General Qualifying Exam Solutions: Physics and Fundamentals

Starkman, Nathaniel Lokken, Martine Ludwig, Bethany
Winch, Harrison

October 17, 2019

Contents

Q1) Interferometry

A two-element interferometer consists of two telescopes whose light is combined and interfered. Explain how this might be accomplished in practice, and sketch the response of such an interferometer to a nearby red giant star, as a function of the(projected) separation between the two telescopes.

Campbell Physics Q2

1.2 Question 2

A two-element interferometer consists of two telescopes whose light is combined and interfered. Sketch the response of such an interferometer to a nearby red giant star, as a function of the (projected) separation between the two telescopes. The red giant subtends one-fiftieth of an arc second on the sky, and the telescope operates at a wavelength of 2 microns.

1.2.1 Short answer

In a similar way that a single slit experiment results in an interference pattern depending on the slit size, an analogous interference pattern will emerge from a two-element interferometer and will depend on the projected baseline. Since the red giant star is resolved to 1/50 arcseconds, let's determine the projected baseline this corresponds to. Before we can do this, we must first convert the angular size from arcseconds into radians:

$$\theta = \left(\frac{1}{50}\,\text{arcsec}\right)\left(\frac{1\,\text{deg}}{3600\,\text{arcsec}}\right)\left(\frac{1\,\text{rad}}{57\,\text{deg}}\right) = 10^{-7}\,\text{rad}.$$

Now to determine the projected baseline:

$$\begin{aligned}\theta &\sim 2.2\left(\frac{\lambda}{D}\right)\ [\text{rad}]\\ D &\sim 2.2\left(\frac{\lambda}{\theta}\right)\ [\text{m}]\\ &= 2.2\left(\frac{2\times 10^{-6}\,\text{m}}{10^{-7}\,\text{rad}}\right) = 44\ [\text{m}].\end{aligned}$$

Thus, the resulting response function will be a Bessel function whose minima are at integer multiples of the 44 m projected baseline:

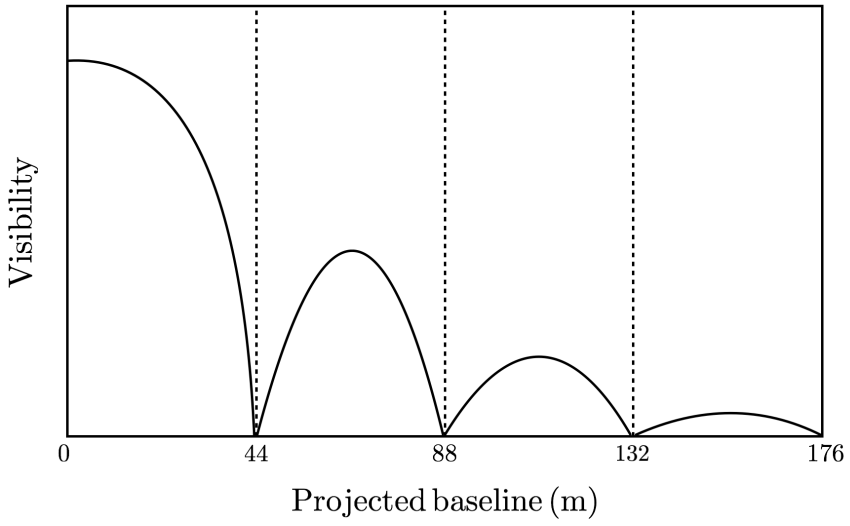


Figure 5: Response function of a two-element interferometer as a function of the projected baseline observing a resolved red giant star with an angular resolution of 1/50 arcseconds.

1.2.2 Additional context

The two-element quasi-monochromatic interferometer: The simplest radio interferometer is a pair of radio telescopes whose voltage outputs are correlated (multiplied and averaged), and even the most elaborate interferometers with $N \gg 2$ antennas, often called **elements**, can be treated as $N(N-1)/2$ independent two-element interferometers. Figure 6 shows two identical dishes separated by the **baseline vector** $\vec{b}$ of length b that points from antenna 1 to antenna 2. Both dishes point in the same direction specified by the unit vector $\hat{s}$, and θ is the angle between $\vec{b}$ and $\hat{s}$. Plane waves from a distant point source in this direction must travel an extra distance $\vec{b}\cdot\hat{s} = b\cos\theta$ to reach antenna 1, so the output of antenna 1 is the same as that of antenna 2, but it lags in time by the geometric delay

$$\tau_g = \frac{\vec{b} \cdot \hat{s}}{c} \text{ [s]}.$$

For simplicity, we first consider a quasi-monochromatic interferometer, one that responds only to radiation in a very narrow band $\Delta\nu \ll 2\pi/\tau_g$ centered on frequency $\nu = \omega/(2\pi)$. Then the output voltages of antennas 1 and 2 at time t can be written as

$$V_1 = V\cos[\omega(t-\tau_g)] \text{ [V]}$$
$$V_2 = V\cos(\omega t) \text{ [V]}.$$

These output voltages are amplified versions of the antenna input voltages; they have not passed through square-law detectors. Instead, a **correlator** multiplies these two voltages to yield the product

$$V_1V_2 = V^2\cos[\omega(t-\tau_g)]\cos(\omega t) = \left(\frac{V^2}{2}\right)[\cos(2\omega t - \omega\tau)g) + \cos(\omega\tau_g)] \text{ [V]}$$

which follows directly from the trigonometric identity $\cos x \cos y = [\cos(x+y) + \cos(x-y)]/2$. The correlator also takes a time average long enough ($\Delta t \gg (2\omega)^{-1}$) to remove the high-frequency term $\cos(2\omega t - \omega\tau_g)$ from the **correlator response** (output voltage) R and keep only the slowly varying term

$$R = \langle V_1V_2\rangle = \left(\frac{V^2}{2}\right)\cos(\omega\tau_g) \text{ [V}^2\text{]}.$$

The voltages V_1 and V_2 are proportional to the electric field produced by the source multiplied by the voltage gains of the two antennas and receivers. Thus the correlator output amplitude $V^2/2$ is proportional to the flux density S of the point source multiplied by $\sqrt{A_1A_2}$, where A_1 and A_2 are the effective collecting areas of the two antennas.

Notice that the time-averaged response R of a multiplying interferometer is zero. There is no DC output, so fluctuations in receiver gain do not act on the whole system temperature T_s as for a total-power observation with a single dish. Uncorrelated noise power from very extended radio sources such as the CMB and the atmosphere over the telescopes, also averages to zero in the correlator response. Short interference pulses with duration $t \ll |b|/c$ are also suppressed because each pulse does not reach both telescopes simultaneously. Likewise, a multiplying radio interferometer differs from a classical **adding interferometer**, such as the optical Michelson interferometer, that adds the uncorrelated noise power contributions.

The correlator output voltage $R = (V^2/2)\cos(\omega\tau_g)$ varies sinusoidally as the Earth's rotation changes the source direction relative to the baseline vector. These sinusoids are called **fringes**, and the **fringe phase**

$$\phi = \omega\tau_g = \frac{\omega}{c}b\cos\theta \text{ [rad]}$$

depends on θ as follows:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\theta} = \frac{\omega}{c}b\sin\theta = 2\pi\left(\frac{b\sin\theta}{\lambda}\right) \text{ [dimensionless]}.$$

The **fringe period** $\phi = 2\pi$ corresponds to an angular shift $\theta = \lambda/(b\sin\theta)$. The fringe phase is an exquisitely sensitive measure of source position if the **projected baseline** $b\sin\theta$ is many wavelengths long. Note that fringe phase and hence measured source position is not affected by small tracking errors of the individual telescopes. It depends on time, and times can be measured by clocks with much higher accuracy than angles (ratios of lengths of moving telescope parts) can be measured by rulers. Also, an interferometer whose baseline is horizontal is not affected by the plane-parallel component of atmospheric refraction, which delays the signals reaching both telescopes equally. Consequently, interferometers can determine the positions of compact radio sources with unmatched accuracy. Absolute positions with errors as small as $\sigma_\theta \approx 10^{-3}$ arcsec and differential positions with errors down to $\sigma_\theta \approx 10^{-5}$ arcsec $< 10^{-10}$ rad have frequently been measured.

