, AZ, where the 2×8.4 m Large Binocular Telescope (LBT) is located, have a sky surface brightness that is comparable to the darkest nights at Cerro Tollolo Inter-american Observatory (CTIO) in Chili.**

Atmospheric Transmission, Emission, and Scattering (cont'd)

- **Loss of intensity during the passage of light through the atmosphere results from *extinction* — the combined effects of *absorption* and *scattering*. For plane parallel layers and for monochromatic light we have at zenith (see Fig. 31(a)):**

$$\mathrm{d}I = -I \cdot \kappa \, \mathrm{d}h \qquad \text{or:} \qquad \frac{\mathrm{d}I}{I} = -\kappa \, \mathrm{d}h$$

$$\Longrightarrow \quad \ln I = -\int_0^\infty \kappa(h) \, \mathrm{d}h \; + \; \text{const.}$$

where κ is the extinction coefficient per cm — which will depend on the density of the atmosphere and, hence, on h —, and const. $= \ln(I_0)$. If we denote the extinction coefficient for one *air mass* (i.e., a full thickness of the Earth's atmosphere) by K, then the *transmission* through the atmosphere will be given by $I/I_0 = e^{-K}$.

For a *zenith distance* (angle from zenith) z this becomes:

$$I/I_0 = e^{-K \, \sec z} \quad \text{, with:} \quad \sec z \equiv \frac{1}{\cos z} \tag{10}$$

And expressed in magnitudes:

$$\begin{aligned} m - m_0 \; &\equiv \; -2.5 \cdot \log(I/I_0) \; = \; -2.5 \cdot \log(e^{-K \, \sec z}) \\ &= \; -2.5 \cdot (-K \, \sec z) \cdot \log e \; \simeq \; 1.0857 \cdot K \, \sec z \end{aligned} \tag{11}$$

