

Chapter 6

Measurements by optical methods

by Urban Brändström, Carl-Fredrik Enell,
Kerstin Stebel, Uwe Raffalski and Thomas Wagner

This chapter tries to summarize detectors and instruments used for atmospheric studies by optical methods including studies of the aurora. The main focus is on instruments and techniques employed by the Swedish Institute of Space Physics in Kiruna, Sweden but an attempt is made to cover as many related fields as possible. Most of the material is compiled from other sources.

6.1 Introduction

Optical detectors and instruments is an extremely extensive field, and even after limiting the scope to atmospheric studies by optical methods, an enormous span of material remains, ranging from simple photodiodes to complex CCD detector systems.

6.1.1 A quite frequently used optical instrument

We will begin our survey of scientific optical instrumentation by investigating the oldest and most widely used optical instrument known to mankind — the human eye (figure 6.1). This device possesses a number of features that are important to the understanding of the instruments and detectors presented later. The human eye consists of an aperture (iris) and an optical system (lens, vitreous humor ...) that focuses the image on the detector (retina) (Campbell 1987). The detector itself employs “on-chip” selective imaging techniques (SIT¹) which largely reduces the amount of data sent to the neural network-based analysis computer (a.k.a. the brain). Actually the detector itself is an integral part of the cerebral cortex. The detector resolution also varies over the field of view. External optical systems (glasses, binoculars, telescopes, e.t.c.) and filters (sun-glasses) can easily be mounted to enhance the measurement system. The two-instrument design provides some redundancy and facilitates stereoscopic imaging. A precision imager positioning system is used to center the

¹The acronym SIT unfortunately has two completely different meanings: (1) Selective Imaging Techniques in the context of state-of-the-art detectors and image processing; and (2) Silicon Intensified Target vidicon, as described in section 6.6.5

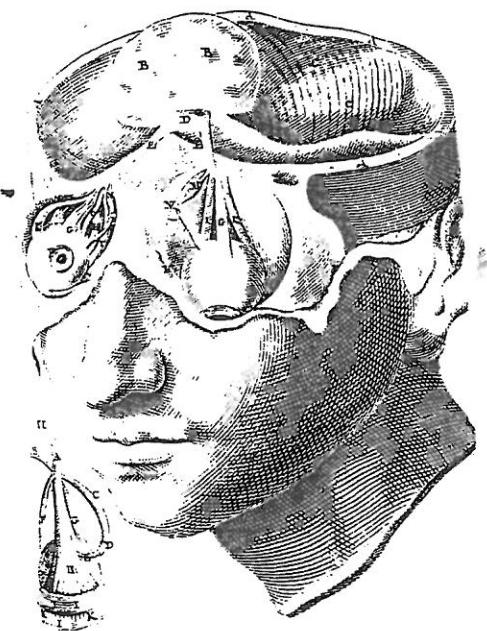


Figure 6.1: An advanced bistatic instrument for optical observations

field of view on the object being studied. The optics are protected by a fast closing mechanical lens-cover that also cleans the optics, and is sometimes used as a secondary signaling system. The system sensitivity is mainly in the visible part of the electromagnetic spectrum, and varies with wavelength. Typical sensitivity is around 1-3 kR (Rayleigh, see section 6.1.2) for 557.7 nm. Occasionally single photons might be observed by this extremely sensitive instrument. The major disadvantage of this excellent design, is that it is delivered completely without documentation. For further reading about the human eye see for example (Widén 1995).

6.1.2 Some basic concepts and units of measurement

This section describes a number of basic concepts and units of measurement that are often used in this field of science. Some of them (e.g. the Ångström) are not standard SI units, but are so frequently used that they are impossible to neglect.

The Ångström

Wavelengths are often given in Ångströms for the optical part of the electromagnetic spectra. 1 Ångström is defined by:

$$1 \text{ \AA} = 10^{-10} \text{ m} = 10 \text{ nm.} \quad (6.1)$$

The Rayleigh

The Rayleigh (R) is a unit agreed upon for photometric studies of airglow and aurora. The quantity actually measured is usually surface brightness B, which is most usefully expressed

in such units as *quanta/(m²s sterad)*. However the important quantity when results are interpreted in terms of physical processes is usually volume emission rate, in *quanta/(m³s)*. This emission rate cannot be found directly from surface brightness measurements, but the rate of emission from a cm² column along the line of sight is usually just $4\pi B$. 1 Rayleigh is defined as:

$$1R = \frac{10^6 \text{ quanta}}{\text{cm}^2 \text{s}(\text{column})} \quad (6.2)$$

For more information about definition of the Rayleigh, see the original paper defining the Rayleigh (Hunten et al. 1956).

Photon statistics

Low light level photon arrival rates are quite well represented by Poisson statistics, which show that if the mean rate of an occurrence is k per unit time, then the probability of r occurrences in any time unit interval is $P(r) = e^{-k} k^r / r!$. It is easily shown that this probability distribution has a mean value k (Eccles et al. 1983).

Quantum efficiency and sensitivity

The sensitivity of a detector may simply be defined as the ratio of the output signal to the input signal. Although useful enough in assessing the response of any one detector, the sensitivity gives no information about how efficient it is at recording photons. The quantum efficiency (QE) of a detector is defined as the number of photons recorded by the detector divided by the number of photons that would have been recorded by a perfect detector under the same conditions. This quantity is therefore a ratio and is usually expressed as a percentage. In a multi-stage detector, the quantum efficiency of each stage may be considered separately. The quantum efficiency is a useful general guide to potential sensitivity of a detector, but is most often used in comparing the performances of detectors of the same type. The concept of quantum efficiency does not allow for photons recorded and subsequently lost. It is therefore not very useful in dealing with the photographic emulsion or other detectors where it is difficult to quantify the number of photons detected.