If the individual antennas comprising an interferometer were isotropic, the interferometer point-source response would be a sinusoid spanning the sky. Such an interferometer is sensitive to only one Fourier component of the sky brightness distribution: the component with angular period $\lambda/(b\sin\theta)$. The response R of a two-element interferometer with directive antennas is that sinusoid multiplied by the product of the voltage patterns of the individual antennas. Normally the two antennas are identical, so this product is the power pattern of the individual antennas and is called the **primary beam** of the interferometer. The primary beam is usually a Gaussian much wider than a fringe period, as indicated in Figure 6. The convolution theorem states that the Fourier transform of the product of two functions is the convolution of their Fourier transforms, so the interferometer with directive antennas responds to

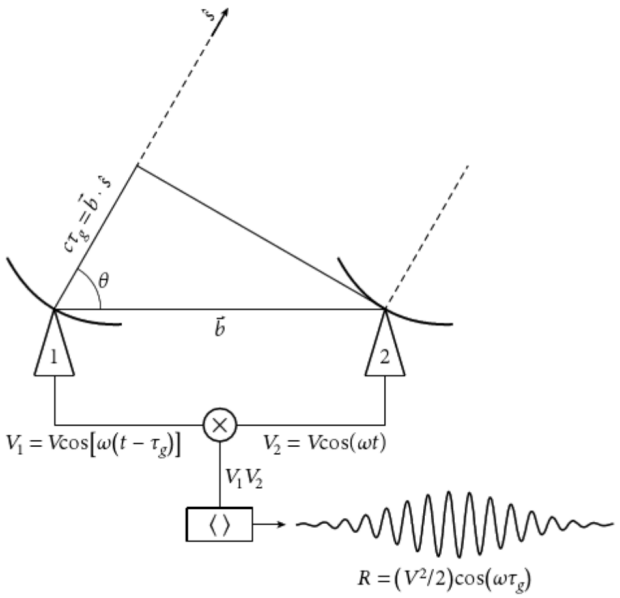


Figure 6: This block diagram shows the components of a two-element quasi-monochromatic **multiplying interferometer** observing in a very narrow radio frequency range centered on $\nu = \omega/(2\pi)$. $\hat{s}$ is the unit vector in the direction of a distant point source and $\vec{b}$ is the baseline vector pointing from antenna 1 to antenna 2. The output voltage V_1 of antenna 1 is the same as the output voltage V_2 of antenna 2, but it is retarded by the geometric delay $\tau_g = \vec{b}\cdot\hat{s}/c$ representing the additional light-travel delay to antenna 1 for a plane wavefront from a source at angle θ from the **baseline vector**. These voltages are amplified, multiplied ($\times$), and time averaged ($||$) by the **correlator** to yield an output response whose amplitude R is proportional to the flux density of the point source and whose phase ($\omega\tau_g$) depends on the delay and the frequency. The quasi-sinusoidal output **fringe** shown occurs if the source direction in the interferometer frame is changing at a constant rate $\mathrm{d}\theta/\mathrm{d}t$. The broad Gaussian envelope of the fringe shows the primary-beam attenuation as the source passes through the beam of the dishes. Figure taken from Condon & Random (2016).

a finite range of angular frequencies centered on $b\sin\theta/\lambda$. Because the antenna diameters D must be smaller than the baseline b (else the antennas would overlap), the angular frequency response cannot extend to zero and the interferometer cannot detect an isotropic source – the bulk of the 3 K CMB for example. The missing short spacings ($b < D$) can be provided by a single-dish telescope with diameter $D > b$. Thus the $D = 100$ m GBT can fill in the missing baselines $b < 25$ m that the $D = 25$ m VLA dishes cannot obtain.

Improving the instantaneous point-source response pattern of an interferometer requires more Fourier components; that is, more baselines. An interferometer with N antennas contains $N(N-1)/2$ pairs of antennas, each of which is a two-element interferometer, so the instantaneous **synthesized beam** (the point-source response obtained by averaging the outputs of all of the two-element interferometers) rapidly approaches a Gaussian as N increases. The instantaneous point-source responses of a two-element interferometer with projected baseline length b, a three-element interferometer with three baselines (projected lengths $b/3, 2b/3$, and b), and a four-element interferometer with six baselines (projected lengths $b/6, 2b/6, 3b/6, 4b/6, 5b/6$, and b) are shown in Figure 7.

Most radio sources are stationary; that is, their brightness distributions do not change significantly on the timescales of astronomical observations. For stationary sources, a two-element interferometer with movable antennas could make $N(N-1)/2$ observations to duplicate one observation with an N-element interferometer.

Slightly extended sources and the complex correlator: The response $R = (V^2/2)\cos(\omega\tau_g)$ of the quasi-monochromatic two-element interferometer with a "cosine" correlator (Figure 6) to a spatially incoherent slightly extended (much smaller than the primary beamwidth) source with sky brightness distribution $I_\nu(\hat{s})$ near frequency $\nu = \omega/(2\pi)$ is obtained by treating the extended source as the sum of independent point sources:

$$R_c = \int I(\hat{s})\cos(2\pi\nu\vec{b}\cdot\hat{s}/c)\mathrm{d}\Omega = \int I(\hat{s})\cos(2\pi\vec{b}\cdot\hat{s}/\lambda)\mathrm{d}\Omega\ [\mathrm{V}^2].$$

Notice that the even cosine function in this response is sensitive only to the even (inversion-symmetric) part I_E of an arbitrary source brightness distribution, which can be written as the sum of even and odd (anti-symmetric) parts: $I = I_E + I_O$. To detect the odd part I_O we need a "sine" correlator whose output is odd, $R = (V^2/2)\sin(\omega\tau_g)$. This can be implemented by a second correlator that follows a $\pi/2\,\mathrm{rad} = 90°$ phase delay inserted into the output of one antenna because $\sin(\omega\tau_g) = \cos(\omega\tau_g - \pi/2)$. Then

$$R_s = \int I(\hat{s})\sin(2\pi\vec{b}\cdot\hat{s}/\lambda)\mathrm{d}\Omega\ [\mathrm{V}^2]$$

The combination of cosine and sine correlators is called a **complex correlator** because it is mathematically convenient to treat the cosines and sines as complex exponentials using Euler's formula

$$e^{i\phi} = \cos\phi + i\sin\phi\ [\mathrm{rad}].$$

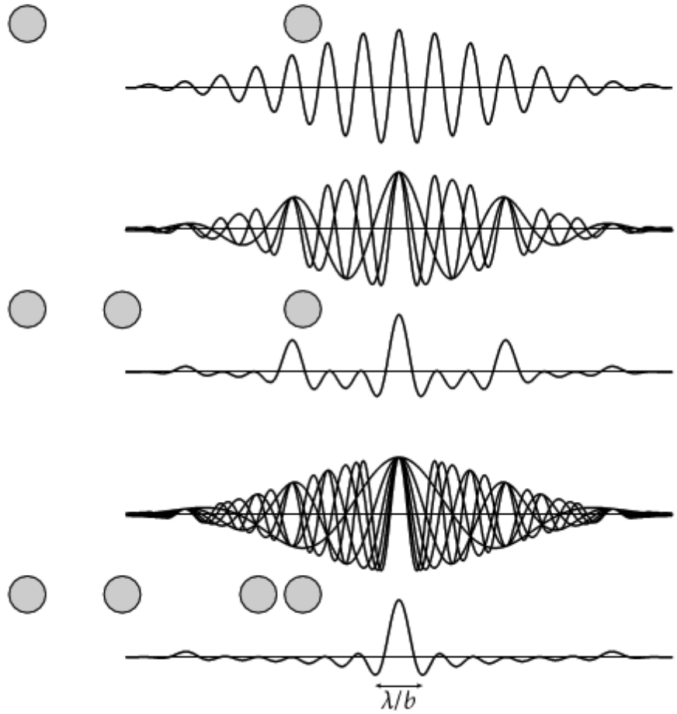


Figure 7: The instantaneous point-source responses of interferometers with overall projected length b and two, three, or four antennas distributed as shown are indicated by the thick curves. The synthesized main beam of the four-element interferometer is nearly Gaussian with angular resolution $\theta \approx \lambda/b$, but the sidelobes are still significant and there is a broad negative "bowl" caused by the lack of spacings shorter than the diameter of an individual antenna. Thus the **synthesized beam** is sometimes called the **dirty beam**. The instantaneous dirty beam of the multi-element interferometer is the arithmetic mean of the individual responses of its component two-element interferometers. The individual responses of the three two-element interferometers comprising the three-element interferometer and of the six two-element interferometers comprising the four-element interferometer are plotted as thin curves. Figure taken from Condon & Random (2016).