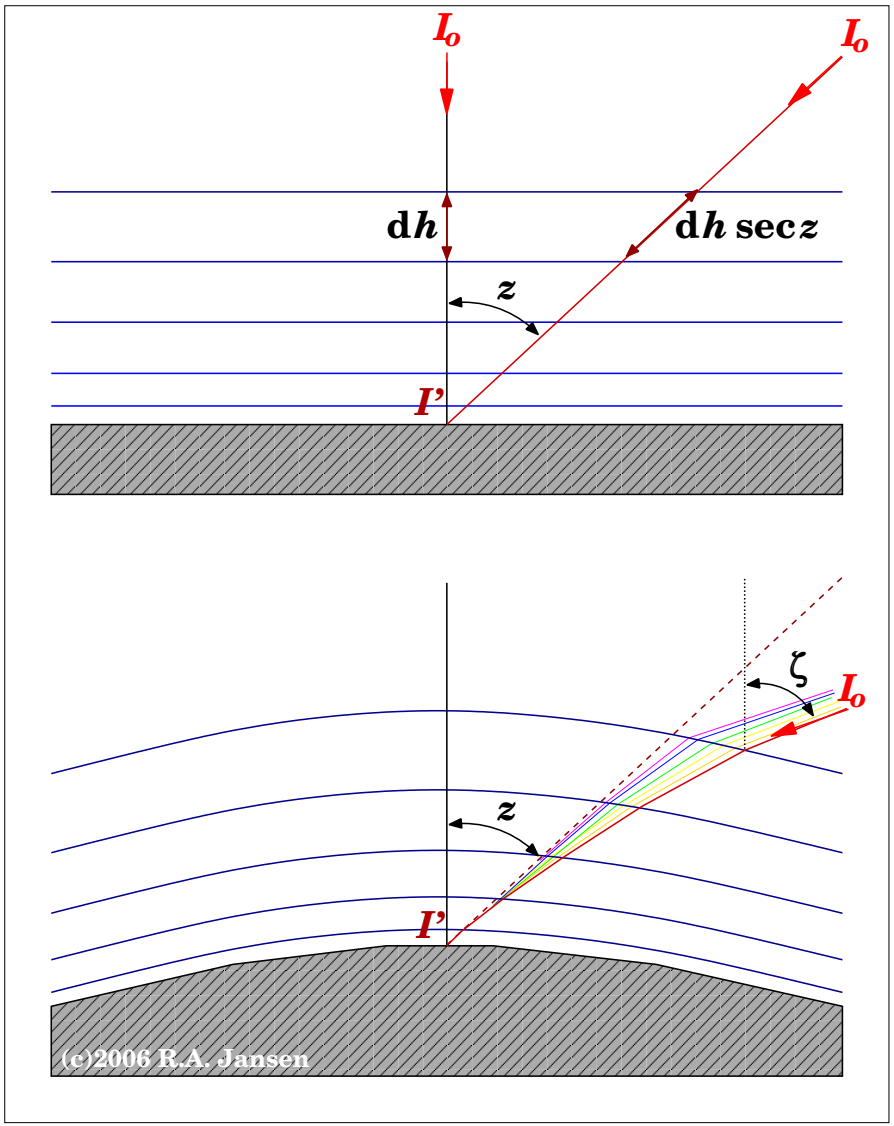


Figure 31: Schematic representation of the geometry of the Earth's atmosphere (*a*) in the approximation of plane parallel layers; (*b*) when the curvature of the atmosphere and refraction are taken into account.

Atmospheric Refraction and Dispersion

- **In reality, however, the Earth's atmosphere is curved, not plane parallel, and the path followed by a beam of light will be curved due to *atmospheric refraction*. This is illustrated in the bottom panel of Fig. 31. Generally, it may be necessary to use the actual air mass, $M(z)$, instead of $\sec z$.**

  To evaluate the importance of the difference between $\sec z$ and $M(z)$ and of atmospheric refraction, in Table 2 we compare the relative intensity losses I/I_0 for each at two different wavelengths. The correct form of Eq. (10) is:

$$I/I_0 = e^{-K\cdot M(z)} \quad \textbf{with:} \quad M(0) \equiv \sec 0 \equiv 1 \tag{12}$$

  It is important to note that the values for $M(z)$ listed in Table 2 are valid for standard temperature and pressure *at sealevel*; $M(z)$ varies with altitude and temperature. Nonetheless, we find that the intensity losses do not differ significantly for zenith distances $\lesssim 80°$. The extinction coefficient $K(\lambda)$, however, is — as we already found from Fig. 28 — a strong function of wavelength, as evidenced here by the large difference in losses at 5500Å and 3200Å.

Atmospheric Refraction and Dispersion (cont'd)

- **The zenith distance z of an astronomical object as observed at the ground is always *smaller* than the true zenith distance ζ that one would measure outside of the Earth's atmosphere. For a precise calculation of the refraction one would need to know the geometry of the atmosphere and the change in refractive index as a function of density (height) along the curved(!) light path. And the refractive index itself is also a function of wavelength. The relation between refractive index and wavelength is called *atmospheric dispersion*.**

z	$\zeta - z$	$\sec z$	$M(z)$	$(I/I_0)_{\sec z}$ 5500Å	$(I/I_0)_{\sec z}$ 3200Å	$(I/I_0)_{M(z)}$ 5500Å	$(I/I_0)_{M(z)}$ 3200Å
0°	$0''$	1.000	1.000	0.83	0.40	0.83	0.40
10°	$10''$	1.015	1.015	0.83	0.39	0.83	0.39
20°	$21''$	1.064	1.064	0.82	0.38	0.82	0.38
30°	$34''$	1.155	1.154	0.81	0.35	0.81	0.35
40°	$49''$	1.305	1.304	0.78	0.30	0.78	0.30
50°	$1'10''$	1.556	1.553	0.75	0.24	0.75	0.24
60°	$1'41''$	2.000	1.995	0.69	0.16	0.69	0.16
70°	$2'39''$	2.924	2.904	0.58	0.069	0.58	0.070
80°	$5'19''$	5.76	5.60	0.34	0.005	0.35	0.006
85°	$9'52''$	11.47	10.40	0.12	0.000	0.14	0.000
90°	$35'22''$	∞	38.	0	0	0.0008	0.000

Table 2: Transmission of the Earth's atmosphere for different zenith distances z and wavelengths λ. The 2nd column lists the angular deviation, $(\zeta - z)$, for $\lambda = 5500$Å (as defined in Fig. 31).