The detective quantum efficiency (DQE) of a complete detector system takes into account any loss of detected photons and the quantum efficiency of the entire system. In a loss-free single stage system $DQE = QE$ but DQE will be less than QE for most other systems. Neither DQE nor QE gives any information about the quality of detector output, since the precision with which a signal can be recorded depends on its signal-to-noise characteristics. DQE can be expressed as in terms of signal-to-noise characteristics of the input (S/N_{in}) and output (S/N_{out}) signals as follows:

$$DQE = \frac{(S/N_{out})^2}{(S/N_{in})^2} \quad (6.3)$$

Modulation transfer function

The spatial resolving power of a detector is often specified in terms of the number of line pairs per millimeter that it can resolve. A more objective measure is the modulation transfer function. Assume an image which takes the form of one-dimensional sinusoidal modulations,

e.g. $E(x) = a + b \sin 2\pi\omega x$. The modulation of this image is defined as

$$M = \frac{(E_{max} - E_{min})}{(E_{max} + E_{min})} \quad (6.4)$$

which in the example is b/a . The output image will have a similar form but will be degraded by the detector to a lower modulation. The modulation transfer is defined as M_{out}/M_{in} , where M_{out} is evaluated in input signal units. This will be a function of spatial frequency ω , falling from 100 % at low frequencies to 0 % at high frequencies. This function of ω is called the modulation transfer function.

The resolving power is closely linked with the concept of a pixel. This can be defined as the size of the smallest element of a detector to which a definite signal level can be attributed. For detector arrays such as CCDs the meaning of a pixel is obvious. This is not the case for e.g. television tubes or photographic emulsion where the reciprocal of the spatial resolving power is often used. E.g. a resolution of 200 line pairs per millimeter implies a characteristic pixel size of $2.5\mu m$.

Dynamic range

The information storage capacity of a pixel is the number of distinguishable values it can assume. This is often measured in bits of information. The dynamic range (see section 2.1) is the ratio of the lowest detectable level to the highest level. If the dynamic range is exceeded the detector is said to be saturated. Note that saturation can occur for many reasons, ranging from overflow in the ADC converter to that the maximum number of photons that the detector can handle has been exceeded (for CCD detectors this is called the full well capacity, and denotes the maximum charge that can be accommodated in a single pixel).

6.1.3 The optical environment

The environment in which an optical instrument is mounted is often subject to a lot of disturbances. For low light level imaging, the signal is photon limited and each disturbing photon detected from other sources degrades the data quality. For ground-based measurements, the high background level caused by artificial illumination sources is becoming a more and more annoying problem. Modern civilization with its ever-increasing energy consumption pollutes the optical environment with both smog and artificial illumination sources (optical pollution), even in the most remote places, making good sites for optical observatories a very rare resource. In space, measurements of atmospheric phenomena are often complicated by the unknown effective albedo of the ground. The orbital motion of a spacecraft also causes a degradation of the resolution, but on the other hand only spacecraft can image dayside phenomena, and give a global perspective of an event. Space- and ground-based observations should therefore be regarded as complementary observation techniques.

6.2 Photographic emulsions

Although the photographic film is now almost regarded as obsolete for scientific studies, it has played a major role until quite recently and in some applications it is still superior to other techniques. The dimensional stability of a photographic plate, combined with the large field and permanence of record, allow very accurate positional measurements to be made of a large

number of objects on one plate. However, the photographic emulsion suffers from several disadvantages compared to electronic detectors. It is very inefficient. Even the best emulsion has a peak detective quantum efficiency (DQE) of only 4% as compared with the Charge Coupled Device (CCD) which may have a DQE of up to 90%. Consequently photographic observations require considerable longer exposure times than those required for electronic systems. The response of photographic emulsion to light is not linear, although this can be overcome to a considerable extent by careful calibration of each plate during exposure.

The degree of blackening of a photographic emulsion due to exposure to light and subsequent chemical processing is directly related to the intensity of the exposing light source. The characteristic curve of the emulsion-developer combination is typically of the form shown in figure 6.2 where D is diffuse density and E exposure.

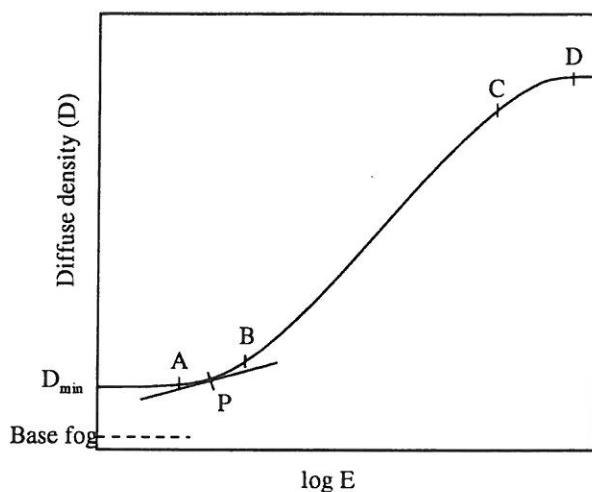


Figure 6.2: The form of the characteristic curve for a photographic emulsion. It shows the exposure, E , versus the diffusive density, D . The contrast or gamma of the emulsion is the slope of the straight-line portion of the curve. The local gamma at any point P is obtained by the slope of the tangent to the curve at that point. (Adapted from Eccles et al. (1983).)

More information about the photographic emulsion as a detector can be found in Eccles et al. (1983). This book also includes an extensive set of further references.