The **complex visibility** is defined by

$$\mathcal{V} \equiv R_c - iR_s \text{ [V}^2\text{]}$$

which can be written in the form

$$\mathcal{V} = Ae^{-i\phi} \text{ [rad]},$$

where

$$A = \sqrt{R_c^2 + R_s^2} \text{ [dimensionless]}$$

is the **visibility amplitude** and

$$\phi = \arctan\left(\frac{R_s}{c}\right) \text{ [rad]}$$

is the **visibility phase**. The response to an extended source with brightness distribution $I(\hat{s})$ of the two-element quasi-monochromatic interferometer with a complex correlator is the complex visibility

$$\mathcal{V} = \int I(\hat{s}) \exp(-i2\pi\vec{b} \cdot \hat{s}/\lambda) \mathrm{d}\Omega \text{ [V}^2\text{]}.$$

Fringe patterns and Fourier transforms: Interferometry begins with the Youngs slits fringe pattern (Fig. 1). With a single point source emitting coherent radiation, interference fringes are observed, with constructive and destructive interference observed as the relative delay of the two interfering rays changes; the separation of the fringes is /d, the wavelength of the light divided by the slit separation.
If the source is made wider (Figure 8b), we can think of it as a sequence of point sources each of which emit radiation which is uncorrelated with the emission from the others. It follows that the total interference intensity pattern is the sum of the individual patterns. Since an angular displacement in the source produces an equal angular displacement in the fringe pattern, as the source size approaches λ/d the fringe patterns will add to give a constant illumination (Figure 8c). In this case, the fringe visibility (defined as the difference between maximum and minimum intensity, normalized by the sum of maximum and minimum intensity) drops to zero. Conversely, when the angular size of the source is λ/d, the fringe visibility is 1; this corresponds to a situation in which the source size is smaller than the angular resolution of the interferometer, and only an upper limit of order λ/d can be obtained on it.
Now suppose that the slit spacing d is decreased. For the same size of source, this produces less "washing-out" of the fringes, because the same displacement of the source now produces much less displacement of the fringe patterns as a fraction of the fringe separation λ/d (Figure 8d). The smaller the slit separation,

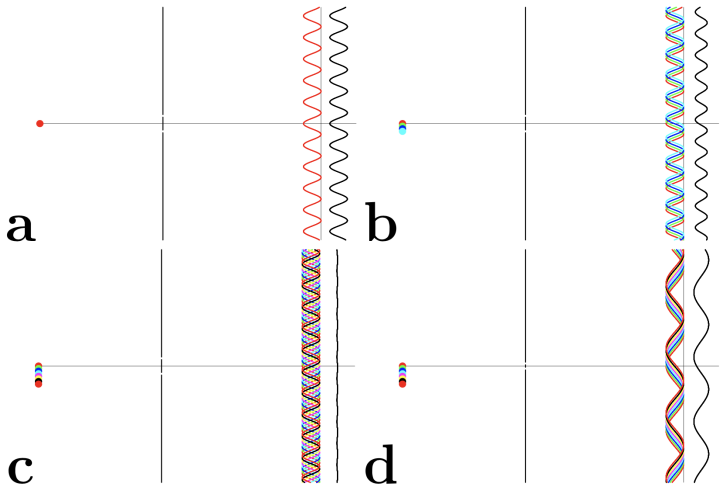
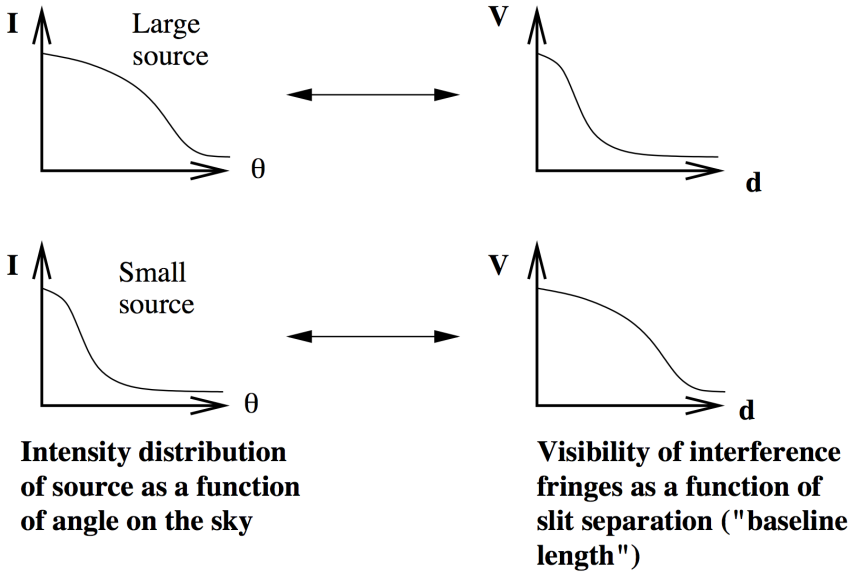


Figure 8: Young's slits in various situations. In each panel the source is shown on the left, and on the right of the slit are shown the fringe patterns separately for each part of the source and then the added fringe pattern. **(a)**: The basic two-slit pattern, showing fringes an angular distance λ/d apart. **(b)**: The effect of increasing the source size. An angular shift of the source position by θ shifts the fringe patterns by θ the other way. Since the patterns come from mutually incoherent sources, the intensity patterns add to give a pattern of reduced visibility. **(c)**: When the size of the source reaches λ/d, the fringes add to give zero visibility. **(d)**: If the slit spacing is then reduced, the fringe spacing increases, and the same size of source is still able to give visible fringes: the source would need to be increased in size to λ/d_{new} in order to wash out the fringes. Figure taken from Jackson.

Figure 9:
Relation between source brightness as a function of angular distance and visibility of interference fringes as a function of slit separation (baseline length). Figure taken from Jackson.



the larger the source size that can be probed using interferometry.
The situation is summarized in Figure 9. If we plot, for a given source distribution, the way in which visibility varies with slit separation, it can be seen that for small sources the visibility remains high out to large slit separation (in the limit of point sources, to infinite slit separation), whereas large sources produce visibility patterns which fall off quickly as the slit separation increases.
The relation between $I(\theta)$ and $V(d)$ represented here is one which maps a large Gaussian into a small Gaussian, and vice versa, and it is fairly obvious that it is a Fourier transform.[1]

1.2.3 Follow-up Questions

- What do the minima in the response function tell you?

[1] This is known as the Van **Cittert-Zernicke theorem**.

1.3 Question 3

What's the minimum mass of a black hole you could survive a fall through the event horizon without being ripped to shreds? Why would you be ripped to shreds for smaller black holes? How does this relate to the BH mass range for which we expect tidal disruption flares caused by shredding main-sequence stars?

1.3.1 Short answer

We can find the minimum mass using the equation for the tidal force one would experience:

$$F_T = \left(\frac{\mathrm{d}g}{\mathrm{d}r}\right)\ell = \frac{\mathrm{d}}{\mathrm{d}r}\left(-\frac{GM}{r^2}\right)\ell = -GM\ell\frac{\mathrm{d}}{\mathrm{d}r}r^{-2} = -GM\ell(-2)r^{-3} = \frac{2GM\ell}{r^3}\ [\mathrm{N}].$$

Setting the radius equal to the Schwarzschild radius $r_s \equiv 2GM/c^2$,

$$F_T = \frac{2GM\ell}{r_S^3} = 2GM\ell\left(\frac{c^2}{2GM}\right)^3 = \frac{\ell c^6}{4G^2M^2}\ [\mathrm{N}],$$

we can now solve for the mass of the black hole M:

$$M = \sqrt{\frac{\ell c^6}{4G^2F_T}}\ [\mathrm{kg}].$$

Let's set $\ell \sim 1\,\mathrm{m}$ for an order-of-magnitude length of the human body, and $F_T \sim 100\,\mathrm{g} = 10^4\,\mathrm{N}$ for the maximum force that the body can withstand. We can now use this to find the minimum mass of a black hole that you could survive a fall through the event horizon without being ripped to shreds:

$$M = \sqrt{\frac{(1\,\mathrm{m})(2.998\times10^8\,\mathrm{m\,s^{-1}})^6}{4(6.67^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}})^2(10^4\,\mathrm{N})}} = 10^{34}\,\mathrm{kg} \sim 10^3\,\mathrm{M_\odot}.$$

Since we can see that the tidal force $F_T \sim M^{-2}$, you would be ripped to shreds for smaller black holes because the resulting tidal force on the body would be greater the smaller the black hole mass is.