Atmospheric Refraction and Dispersion (cont'd)

- **Due to *atmospheric dispersion*, any object that isn't exactly at zenith will appear to the observer as a small spectra. Since the refraction will be stronger for bluer wavelengths than for redder wavelengths, the blue side of that spectrum will point to zenith and the red side away from it (i.e., the blue side points *up*).**

A good approximation for the difference between true and apparent zenith distance to $\zeta \simeq 75^\circ$ (for $\lambda = 5500$Å, 750 mbar, 0 °C) is given by the expression:

$$(\zeta - z) = 60\farcs4 \tan\zeta - 0\farcs064 \tan^3\zeta \tag{13}$$

In view of Snell's law, the wavelength dependence of this difference between true and apparent zenith distance with respect to the dispersion at 5500Å is given by:

$$(\zeta - z)_\lambda \;=\; (\zeta - z)_{5500\text{Å}} \cdot \frac{(n_\lambda - n_{\text{vac}})}{(n_{5500\text{Å}} - n_{\text{vac}})} \quad \textbf{, where} \quad n_{\text{vac}} \equiv 1 \tag{14}$$

Atmospheric Refraction and Dispersion (cont'd)

λ (Å)	3000	4000	5000	6000	8000
$\left(\frac{n_\lambda - 1}{n_{5500} - 1}\right)$	1.047	1.014	1.001	0.995	0.989

Table 3: Atmospheric dispersion with respect to the dispersion at $\lambda = 5500$Å

Table 3 tabulates the atmospheric dispersion coefficient with respect to that at 5500Å for five wavelengths from the atmospheric cut-off in the near-UV to the near-IR (800 nm).

From this table we can evaluate how severe the distortion of the image of that astronomical object into a little spectrum will be, for a moderate zenith distance of, say, 45°:

$$\left.\begin{aligned} (\zeta - z)_{3000} &= 1.047 \cdot (60\farcs4 \tan 45^\circ - 0\farcs064 \tan^3 45^\circ) \\ (\zeta - z)_{8000} &= 0.989 \cdot (60\farcs4 \tan 45^\circ - 0\farcs064 \tan^3 45^\circ) \\ \tan 45^\circ &\equiv 1 \end{aligned}\right\} \Longrightarrow$$

$$(\zeta - z)_{3000-8000} = (1.047 - 0.989) \cdot (60\farcs4 - 0\farcs064) \simeq 3\farcs50$$

And at an air mass of 2 (zenith angle of 60°), the length of the same spectrum would have grown to $\sim 6\farcs0$.

Atmospheric Refraction and Dispersion (cont'd)

- **When trying to obtain a spectrum of an object through a typical spectrograph slit of $\sim 1''$ width, one might completely lose the blue and red parts of the spectrum when the green light is centered in the slit (Filippenko 1982).**
  - **To avoid this, the spectrograph slit should be aligned with the *parallactic angle*: the angle at which the spectrograph slit is normal to the horizon.**

Although in imaging applications one never covers such a large range in wavelength, a stellar image observed in, say, the B filter would still be elongated by $\sim 0\farcs5$ at z=45° and $\sim 1\farcs0$ at z=60°. This is sufficiently noticeable that large telescopes at good sites have implemented *atmospheric dispersion correctors*.