6.3 The image intensifier

Although not a detector itself, the image intensifier is a very important part of many detector systems and instruments. An image intensifier can be used together with any detector (including photographic emulsions, and the human eye).

The most important part of the image intensifier is a chemically-coated glass plate called photo-cathode (see section 6.5.1), which emits electrons when exposed to incident photons. The basic process of image intensification, or optical amplification as it is sometimes called, starts with the production of such electrons from light incident on a photocathode and proceeds with acceleration of these electrons to a high energy. The electrons finally impinge on a target, generally a phosphor screen, which emits photons. Figure 6.3 shows a single stage electrostatically-focused image intensifier.

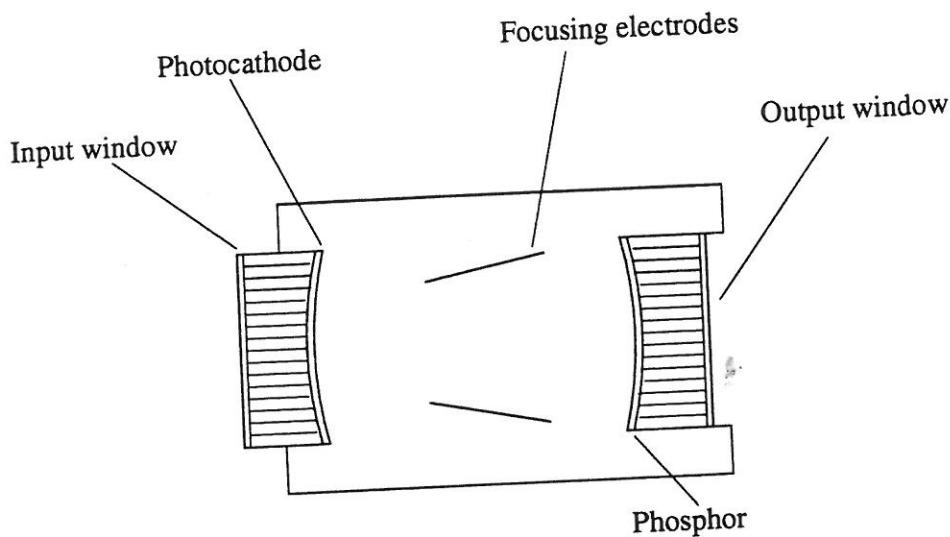


Figure 6.3: A single-stage electro-statically focused image intensifier. (Adapted from Eccles et al. (1983).)

The image intensifier usually consists of an evacuated cylinder with a photocathode at one end of the cylinder, some form of target at the opposite end, and a potential difference of typically 20 kV maintained between the two ends. The incident photons are focused onto a glass screen and transferred to the photocathode via a fiber optic channel plate. The electrostatic focusing electrodes are such that both the object and image planes are curved and the image reduced in size. The image is formed on a curved phosphor screen and transferred to a flat face plate in contact with the detector (or a secondary optical system). Typical gain of a single image tube is around 40, and it is quite common to find three similar tubes cascaded, giving a gain of over 6000.

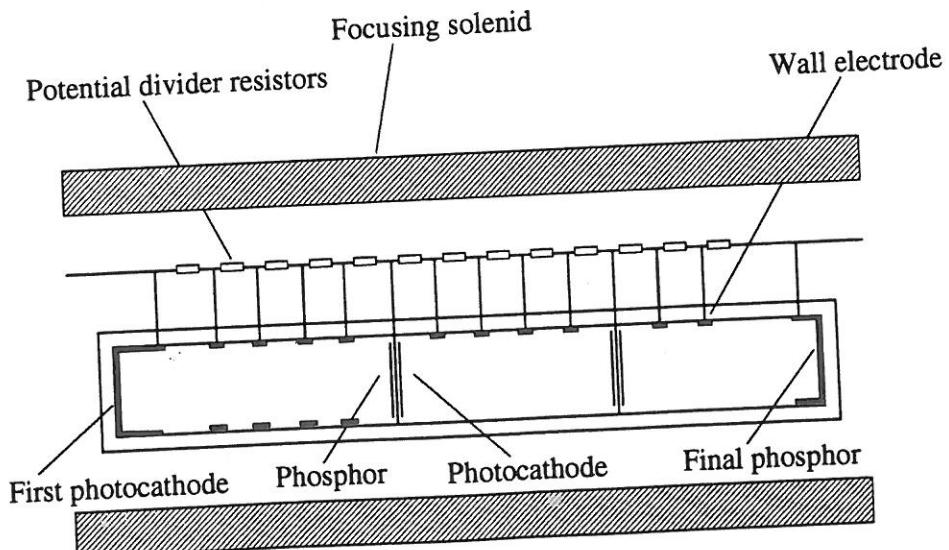


Figure 6.4: A three-stage magnetically focused image intensifier. (Adapted from Eccles et al. (1983).)

Most of the work on image intensifier development has been for military (night-vision, etc) applications, where strength and simplicity of operation have taken precedence over image quality, and therefore favored tubes with electrostatic focusing. Figure 6.4 shows a magnetically focused image intensifier with three cascaded intensifiers. In this case the object and image fields are flat, and there is no need for optical image transfer. The solenoid is wound so as to produce as nearly uniform an axial magnetic field as possible. Light focused on the first photocathode generates photoelectrons which are focused onto the first phosphor; this stimulates the second photocathode and so on to the flat glass output screen.