1.3.2 Follow-up Questions

- How would you estimate the maximum tidal acceleration a star can withstand?
- Why is it enough to know if the surface of the star will be disrupted?

Q2) Detector Cooling

You don't usually need to cool down the detectors for short wavelength (e.g., X-ray) observations, but it's critical to cool down the detectors in long wavelength (e.g.,far-IR) observations. Why is this, and why is it usually less essential or unnecessary for radio observations?

Campbell Physics Q11

1.11 Question 11

You don't usually need to cool down the detectors for short wavelength (e.g., X-ray) observations, but it's critical to cool down the detectors in long wavelength (e.g., far-IR) observations. Why is this, and why is it not necessary for radio observations?

1.11.1 Short answer

IR detectors use CCDs which are made up of semi-conducting silicon material; these detectors, when hit with a photon of the 'right' energy, will release a photoelectron which is later read by the CCD. The required energy for an incoming photon to release a photoelectron must be greater than silicon's band gap energy ($> 1.14\,\text{eV}$). Silicon has a useful photoelectric effect range of $1.1 - 10\,\text{eV}$, covering the range of near-IR to soft x-ray. While CCDs are generally more sensitive at higher temperatures, this results in a dark current of photoelectrons which cannot be distinguished from astrophysical sources. Thus, cooling detectors for short wavelength observations is necessary to reduce these dark currents specifically because they are sensitive to the IR whereas short wavelength observations are not. This is not necessary for radio observations because these observations use bolometers (instead of CCDs) which detect electromagnetic fields (instead of photons).

1.11.2 Additional context

Why use CCDs?: Most astronomical detectors in use today at professional observatories, as well as with many amateur telescopes, are CCDs. This fact alone gives an impression that there must be something very special or useful about CCDs; otherwise why all the fuss? CCDs have revolutionized modern astronomy. They will take their place in astronomical history along with other important discoveries such as the telescope, photographic plates, prisms, and spectroscopy. The contribution to our knowledge of the heavens brought about by CCDs is astounding, even more so when one considers that they have been in use for only about thirty years.

First introduced as electronic analogs to magnetic bubble memory at Bell labs, CCDs provided their first astronomical image in 1975 when scientists from the Jet Propulsion Laboratory imaged the planet Uranus at a wavelength of 8900Å. This observation used the University of Arizona $61''$ telescope atop Mt. Lemmon and a prototype version of a CCD made by Texas Instruments Corporation as part of a development project for NASA spacecraft missions.

During the past ten years, tremendous progress has been made in the manufacturing process and therefore in the properties of the CCD itself. These improvements have allowed much lower noise properties for CCDs, thereby increasing their overall efficiency in astronomy. In addition, larger format devices have been produced and the readout times are much shorter, approaching $1 - 2$ seconds even for arrays as large as 1024 pixels square. This latter advance is mainly due to the availability of high-speed, low-power and low-noise CCD controllers. The driving technology for CCD manufacturing is for items such as copy machines, TV cameras, and digital cameras, but the requirements for low noise, excellent pixel cosmetics, and nearly perfect performance is still firmly rooted in astronomy. We outline below two of the important reasons why CCDs are considered as essentially the perfect imaging device.

Noise properties: The usefulness of a detector is very often determined by the amount of inherent noise within the device itself. Suffice it to say that modern astronomical CCDs are almost noise free. The original line of photosensitive electronic array detectors, such as television-type imaging detectors, vidicons, silicon intensified targets, and image-dissector scanners, all had very high noise properties. For comparison, silicon intensified target imagers (SITs) had a noise level upon readout of 800 electrons per picture element. Some very good systems of this type could be produced with read noise values of only 200 electrons. The first CCDs had readout noise levels similar to this latter value, while modern CCDs have noise values of ten down to two electrons per pixel per readout. The large noise levels present in early array detectors not only limited the SNR obtainable for a given measurement, they also severely limited the total dynamic range available to the camera. Another "feature" of older, higher noise CCDs was the decision an astronomer had to make about co-addition of frames. Since the read noise adds as its square to the total noise budget, adding two frames resulted in a much higher read noise contribution. Today, with typical read noise values of $2-5$ electrons, co-addition is essentially equal to a single exposure of longer integration time.

CCD operation: The simplest and very understandable analogy for the operation of a CCD is also one that has been used numerous times for this purpose. This is the "water bucket" idea in which buckets represent pixels on the CCD array, and a rainstorm provides the incoming photons (rain drops). Imagine a field covered with buckets aligned neatly in rows and columns throughout the entirety of the area (Figure 18). After the rainstorm (CCD integration), each bucket is transferred in turn and metered to

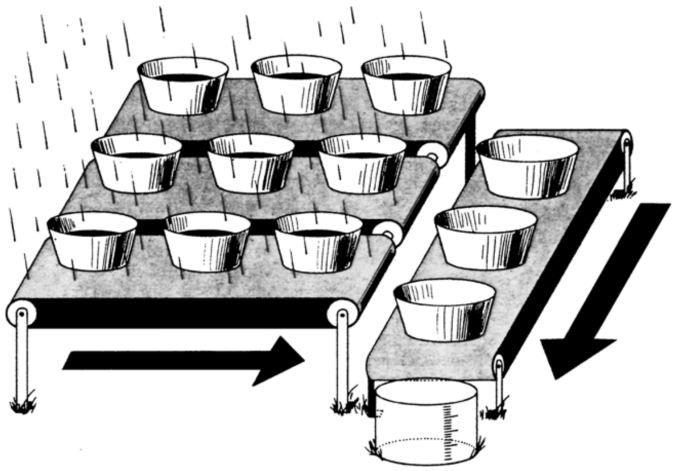


Figure 18:
CCDs can be likened to an array of buckets that are placed in a field and collect water during a rainstorm. After the storm, each bucket is moved along conveyor belts until it reaches a metering station. The water collected in each field bucket is then emptied into the metering bucket within which it can be measured. From Janesick & Blouke (1987). Figure taken from Howell (2006).

determine the amount of water collected. A written record (final CCD image) of the amount of water in each bucket will thus provide a two-dimensional record of the rainfall within the field.

Referring to the actual mechanisms at work within a CCD, we start with the method of charge generation within a pixel: the photoelectric effect.[3] Incoming photons strike the silicon within a pixel and are easily absorbed if they possess the correct wavelength (energy). Silicon has a band gap energy of 1.14 eV, and so it easily absorbs light of energy 1.1 to 4 eV (3000 to 11000Å).[4] Photon absorption causes the silicon to give up a valence electron and move it into the conduction band. Photons of energy 1.1 eV to near 4 or so eV generate single **electron-hole pairs**, whereas those of higher energy produce multiple pairs. Left to themselves, these conduction band electrons would recombine back into the valence level within approximately 100 μs. Silicon has a useful photoelectric effect range of 1.1 to about 10 eV, which covers the near-IR to soft X-ray region. Above and below these limits, the CCD material appears transparent to the incoming photons.

Once electrons have been freed to the conduction band of the silicon, they must be collected and held in place until readout occurs. The details of the actual construction of each pixel within a CCD (that is, the formation of the MIS capacitor with its doped silicon, layers of silicon dioxide, etc.) are beyond the scope of this discussion, but suffice it to say that each pixel has a structure allowing applied voltages to be placed on sub-pixel sized electrodes called **gates**. These gate structures provide each pixel with the ability to collect the freed electrons and hold them in a potential well until the end of the exposure. In a typical arrangement, each pixel has associated with it three gates, each of which can be set to a different voltage potential. The voltages are controlled by clock circuits with every third gate connected to the same clock. Figure 19 illustrates this clocking scheme for a typical three-phase device.

We note in Figure 19 that, when an exposure ends, the clock voltages are manipulated such that the electrons that have been collected and held in each pixel's +10 volt potential well by clock voltage V3 can now be shifted within the device. Note that electrons created anywhere within the pixel during the exposure (where each pixel has a surface area equal to the total area under all three gates) will be forced to migrate toward the deepest potential well. When the exposure is terminated and CCD readout begins, the voltages applied to each gate are cycled (this process is called **clocking the device**) such that the charge stored within each pixel during the integration is electronically shifted. A simple change in the voltage potentials (V3 goes to +5 volts, while V1 becomes +10 volts and so on) allows the charge to be shifted in a serial fashion along columns from one CCD pixel to another throughout the array. The transfer of the total charge from location to location within the array is not without losses. Each charge transfer (one of which occurs for each voltage change or clock cycle) has an associated efficiency. This efficiency value is the percent of charge transferred compared with that which was actually collected. Modern values for the charge transfer efficiency (CTE) are approaching 0.999999 (i.e., 99.9999% efficient) for each transfer.