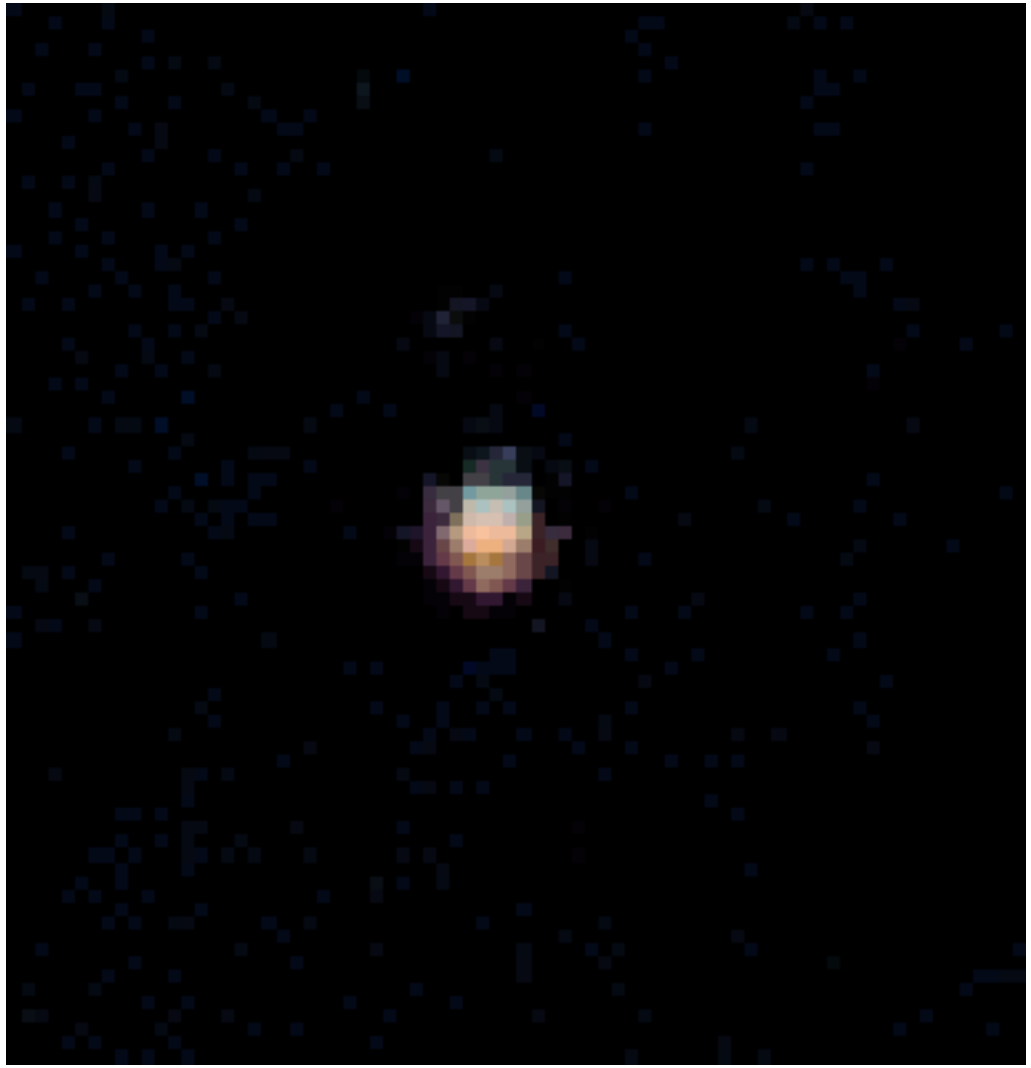


Figure 32: Example of the effects of atmospheric dispersion. The stellar image above was observed in a quick succession of $\boldsymbol{B\,V\,R}$ exposures at the 6.5 m Magellan '*Baade*' telescope with the IMACS instrument in Dec 2003 (R. Jansen & R. Windhorst) — before the atmospheric dispersion corrector was commissioned. The air mass was 1.33 (i.e., $z \simeq 41°$) and the seeing $0\farcs75$ (3.38 pixels) FWHM. Atmospheric dispersion caused a noticeable differential shift of the centroids of the $\boldsymbol{B}$ and $\boldsymbol{R}$ images on the CCD of $\sim 1\farcs0$.