The majority of photocathodes operate most efficiently towards the blue end of the optical spectrum and have quantum efficiencies of 10 to 20 %. It should be stressed that the inherent high quantum efficiency of some CCD devices is lost if combined with an image intensifier. Other problems in using image intensifiers for quantitative measurements are small or blemished fields, geometrical distortion, and ultimately non-linear responses. Another major disadvantage of image intensifiers is that they are easily damaged by light, and therefore unsuited for long-term unmanned operation (e.g. all-sky cameras). Although the image intensifier is still used in many measurement systems today, they are avoided in the design of modern high quality imaging systems.

More information about image intensifiers and their applications can be found in Eccles et al. (1983) and Csorba (1990) with additional references found therein.

6.4 Photoconductive detectors

The intrinsic photoconductive process is the absorption of a sufficiently energetic photon by a semiconducting crystal lattice such that one of the covalent bonds are broken and an electron-hole pair is created. To be specific, for an electron-hole pair to be created the photon wavelength must be shorter than a threshold wavelength λ_t , given by:

$$\lambda_t = \frac{hc}{E_G}, \quad (6.5)$$

where E_G is the energy gap, h is Planck's constant, and c is the speed of light. Table 6.1 lists the long wavelength threshold and energy gap at 300 K for several semiconductors with important applications in optical astronomy and space physics. The creation of an electron-hole pair temporarily increases the local conductivity. To detect this externally a uniform electric field is established in the semiconductor crystal by connecting a voltage in series with a load resistor R_L between the ends of the crystal as shown in figure 6.5. When a charge pair is created, the more mobile electrons diffuse away from the surface faster than the holes and so create a carrier population gradient with its associated electric field perpendicular to the longitudinal externally applied field. If the two fields become comparable the diffusion of electrons will be inhibited and a non-linearity will result.

The term intrinsic photoconductor refers to those photoconductors just described including doped semiconductors, while the term extrinsic photoconductors refers to semiconductors where impurity atoms are introduced into the crystal lattice such that additional energy levels are created within the forbidden energy band. In this way, very long wavelength photons can be detected, provided that the crystal is cooled sufficiently so as to eliminate thermally-induced transitions. Table 6.2 lists the ionization energy, E_i , and corresponding long wavelength threshold, λ_t , for some typical semiconductor-impurity combinations.

Table 6.1: Long wavelength threshold and energy gap at room temperature for electron-hole pair production in various semiconductors. (Adapted from Eccles et al. (1983).)

Semiconductor	E_G (eV)	λ_t (μm)
Si	1.12	1.10
Ge	0.68	1.82
GaAs	1.35	0.92
GaSb	0.78	1.59
InAs	0.33	3.8
InP	1.25	1.00
InSb	0.18	6.9
PbO	1.90	0.65

Table 6.2: Ionization energy and corresponding long wavelength threshold for some typical semiconductor-impurity combinations. (Adapted from Eccles et al. (1983).)

Semiconductor-impurity	E_i (eV)	λ_t (μm)
Ge-In	0.011	113
Ge-Ga	0.011	113
Ge-As	0.013	95
Ge-P	0.012	103
Si-In	0.16	8
Si-Ga	0.068	18
Si-As	0.053	23
Si-P	0.045	28

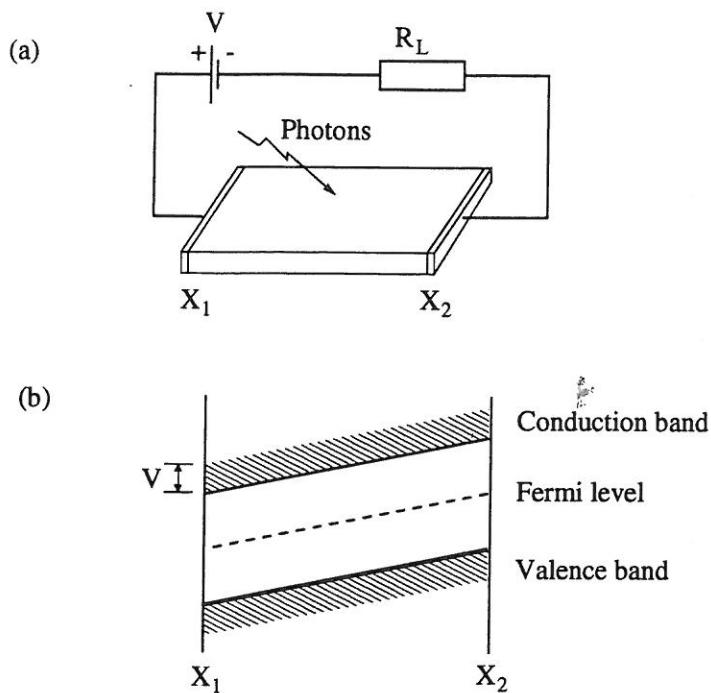


Figure 6.5: (a) Circuit diagram showing an intrinsic photoconductor across a maintained voltage in series with a load resistor. (b) Corresponding energy band diagram showing all energy levels tilted by V . (Adapted from Eccles et al. (1983).)

6.4.1 Photo diodes

The most commonly encountered junction diode is the simple p-n diode which is manufactured by differentially doping a homogeneous semiconductor crystal. There is a very narrow region in which the doping changes from p-type to n-type known as the p-n junction. This region determines the essential behavior of the p-n diode. Applying a potential difference where the p-side is positive with respect to the n-side forward biases the diode with a voltage drop (0.6V for silicon) across the junction. The current is limited only by the external circuit. If the potential is reversed, the diode is said to be reverse biased. It can be seen that at a sufficiently high reverse bias, valence band electrons on the p-side will have the same energy as unfilled energy states in the conduction band on the n-side. Under this condition tunneling can occur directly across the junction. This phenomenon is called Zener breakdown (cf. Zener diodes in voltage regulators) and limits the maximum reverse voltage to a few tens of volts.