Each column in the array is connected in parallel and thus each pixel shift is mimicked throughout the entire array simultaneously. One clock cycle moves each row of pixels up one column, with the top row being shifted off the array into what is called the **output shift register** or **horizontal shift register**. This register is simply another row of pixels hidden from view (i.e., not exposed to incident light) and serves as the transition between active rows on the array and the output of the device. Once an entire

[3] Albert Einstein received his Nobel Prize mainly for his work on the photoelectric effect, not, as many think, for relativity.
[4] The energy of a photon of a given wavelength (in electron volts) is given by $E_{\mathrm{eV}} = 12407/\lambda(\text{Å})$.

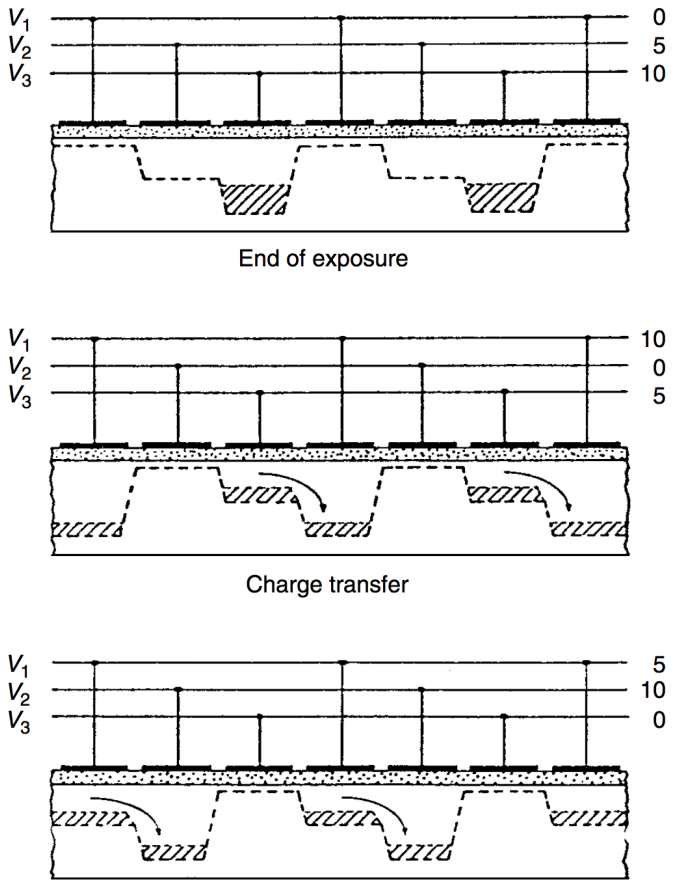


Figure 19:
Schematic voltage operation of a typical three-phase CCD. The clock voltages are shown at three times during the readout process, indicating their clock cycle of 0, 10, and 5 volts. One clock cycle causes the stored charge within a pixel to be transferred to its neighboring pixel. CCD readout continues until all the pixels have had their charge transferred completely out of the array and through the A/D converter. From Walker (1987). Figure taken from Howell (2006).

row is shifted into the output register, and before any further row shifts on the active area occur, each pixel in the output register is shifted out one at a time (in a similar manner as before) into the output electronics. Here, the charge collected within each pixel is measured as a voltage and converted into an output digital number. Each pixel's collected charge is sensed and amplified by an **output amplifier**. CCD output amplifiers are designed to have low noise and are built directly into the silicon circuitry; thus they are often referred to as **on-chip amplifiers**. These amplifiers must work with extremely small voltages and are rated, as to their sensitivity, in volts per electron. Typical values are in the range of 0.5 to 4 microvolts per electron.

The output voltage from a given pixel is converted to a **digital number** (DN) and is typically discussed from then on as either **counts** or **analog-to-digital units** (ADUs). The amount of voltage needed (i.e., the number of collected electrons or received photons) to produce 1 ADU is termed the **gain** of the device. A typical CCD gain might be 10 electrons/ADU, which means that for every 10 electrons collected within a pixel, the output from that pixel will produce, on average, a count or DN value of 1. For example, with this gain value if a pixel collects 1000 electrons (photons), the output pixel value stored in the computer would be 100 ADUs. For 1500 electrons 150 ADUs would be produced and for 17234 electrons, the output pixel value would be 1723 ADUs (note, not 1723.4). Digital output values can only be integer numbers and it is clear already that the discrimination between different pixel values can only be as good as the resolution of the gain and digital conversion of the device.

Conversion of the output voltage signal into a DN is performed within a device called an **analog-to-digital converter** (A/D or ADC). There is an intimate connection between the number of digital bits available in the A/D and the value of the gain that can or should be used for the CCD. The output DNs are usually stored initially in computer memory and then moved to disk for storage and later manipulation. The process of shifting each entire CCD row into the output register, shifting each pixel along within this register, and finally performing the voltage conversion of each pixel's stored charge by the A/D to produce a DN value is continued until the entire array of pixels has been read out. For large-format CCD arrays, this process can take upwards of a few minutes to complete a single read out of the entire device. Note that for a 2048×2048 CCD, the charge collected in the last pixel to be read out has to be transferred over four thousand times. However, most modern large-format CCDs or mosaic cameras containing many large CCDs use a few tricks to readout faster. Single monolithic CCDs usually have 2

or 4 output amplifiers available (one in each corner) and given the proper electronic setup, these large chips are often read out from 2 or 4 corners simultaneously, thus decreasing the total readout time by $2-4$. For a mosaic of CCDs, this same process can read the entire array (using multiple amplifiers on each CCD) much faster than even one single large CCD.

The array size of a single CCD, as well as the size of a given pixel on a device, is controlled by the current limitations of manufacturing. How large one can make a good quality, large-scale integrated circuit and how small one can make an MIS capacitor, both of which have demanding requirements for near perfect operation, set the scale of CCD and pixel sizes that are available. CCDs as large as 5040×10080 and 7168×9216 pixels and pixels as small as $2-10\,\mu$m have been successfully produced.

Modern CCDs have much higher processing standards than even five years ago. Items such as multi-layer registration on the silicon wafer on the photomasks used in the production of the CCD integrated circuit and the ability to make smaller electrical component parts on the wafers (such as output amplifiers) lead to much lower noise characteristics, better pixel charge manipulation, and the ability for faster readout speeds with lower noise. For example, better alignment of the CCD layers in each pixel allow lower clock voltages to be used (as low as 2 volts has been demonstrated) leading to lower overall power consumption. This fact, in turn, allows for items such as plastic packaging instead of ceramic, reducing overall packaging costs, a cost that often rivals that of the CCD itself.

As you might imagine, astronomy is not the driving force for CCD manufacturing. Video devices, cell phones, security cameras, Xerox machines, etc. are the global markets boosting the economy of CCD makers. The trend today is to produce CCDs with small pixels ($10-12\,\mu$m for astronomy down to $\sim 2\,\mu$m for other applications) in order to increase image resolution. Small pixels (and small CCDs) have lower cost and higher yield but the small pixels have shallow well depths. This is somewhat compensated for using fast readout techniques and/or microlens arrays, which focus light from an incoming source onto each small CCD pixel. Not all CCD pixels are desired to have shallow wells. The CCDs produced by E2V for the NASA Kepler Discovery mission have $27\,\mu$m pixels with well depths of nearly 1 million electrons each and a capacity of > 50000 electrons per pixel is quite common in astronomy. Even CCDs with built-in electronic shutters are being experimented with. Each pixel contains a $p^+ - n - p^-$ vertical overflow drain (VOD) photodiode structure on its top through which the incoming light passes. The absorption of incoming light when the "shutter" is open is minimal and, within a few hundred nanoseconds, the electronic shutter can be biased and become opaque.

Quantum efficiency: The composition of a CCD is essentially pure silicon. This element is thus ultimately responsible for the response of the detector to various wavelengths of light. The wavelength dependence of silicon can be understood in an instant by glancing at Figure 20. Shown here is the length of silicon needed for a photon of a specific wavelength to be absorbed. Absorption length is defined as the distance for which 63% $(1/e)$ of the incoming photons will be absorbed. Figure 20 clearly shows that, for light outside the range of about 3500 to over 8000Å, the photons (1) pass right through the silicon, (2) get absorbed within the thin surface layers or gate structures, or (3) simply reflect off the CCD surface. At short wavelengths, 70% or more of the photons are reflected, and for very short wavelengths (as for long wavelengths) the CCD becomes completely transparent. Thus the quantum efficiency of a typical CCD device will approximately mirror the photon absorption curve for silicon. Shortward of ~ 2500Å (for thinned devices) or about ~ 25Å (for thick devices) the detection probability for photons increases again. However, owing to their much higher energy, these photons lead to the production of multiple electron-hole pairs within the silicon and may also produce damage to the CCD itself.