Seeing and Scintillation

- **Along the path of a beam of light through the atmosphere, changes in intensity and direction not only occur due to extinction and refraction, but also due to *atmospheric turbulence*. Turbulence causes fluctuations in density, ρ, of the air — mostly as a result of temperature differences. Pressure fluctuations at small scales are negligible, since these would rapidly (at the sound speed!) smooth out. Temperature differences can be much longer-lived.**

  Let's estimate the effects due to such fluctuations. For changes in direction of a beam of light, any incremental change in refractive index, $\Delta n \equiv n - n_0$, is important. The refractive index of air is proportional to its density: $(n-1) \propto \rho$ and $n_{|\rho=0} \equiv n_{\text{vac}} \equiv 1$. For an ideal gas, we have:

$$\frac{n-1}{n_0-1} = \frac{\rho}{\rho_0} = \frac{P/T}{P_0/T_0} = \frac{P\,T_0}{P_0\,T} \tag{15}$$

  Because pressure differences can be ignored, $P = P_0$, from which follows (if we take $\Delta n/\Delta T \simeq \mathrm{d}n/\mathrm{d}T$):

$$\Delta n = -(n_0-1)\cdot\frac{T_0}{T^2}\cdot\Delta T \tag{16}$$

Seeing and Scintillation (cont'd)

At sea level, $T \simeq T_0 \simeq 300\,\mathrm{K}$ and $n_0 = 1.000293$, which gives $\Delta n \simeq -10^{-6}\Delta T$. At altitude h, the pressure and, hence, the density of the air and its refractive index decrease, and they do so roughly exponentially:

$$\frac{n_h - 1}{n_0 - 1} = e^{-h/H} \tag{17}$$

where H is the exponential scale height of the Earth's atmosphere ($H \simeq 8\,\mathrm{km}$). This then gives us (Eq. 17 $\longrightarrow$ Eq. 15 $\Rightarrow$ Eq. 16):

$$\Delta n \simeq -10^{-6} \cdot e^{-h/H} \cdot \Delta T \tag{18}$$

As it turns out, turbulence in the atmosphere results in *two* separate effects:

1. **it causes changes in direction of the incident beam of light, which we call *seeing*;**
2. **it causes fluctuations in intensity, which we call *scintillation*.**

In the following, we will consider a much simplified model of a single turbulence element (*turbulence cell*) of diameter L and refractive index $n_0 + \Delta n$, embedded in the ambient air (with index n_0), and reduce the problem to the two-dimensional geometry illustrated in Fig. 33.