The junction can be thought of as a capacitance with the rest of the diode as a series resistance. Forward biasing decreases the width of the depletion layer thus increasing capacitance while reverse biasing does the contrary. For this reason most diodes intended for the detection of rapidly modulated light are reverse biased.

Photons of sufficient energy incident on the diode will create electron-hole pairs as before. In the absence of any external load or bias across the diode there will be a buildup of electrons on the n-side and holes on the p-side such that the Fermi level will move up to a higher energy on the n-side relative to the p-side. This mode of operation is called photovoltaic; the potential across the ends of the diode is proportional to the number of incident photons, provided the charge collected is not so high as to neutralize the depletion layer electric field. If a load

resistor is connected a current will flow until the Fermi level is the same on both sides of the junction. However, the normal mode of operation is with a reverse bias in series with a load resistor connected across the diode. In this photoconductive mode of operation, charge pairs are removed from the region of the junction as soon as they are formed, which leads to a highly linear response. Typical diodes are linear over eight decades of incident photon flux. The presence of the reverse current gives rise to a leakage or dark current which is associated with several sources of noise.

The sensitivity of a photo diode is determined by the quantum efficiency and the sum of various noise sources. In general for any semiconductor there are four noise sources to be considered: shot noise, generation-recombination noise, Johnsson noise and flicker noise. Shot noise is caused by quantum fluctuations in the current flowing through the semiconductor device. Generation-recombination (G-R) noise is caused by thermally generated fluctuations in the rates of carrier generation and recombination which in turn causes a fluctuation in the semiconductor resistance and hence a fluctuation in the voltage across the semiconductor when a bias current flows. Associated with any resistance is Johnsson noise which is due to Brownian motion of the charge carriers and is solely a function of temperature and measurement bandwidth. The equivalent Johnsson noise current in a resistance R is given by:

$$i_N = \sqrt{\frac{4kTB}{R}} \quad (6.6)$$

where k is Boltzmann's constant, T is the absolute temperature and B is the measurement bandwidth. Flicker noise is associated with all electronic devices and has the general form:

$$i_N \propto I_0 B^{\frac{1}{2}} f^{-1} \quad (6.7)$$

so that it is often referred to as $1/f$ noise.

Figure 6.6 shows a typical photodiode detector circuit and figure 6.7 shows the spectral response for a typical photodiode. The quantum efficiency approaches 80 % at $0.85\mu m$.

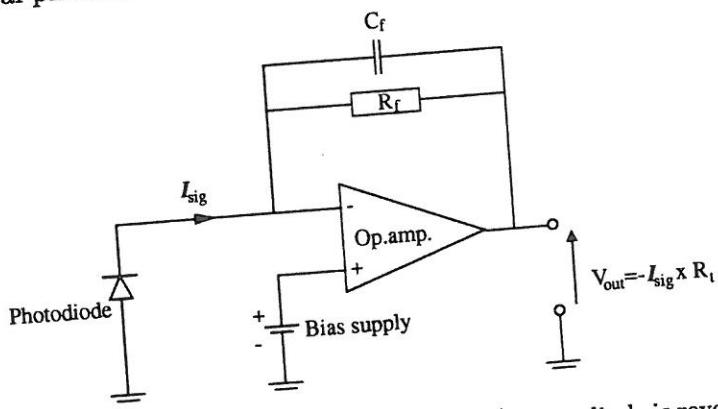


Figure 6.6: A simple photodiode detector circuit. The photodiode is reverse biased.
(Adapted from Eccles et al. (1983).)

Avalanche photodiodes

The avalanche photodiode (APD) is a derivative of the silicon pin photodiode described above, where a high internal gain is introduced by applying a very high voltage to a p-n photodiode.

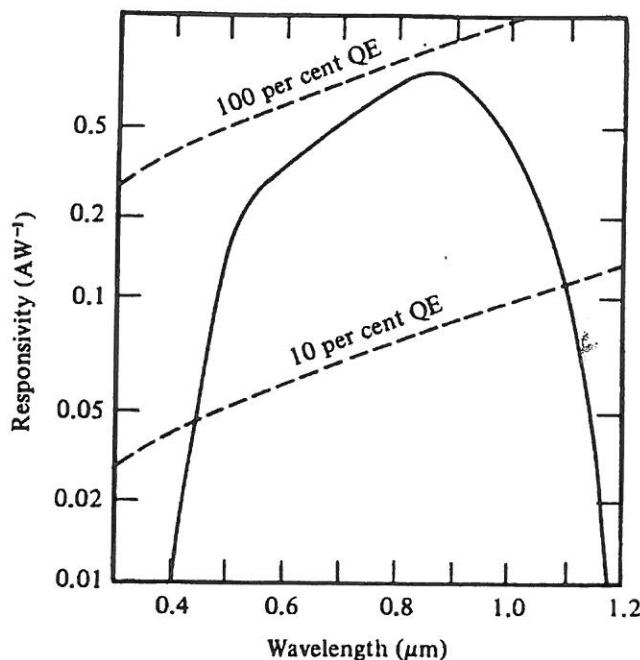


Figure 6.7: Spectral response for a typical silicon p-i-n diode. The quantum efficiency approaches 80 % at $0.85\mu m$. (Adapted from Eccles et al. (1983).)

The corresponding internal electric field accelerates the carriers, mainly the electrons, which in turn create secondary electrons via collisional ionization — an avalanche of electrons. Semiconductor devices tend to be damaged by high voltages so constructing an APD is not an easy task. Figure 6.8 shows two designs of APDs.