CCD quantum efficiencies are therefore very dependent on the thickness of the silicon that intercepts the incoming photons. This relation between absorption probability and CCD thickness is why front-side illuminated (thick) devices are more red sensitive (the photons have a higher chance of absorption) and why they have lower overall (blue) QEs (since the gate structures can be close to or even exceed the necessary absorption depths of as small as only a few atomic layers). A few front-side CCDs have been produced with special gate structures that are transparent to incoming blue and UV photons. In thinned devices, the longest wavelength photons are likely to pass right through the CCD without being absorbed at all.

Figure 21 (left) shows the quantum efficiencies for various imaging devices. Note that the y scale is logarithmic and the much superior QE provided by CCDs over previous detectors. Figure 21 (right) shows a selection of modern CCD QEs. The large difference in QE that used to exist between thinned and thick CCDs is now mostly eliminated due to manufacturing processes and coatings although other differences (such as location of peak QE, cosmic ray detection, etc.) remain. Quantum efficiency or QE curves allow one quickly to evaluate the relative collecting power of the device as a function of wavelength. Measured QE curves, such as in Figure 21 (right) and those shown in the literature, are generally assumed to be representative of each and every pixel on the device, that is, all pixels of a given device are assumed to work identically and have the same wavelength response. This is almost true, but it is the "almost" that makes flat fielding of a CCD necessary. In addition, the QE curves shown or delivered with a

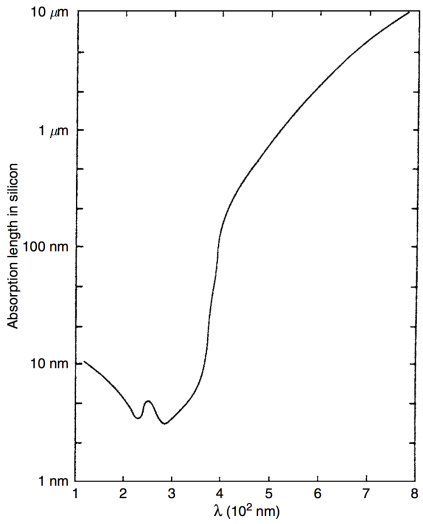


Figure 20:
The photon absorption length in silicon is shown as a function of wavelength in nanometers. From Reicke (1994). Figure taken from Howell (2006).

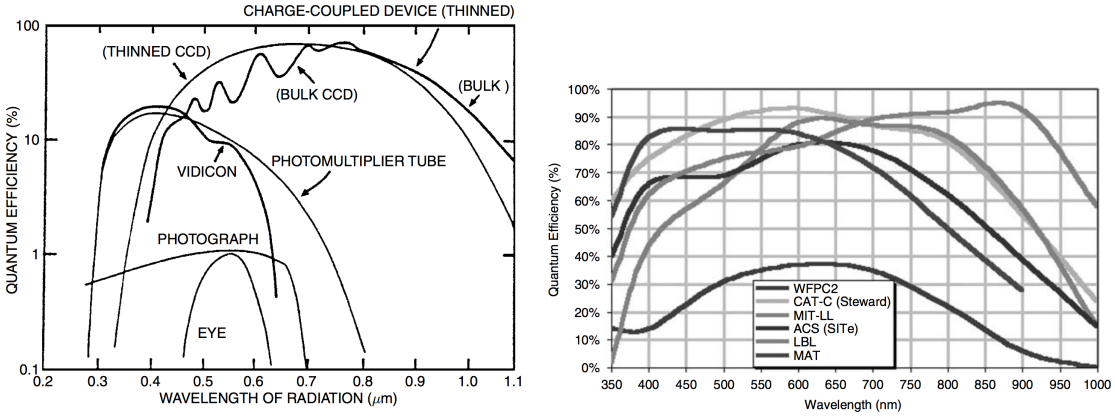


Figure 21: (**left**): QE curves for various devices, indicating why CCDs are a quantum leap above all previous imaging devices. The failure of CCDs at optical wavelengths shorter than about 3500Å has been essentially eliminated via thinning or coating of the devices. (**right**): QE curves for a variety of CCDs. WFPC2 is the second generation wide-field/planetary camera aboard HST, CAT-C is a new generation SITe CCD used in a mosaic imager at the University of Arizona's 90″ telescope on Kitt Peak, MIT-LL is a CCD produced at the MIT Lincoln Laboratories and optimized for red observations, ACS is the Hubble Space Telescope Advanced Camera for Surveys SITe CCD, LBL is a Lawrence Berkeley Lab high resistivity, "deep depletion" CCD with high red QE, and MAT is a front-side, processed CCD showing high blue QE. Figures taken from Howell (2006).

particular device may only be representative of a "typical" device of the same kind, but they may not be 100% correct for the exact device of interest.

The quantum efficiency of a CCD is temperature sensitive especially in the red wavelength region. It has long been known that measurement of the QE at room temperature is a poor approximation to that which it will have when operated cold. Thus QE curves should be measured at or near the operating temperature at which the CCD will be used. As an example of the temperature sensitivity of the efficiency of a CCD, Figure 22 shows three QE measurements of the same CCD for temperatures of $+20^\circ$ C ($\sim$ room temperature), -40° C, and -100° C ($\sim$ operating temperature). Note that the variations are small at 8000Å but increase to 20% at $9000 - 10000$Å. The curves in this plot would lead one to the conclusion that operating a CCD at room temperature is the best thing to do. However, this is *not* the case and a compromise between operating temperature (i.e., dark current) and delivered QE must be used.

Charge diffusion: Once an electron is captured in a CCD pixel, the voltages applied during integration attempt to hold it in place. However, situations arise within a CCD pixel that provide a finite possibility for any given electron to wander out of its collection pixel and into a neighboring pixel. This process is called **charge diffusion** and until recently it was noted but of low significance compared with other noise and readout issues. Today CCDs are of excellent quality and have very low readout noise, good pixel registration on the array, and reside in high quality optical systems. These facts mean that CCD imaging now has the ability to show great detail of any optical aberrations and geometric distortions. Even items such as better mechanical tolerances in instrumentation can reveal noticeable focus variations

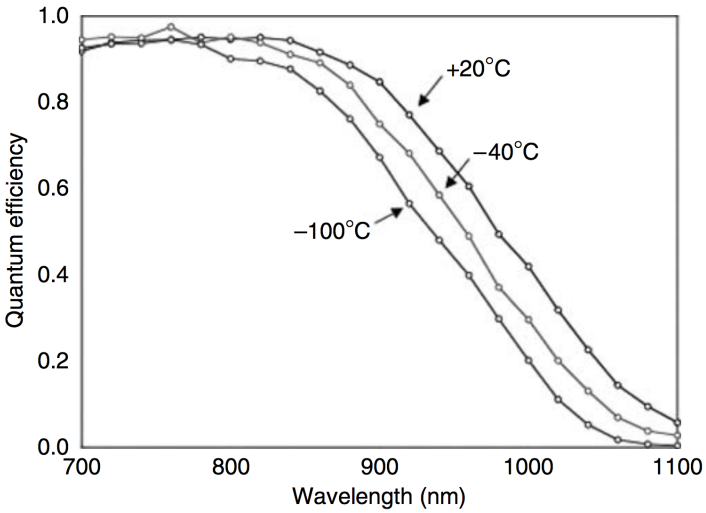


Figure 22:
Sensitivity of the quantum efficiency of a MIT/LL CCD for three operating temperatures. The blue sensitivity is little affected by a change in operating temperature but the red QE can change by a factor of two. The use of such devices requires a balance of higher operating temperature and keeping the dark current under control. Figure taken from Howell (2006).

as materials breathe with thermal changes. Given CCDs with deep pixel wells, large format front-side illuminated thinned devices, and the related improvements to modern astronomical instrumentation, the effects of charge diffusion on the point-spread function are noticeable.