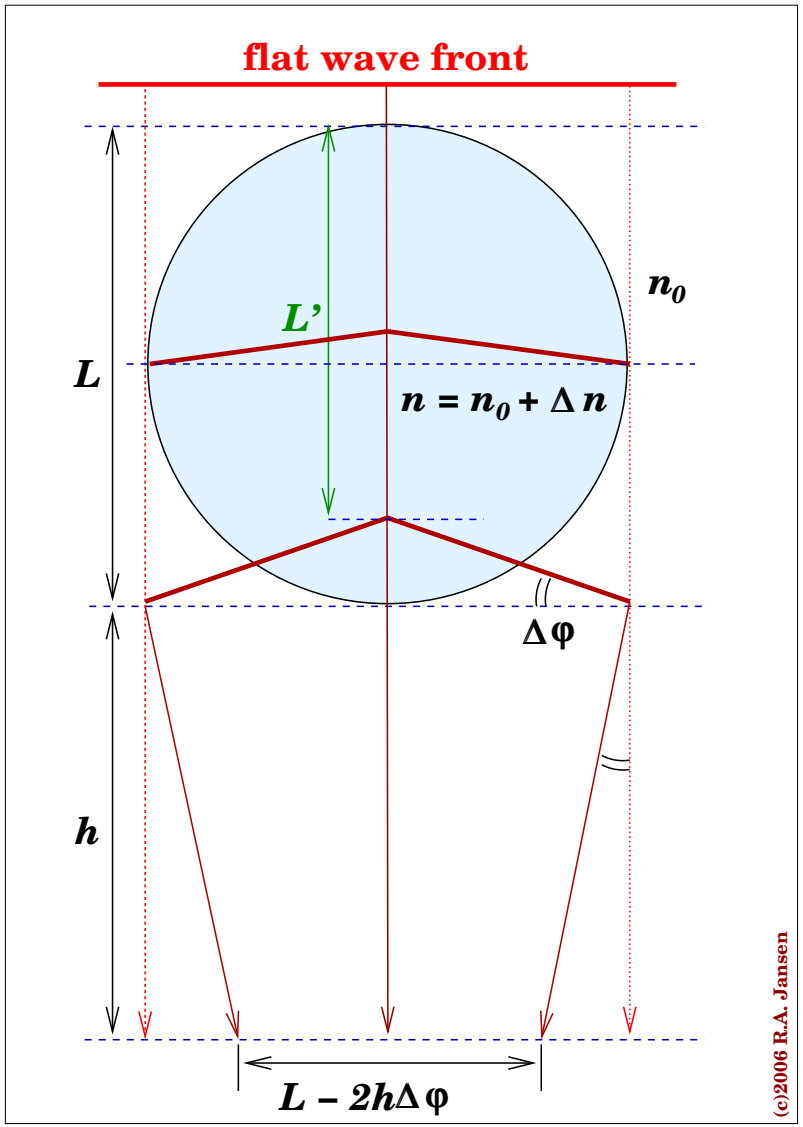


Figure 33: Deformation of a flat wave front by a turbulence cell with diameter L and refractive index $n = n_0 + \Delta n$, embedded in a medium with refractive index n_0.

Seeing and Scintillation (cont'd)

- ***Seeing***: **If a flat wave front is incident on a turbulence element, then the wave front will be distorted as a result of the difference in optical path length for refractive index $n_0 + \Delta n$ compared to that for the ambient index of n_0:**

$$L' \;=\; L \cdot \frac{n_0}{n_0 + \Delta n}$$

From which follows:

$$\Delta\phi \;=\; \frac{L - L'}{\frac{1}{2}\,L} \;=\; 2\,\frac{\Delta n}{n_0 + \Delta n} \;\simeq\; 2\,\Delta n \quad \text{(since } n_0 + \Delta n \simeq n_0 \simeq 1\,\text{)}$$

Hence:

$$\Delta\phi \;\simeq\; 2\,\Delta n \;=\; -2 \times 10^{-6} \cdot e^{-h/H} \cdot \Delta T \tag{19}$$

This explains why seeing is dominated by the turbulence within the first tens of meters above the telescope: due to the exponential decrease in density, $\Delta\phi$ rapidly decreases with increasing altitude h. Moreover, ΔT also diminishes with altitude.