The Vacuum Avalanche Photodiode (VAPD or iAPD) consists of a photocathode mounted in front of a large-area APD (figure 6.9). This device combines high gain $> 10^6$, low noise and excellent temporal resolution, allowing it to perform single photon discrimination better than any other type of photodetector.

The APD and VAPD are mainly a solid state replacement for the photo multiplier tube (described below) in photometers and other similar instruments.

References for further reading: general introduction Eccles et al. (1983) and avalanche photo diodes Clark (1993) with references therein.

6.5 Photoemissive detectors

The previous section has been concerned with the fundamental detection process in which sufficiently energetic photons create electron-hole pairs in the detector material. At very low light levels the photon-induced electric current in the detector will be lower than the noise of any electronic amplifier connected to the detector. For faint object work the detector must employ some gain mechanism.

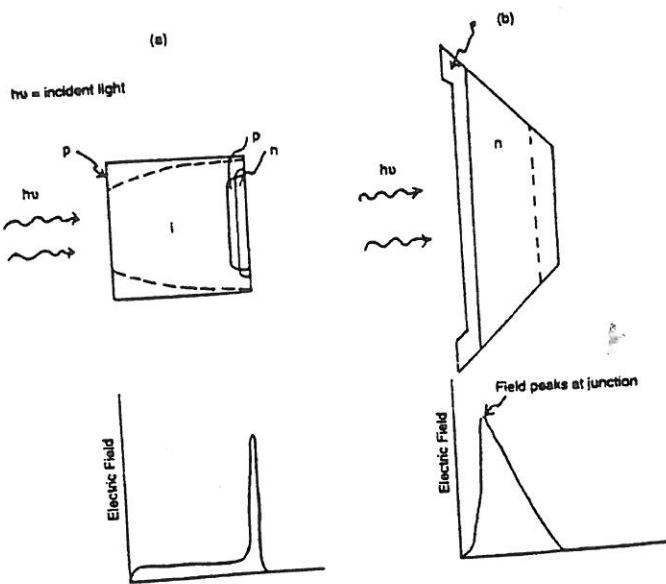


Figure 6.8: The two leading APD designs: (a) reach through and (b) beveled-edge or large-area. In both cases the extent of the depletion region (region of internal field) is indicated by a dotted line.

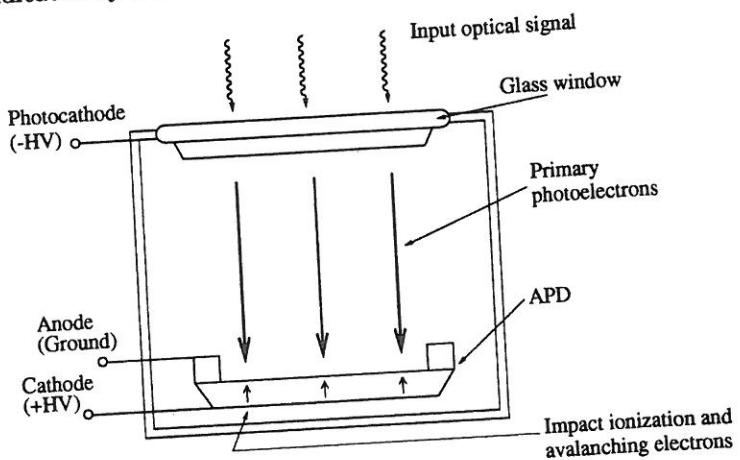


Figure 6.9: Schematic of a VAPD, a successful hybrid.

6.5.1 The photocathode

The photocathode is a particularly important part of any detector system, since it establishes the maximum quantum efficiency and spectral bandwidth of the overall system.

Photoemission can be thought of as comprising of three consecutive steps, the absorption of a photon producing an electron-hole pair in the solid material, the motion of the electron towards the material-vacuum interface, the overcoming of the potential barrier at the interface and the release of the electron into the vacuum. Metals make poor photocathodes because most incoming photons are reflected at the surface rather than absorbed. On the other hand semiconductors can be very efficient in absorbing photons, have very few free electrons and can be made with very small surface potential barriers.

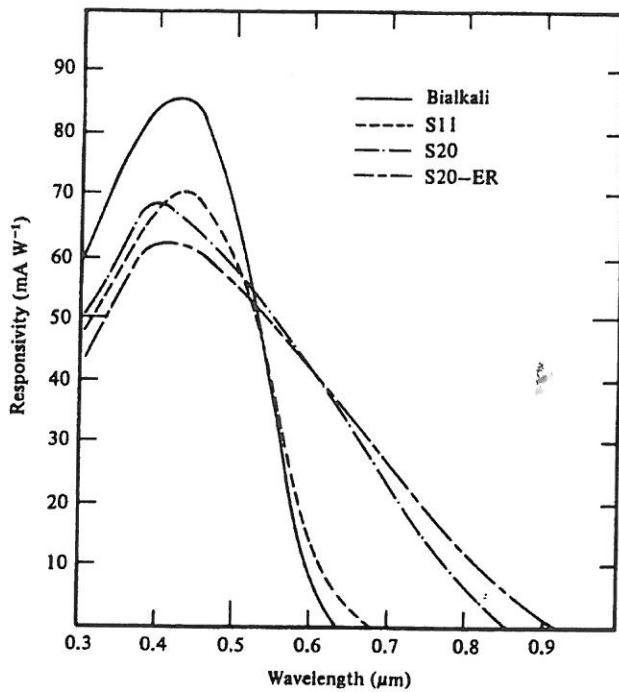


Figure 6.10: Spectral response of some important photocathodes. (Adapted from Eccles et al. (1983).)