A few ways in which charge diffusion can occur may be useful to discuss. Imagine a deep (geometrically long) pixel modeled after that which is shown in Figure 23. Electrons produced by long wavelength photons are captured in the front-side illuminated pixel near the bottom, far from the applied voltages in the front gates. Thus the potential well for these electrons is more like a potential dip. Given the right circumstances, an electron can move into a neighboring pixel. Another example would be impurities in the silicon material the CCD was manufactured from. These variations in the Si lattice can slightly bend or slightly redirect the potential within a pixel and provide weak spots from which electron escape is possible. Ways to mitigate electron loss are the use of higher potential voltages (although this can lead to other issues such as logic glow or shorting on the array), higher resistivity Si to more tightly hold the electrons (the Si lattice) in place, or to use small pixels (but these have lower red QE and small well depths). Again, we see that compromise and specific application come into play and can be tuned into the CCD as a part of its production process.

Charge diffusion often varies significantly across a CCD, especially thinned devices as they are not equal thickness everywhere. For example, the HST ACS wide field camera thinned CCD shows a variation in the core of the PSF, caused by charge diffusion, across the field of view. The variation is $30-40\%$ at 5000Å with larger variations at shorter wavelengths. The effects of charge diffusion are not to be taken lightly. The ACS/WFC suffers a loss of about 0.5 magnitudes in its limiting magnitude at short wavelengths and near 0.2 magnitudes elsewhere. Charge diffusion is especially important in thinned devices that under-sample the PSF.

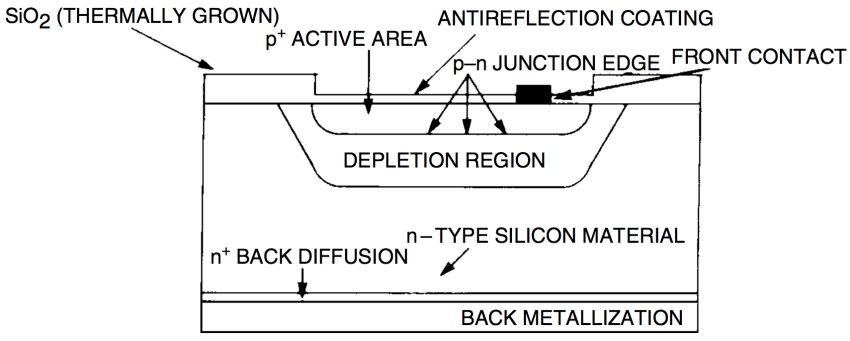


Figure 23:
Schematic view of a single front-side illuminated CCD pixel. The square labeled "front contact" is a representation of part of the overall gate structure. The letters "p" and "n" refer to regions within the pixel consisting of silicon doped with phosphorus and boron respectively. Figure taken from Howell (2006).

Charge transfer efficiency: Charge transfer efficiency or CTE is a measure of the fraction of the charge that is successfully transferred for each pixel transfer. CTE values of 0.999995 or more are typical in good modern CCDs. For a CCD with 1024×1024 pixels, the charge collected in the last pixel readout has shifted 2048 times thus the CTE must be nearly 100% in order to preserve the charge in each pixel during readout. CTI (charge transfer inefficiency) is 1-CTE or numerically near $10-5$ or $10-6$ in value. CTI can be and usually is different in the vertical and horizontal directions. The loss in charge from a CCD

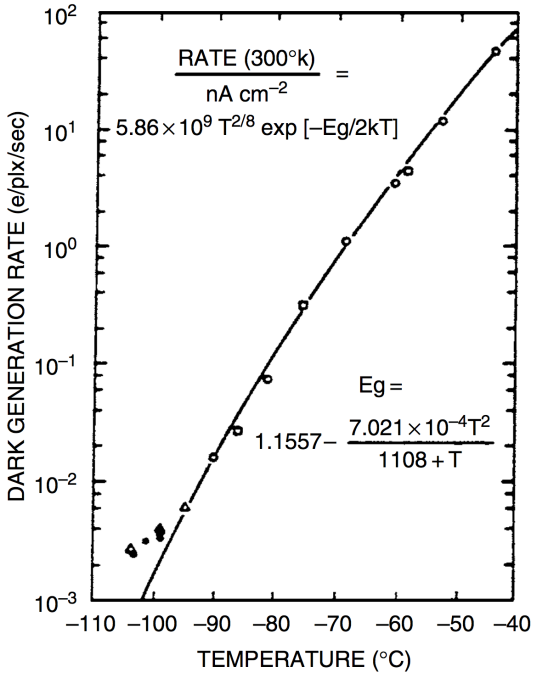


Figure 24:
Experimental (symbols) and theoretical (line) results for the dark current generated in a typical three-phase CCD. The rate of dark current, in electrons generated within each pixel every second, is shown as a function of the CCD operating temperature. E_g is the band gap energy for silicon. From Robinson (1988a). Figure taken from Howell (2006).

pixel containing N electrons that is shifted 1024 times vertically and 1024 times horizontally is given by $L(e) = N(1024 \times \text{CTI}(H) + 1024 \times \text{CTI}(V))$, or, if a single CTI value is given, $L(e) = 2048 \times N \times \text{CTI}$. CCDs with poor CTE generally show charge tails in the direction opposite readout for bright stars. These tails are the charge left behind as the image is shifted out.

The standard method for measuring CTE is to use X-ray stimulation of a CCD with a Fe^{55} source. CCDs are good X-ray detectors and for a specific X-ray source such as Fe^{55}, each X-ray photon collected produces ~ 1620 electrons.[5] A CCD is exposed to X-ray illumination and the resulting image readout. An X-ray transfer plot is made of signal in DN (y-axis) vs. running pixel number on the x-axis. Often hundreds of rows are summed together to increase the signal generated by the weak X-ray source. If the CCD has good CTE, a horizontal line will be seen at 1620 electrons (assuming a gain of 1.0). If the CTE is poor, this line starts at 1620 electrons (for rows close to the output amplifier) but tilts toward lower signal values for pixels further away from the output amplifier. This behavior indicates a loss of charge being transferred due to poor CTE. The CTE of a given CCD will generally degrade rapidly with decreasing operating temperature and is also a function of the clocking pulse shape and speed.

Readout noise: CCDs can be though of as having three noise regimes: **read noise**, **shot noise**, and **fixed pattern noise**. In astronomy, we speak of these as **read noise limited**, **photon noise limited**, and **flat field uncertainties**. A plot of the log of the standard deviation of the signal (y-axis) vs. the log of the signal itself (x-axis) for a CCD is called the **photon transfer curve**. Read noise (or any noise independent of signal level) sets a noise floor for a device. Upon illumination, photon or Poisson noise raises the sigma measured following a $\sqrt{N}$ slope. Finally, for large signal values, pixel to pixel variations due to processing errors and photomask misalignment begin to dominate. This latter noise is proportional to the signal and rises with a slope of 1.0. Full well sets in at some very high illumination and the slope of the photon transfer curve turns over or breaks.

Readout noise, or just read noise, is usually quoted for a CCD in terms of the number of electrons introduced per pixel into your final signal upon readout of the device. Read noise consists of two inseparable components. First is the conversion from an analog signal to a digital number, which is not perfectly repeatable. Each on-chip amplifier and A/D circuit will produce a statistical distribution of possible answers centered on a mean value.[6] Thus, even for the hypothetical case of reading out the same pixel twice, each time with identical charge, a slightly different answer may be produced. Second, the electronics themselves will introduce spurious electrons into the entire process, yielding unwanted random fluctuations in the output. These two effects combine to produce an additive uncertainty in the final output value for each pixel. The average (1σ) level of this uncertainty is the read noise and is limited by the electronic properties of the on-chip output amplifier and the output electronics.

[5] Remember that for optical photon detection, one photon collected produces one photoelectron, regardless of its wavelength. For the much higher energy X-rays, a single photon collected produces multiple electrons in proportion to the photon's energy.