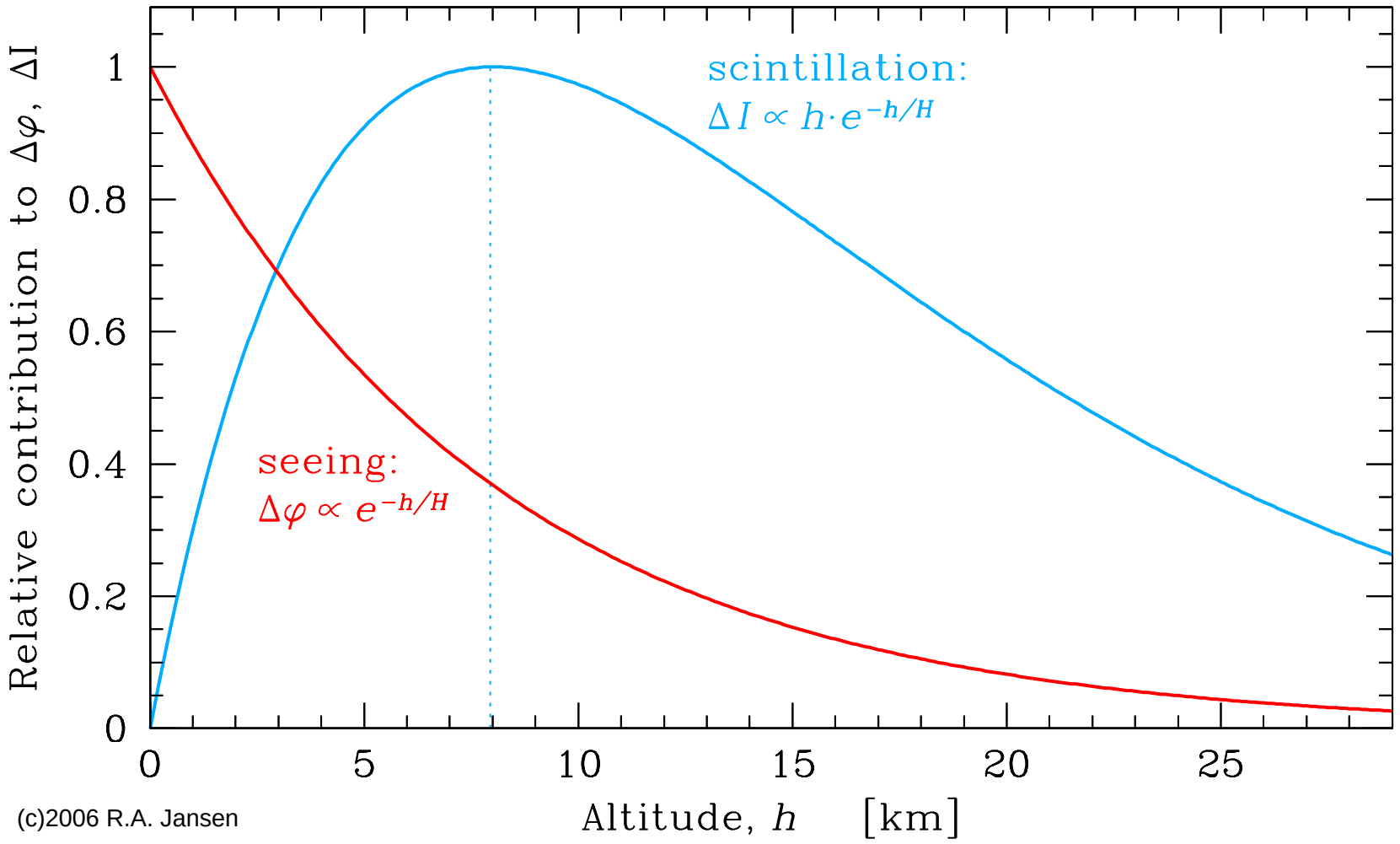


Figure 34: Relative contributions to the seeing and scintillation as a function of the altitude of the turbulence elements. The exponential scale height of the atmosphere is 8 km. The maximum contributions to seeing and scintillation are at quite different altitudes. This means that *seeing and scintillation are uncorrelated*.

Seeing and Scintillation (cont'd)

- ***Scintillation*: The intensity fluctuations as observed at ground-level due to the passage through a turbulence element is proportional to the change in beam diameter, hence $\Delta I \propto h \cdot \Delta\phi$. So now we're dealing with $h \cdot e^{-h/H}$ instead of with $e^{-h/H}$. This means that the maximum effect of scintillation occurs higher in the atmosphere. As it turns out, at an altitude of $\sim$8 km.**

  - **The difference in dominant altitude implies that *seeing and scintillation are uncorrelated*.**

If D denotes the diameter of a telescope, then:

- **if $D \lesssim L$: most of the time only one turbulence element will be within the beam $\Longrightarrow$ the entire (sharp) source image will be displaced but unsmeared at any one time, but that image will show strong fluctuation in intensity (i.e., scintillation);**
- **if $D \simeq L$: At $h \simeq 0$ and for $\Delta T = 1^\circ \quad \Longrightarrow \quad \Delta\phi = 2 \times 10^{-6}$ rad $= 0\farcs41$;**
- **if $D \gg L$: several turbulence cells will be within the beam at any one time: their combined displacements will smear the source image to a seeing-disk, but the net amplitude of scintillation is less severe.**

Seeing and Scintillation (cont'd)

- **A typical diameter of a turbulence cell of $\sim$10 cm at an altitude of 8–10 km subtends an angle of:**

$$\tan\theta \simeq \theta = \frac{10}{(8-10)\times 10^5}\cdot\frac{180}{\pi}\cdot 3600 = 2\farcs58 - 2\farcs06$$

  - **Because most planets subtend angles larger than $2\farcs6$ (Mars: 5–15″, Jupiter: 30–40″, Venus: 10–60″), planets do not scintillate. Twinkle, twinkle is a star after all...**