Incident photon energies must be higher in photoemissive detectors than in photoconductive detectors, and thus photoemissive detectors have a generally poorer red response. Furthermore, photoemissive quantum efficiencies are typically only around 20% as compared to the over 80% quantum efficiency of photoconductive detectors. Figure 6.10 shows spectral response for a number of photocathodes.

6.5.2 Photomultiplier tubes

The simplest photoemissive detector consists of a photocathode and an anode enclosed in evacuated envelope (an electronic valve diode so to say). Photons incident on the photocathode emit electrons (photoelectrons) which are collected by the anode and amplified by external electronic circuitry. This device is suitable only for detecting high light levels as there is no internal gain to overcome the amplifier read-out noise. For this purpose the photomultiplier tube (PMT) has been developed. A typical PMT device used for low light level photon counting is shown in figure 6.11.

Photons incident on the semitransparent photocathode cause photoelectrons to be released from the rear of the photocathode and to be directed by focusing electrodes to the first dynode of an electron multiplier. The first dynode is at a potential of a few hundred volts positive with respect to the photocathode so that photoelectrons strike the dynode with sufficient energy to release several secondary electrons per primary electron. This process is repeated at successive dynodes so that the charge pulse leaving the last dynode will contain at least one million electrons for each primary photoelectron. An anode close to the last dynode collects the charge pulse and passes it to the external circuitry. Figure 6.12 shows a number of dynode

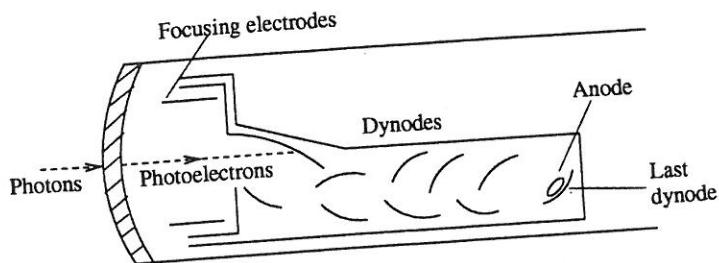


Figure 6.11: A typical photomultiplier tube showing the photocathode, dynodes and anode. This particular electrode configuration is known as linear-focused. (Adapted from Eccles et al. (1983).)

configurations. The PMT is mostly used in photometers, and is now being replaced by APDs and VAPDs and other semiconductor devices.

6.5.3 Channel electron multipliers

The concept of a multi-dynode electron multiplier has been extended to its logical extreme of a continuous distributed dynode in the channel electron multiplier (as described more detailed in section 3.1.1). This device, shown in figure 6.13, consists of a hollow tube made out of semiconducting glass. The inner surface of the tube, the channel, acts as both the photocathode for incoming short wavelength photons, and the continuous distributed dynode. A positive voltage of several kilovolts applied between the ends of the tube establishes an accelerating electric field in the channel. Photons or electrons which strike the channel wall with sufficient energy release electrons which are accelerated along the channel axis while drifting across to strike the wall and release further secondary electrons. The work function of the semiconducting glass is so high that a separate photocathode must be employed for visible light wavelengths.

Micro channel plates

Since the parameter which most strongly defines the channel performance is the ratio of length to diameter of the channel, fiber optic technology has been used to reduce the channel diameter. Large numbers of these micro channels can then be bundled together to form a micro channel plate (MCP), as described in more detail in section 3.1.2. The MCP could be used either as an imaging detector or, if all the anodes are connected together, a large area detector. It is also possible to use a phosphor screen at the output so that the whole MCP acts as an image intensifier. The MCP makes a very compact two-dimensional photon counting detector, particularly suited for space-born applications.

More information about photoemissive detectors and their applications can be found in (Eccles et al. 1983, Csorba 1990) with additional references found therein.

6.6 Television-type detectors

The development of vacuum tube imaging detectors has followed closely the demands for ever better and better television camera tubes by the broadcasting community. For this reason