[6] The distribution of these values is not necessarily Gaussian

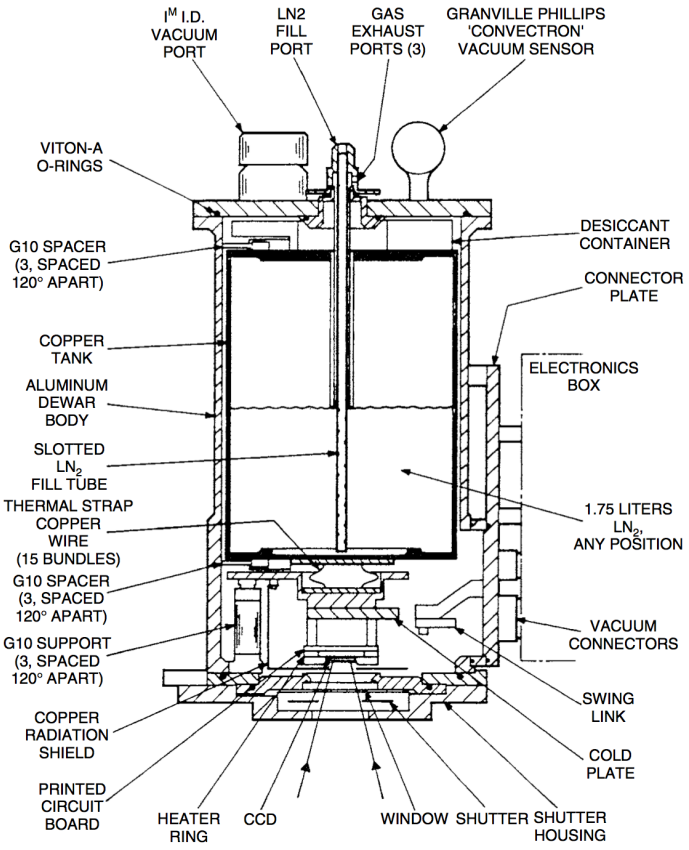


Figure 25:
A typical CCD dewar. This is the Mark-II Universal dewar originally produced in 1984 at Kitt Peak National Observatory. The dewar held 1.75 liters of liquid nitrogen providing a CCD operating time of approximately 12 hours between fillings. This dewar could be used in up-looking, down-looking, and side-looking orientations. From Brar (1984). Figure taken from Howell (2006).

In the output CCD image, read noise is added into every pixel every time the array is readout. This means that a CCD with a read noise of 20 electrons will, on average, contain 20 extra electrons of charge in each pixel upon readout. High read noise CCDs are thus not very good to use if co-addition of two or more images is necessary. The final resultant image will not be quite as good as one long integration of the same total time, as each co-added image will add in one times the read noise to every pixel in the sum. However, for modern CCDs, read noise values are very low and are hardly ever the dominant noise with which one must be concerned. Good read noise values in today's CCDs are in the range of 10 electrons per pixel per read or less. These values are far below read noise levels of ten years ago, which were as high as $50 - 100$ electrons, and even those are well down from values of $300 - 500$ or more electrons/pixel/read present in the first astronomical CCDs.

Dark current: Every material at a temperature much above absolute zero will be subject to thermal noise within. For silicon in a CCD, this means that when the thermal agitation is high enough, electrons will be freed from the valence band and become collected within the potential well of a pixel. When the device is read out, these dark current electrons become part of the signal, indistinguishable from astronomical photons. Thermal generation of electrons in silicon is a strong function of the temperature of the CCD, which is why astronomical use generally demands some form of cooling. Figure 24 shows a typical CCD dark current curve, which relates the amount of thermal dark current to the CCD operating temperature. Within the figure the theoretical relation for the rate of thermal electron production is given.

Dark current for a CCD is usually specified as the number of thermal electrons generated per second per pixel or as the actual current generated per area of the device (i.e., picoamps cm^{-2}). At room temperature, the dark current of a typical CCD is near 2.5×10^4 electrons/pixel/second. Typical values for properly cooled devices range from 2 electrons per second per pixel down to very low levels of approximately 0.04 electrons per second for each pixel. Although 2 electrons of thermal noise generated within a pixel every second sounds very low, a typical 15 minute exposure of a faint astronomical source would include 1800 additional (thermal) electrons within each CCD pixel upon readout. These additional charges cannot, of course, be uniquely separated from the photons of interest after readout. The dark current produced in a CCD provides an inherent limitation on the noise floor of a CCD. Because dark noise has a Poisson distribution, the noise actually introduced by thermal electrons into the signal is proportional to the square root of the dark current.

Campbell Physics Q11

Cooling of CCDs is generally accomplished by one of two methods. The first, and usually the one used for scientific CCDs at major observatories, is via the use of liquid nitrogen (or in some cases liquid air). The CCD and associated electronics (the ones on or very near the actual CCD itself, called the head electronics) are encased in a metal dewar under vacuum. Figure 25 shows a typical astronomical CCD dewar. The liquid nitrogen (LN2) is placed in the dewar and, although not in direct physical contact with the CCD, cools the device to temperatures of near $-100°C$. Since LN2 itself is much colder than this, CCDs are generally kept at a constant temperature $\pm 1°C$ with an on-board heater. In fact, the consistency of the CCD temperature is very important as the dark current is a strong function of temperature (Figure 24) and will vary considerably owing to even modest changes in the CCD temperature.

A less expensive and much less complicated cooling technique makes use of **thermoelectric cooling** methods. These methods are employed in essentially all "off-the-shelf" CCD systems and allow operation at temperatures of $-20°C$ to $-50°C$ or so, simply by plugging the cooler into an electrical outlet. Peltier coolers are the best known form of thermoelectric cooling devices. CCD operation and scientific quality imaging at temperatures near $-30°C$ is possible, even at low light levels, due to advances in CCD design and manufacturing techniques and the use of multipinned phase operation.

The amount of dark current a CCD produces depends primarily on its operating temperature, but there is a secondary dependence upon the bulk properties of the silicon used in the manufacture. Even CCDs produced on the same silicon wafer can have slightly different dark current properties. Today's CCDs are made from high purity epi wafers produced with low occurrences of integrated circuit error. These factors have greatly reduced many of the sources of dark current even at warmer temperatures. As with most of the noise properties of a given CCD, custom tailoring the CCD electronics (such as the bias level and the readout rate) can produce much better or much worse overall dark current and noise performance.

Q3) Camera Field of View

What's the field of view of a 2K x 2K CCD camera on a 5-m telescope with f/16 focal ratio? The pixel size of the CCD is 20 micron. How does the field of view change if we bring it to a 10-m telescope?

Campbell Physics Q14

1.14 Question 14

What's the field of view of a 2K × 2K CCD camera on a 5 m telescope with $f/16$ focal ratio? The pixel size of the CCD is 20 μm. Now, lets bring this to a 10 m telescope with the same focal ratio. Explain how the field of view changes on the 10 m telescope (compared to that of the 5 m telescope) based on the Etendue conservation rule.

1.14.1 Short answer

The FOV of a telescope operating with a CCD is given as follows:

$$\text{FOV} = \frac{1\,\text{rad}}{\text{focal length}} \cdot (\#\,\text{pixels}) \cdot (\text{pixel size}) = \frac{206265''}{f^{\#}D} N\Delta x,$$

where we have used the fact that $1\,\text{rad} = 206265''$ and that the focal length is given by $f = f^{\#}D$, where D is the telescope diameter.
Using this to find the FOV for a 2K × 2K CCD camera on a 5 m telescope with $f/16$ focal ratio:

$$\text{FOV} = \frac{206265''}{(16)(5\,\text{m})}(2000)(20 \times 10^{-6}\,\text{m}) \sim 100''.$$

We can now use the Etendue conservation rule to find the FOV at a 10 m telescope. Etendue is a property of light which characterizes how "spread out" the light is in an area and solid angle, and the Etendue conservation rule says that this property is conserved following the equation

$$\epsilon = \text{FOV}^2 \cdot D^2 = \text{const [dimensionless]}.$$

This allows us to equate the etendue of the first telescope setup to that of the second in order to solve for the new FOV:

$$\begin{aligned}
\epsilon_{5\,\text{m}} &= \epsilon_{10\,\text{m}} \\
\text{FOV}_{5\,\text{m}}^2 \cdot D_{5\,\text{m}}^2 &= \text{FOV}_{10\,\text{m}}^2 \cdot D_{10\,\text{m}}^2 \\
\text{FOV}_{10\,\text{m}} &= \left(\frac{D_{5\,\text{m}}}{D_{10\,\text{m}}}\right) \text{FOV}_{5\,\text{m}} \\
\text{FOV}_{10\,\text{m}} &= \left(\frac{5\,\text{m}}{10\,\text{m}}\right) 100'' \\
\text{FOV}_{10\,\text{m}} &= \frac{1}{2} 100'' \\
\text{FOV}_{10\,\text{m}} &= 50''.
\end{aligned}$$

Thus, using Etendue conservation, we can see that the FOV is reduced by a factor of 2 when brought to a telescope twice the diameter.

1.14.2 Follow-up Questions

- If you wanted a smaller FOV with the same size telescope, what would you change?