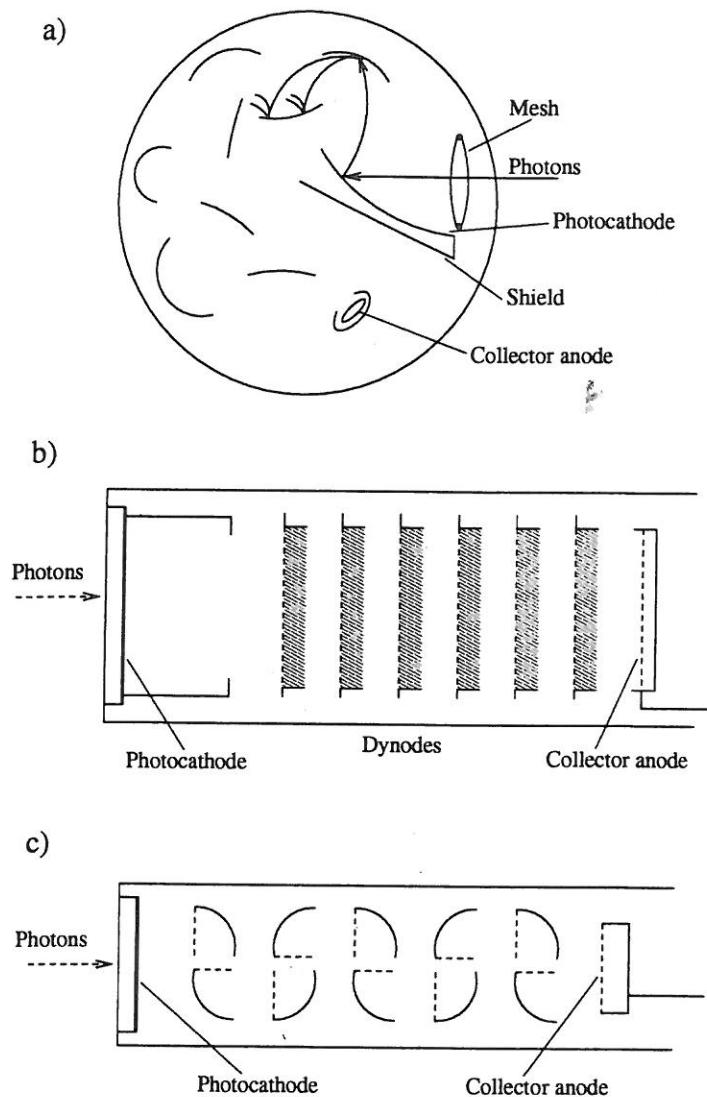


Figure 6.12: Some PMT designs: (a) The squirrel cage PMT which is very compact, (b) the venetian-blind MT and (c) the box-and-grid PMT. (Adapted from Eccles et al. (1983).)

vacuum tube imaging detectors are often referred to as television-type imaging detectors. The common characteristic of all television-type camera tubes is the use of an electron beam to read out the video information which is stored as a charge pattern on an element called the target. It should be noted that this section is cut short because vacuum tube imagers are now almost completely obsoleted by CCD devices. However, some instruments based on camera tubes still remain in operation.

6.6.1 Electron beam target read-out

Figure 6.14 illustrates the basic operation of target storage and read-out for a typical tube. Photons incident on the tube faceplate are carried by a fiber optic bundle to the curved photocathode where they release photoelectrons with a typical quantum efficiency of 20 %.

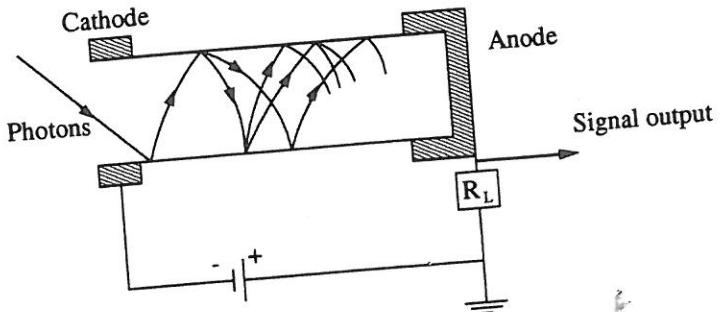


Figure 6.13: Channel electron multiplier. Photons strike the open end of the tube and release electrons which are accelerated towards the anode while drifting across the tube to strike the walls releasing secondary electrons. (Adapted from Eccles et al. (1983).)

The photoelectrons are accelerated through several kilovolts and strike the target. The target is a semiconductor in which the photoelectrons can produce many electron-hole pairs by the same mechanism already described for the dynodes used in PMTs. In the case of most targets, the photoelectrons are of higher energy and thus produce the electron-hole pairs deep within the target. Secondary electron emission is therefore negligible. Prior to the illumination of the photocathode, an electron beam is scanned over the rear of the target. The electron beam is virtually a perfect conductor, so that the rear of the target assumes the potential of the source of the beam, that is the cathode, which is normally at earth potential. The front of the target is held at a fixed potential of a few tens of volts. The target is usually made of a resistive material and can be thought of as a leaky capacitor charged up to the target bias voltage. In the absence of the electron read beam electron-hole pairs produced in the target will locally discharge this capacitor, raising the potential of the rear of the target towards the fixed-bias voltage. Those areas of the target receiving the greatest photoelectron flux will have the most rear-side potential. When the rear of the target is again scanned with the electron read beam, the most positive or "brightest" areas of the target will require the highest number of electrons to restore the potential to earth. This recharging electron current flowing to the target is then the video output signal. The output signal current is often 1nA or less and requires amplification at the camera before being sent elsewhere.

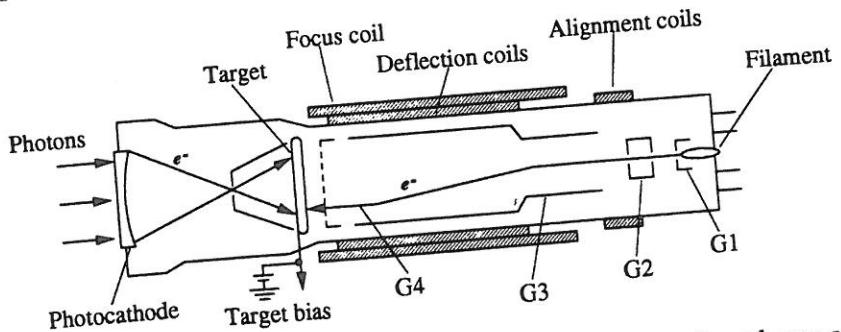


Figure 6.14: Basic operation of target storage and read-out in a direct beam camera tube. G1 is the control grid, G2 is a cylindrical lens, G3 is an accelerating wall electrode and G4 is the decelerating mesh. (Adapted from Eccles et al. (1983).)

While the quantum efficiency is defined by the photocathode, linearity and image latency, or lag, are defined by the target. Figure 6.14 also shows the typical arrangement of electrodes and coils used to generate focus and deflect the electron read beam.

6.6.2 Return-beam tubes

Figure 6.15 illustrates the image orthicon camera tube which was developed in the 1940s. The scanning beam deposits electrons on the target until the positive charge is neutralized and is then reflected back towards the cathode where it strikes the first dynode of an electron multiplier which has sufficient gain to remove the need for a very sensitive amplifier. The video information is then the difference in intensity between the direct beam and the return-beam. This tube is now completely obsolete.

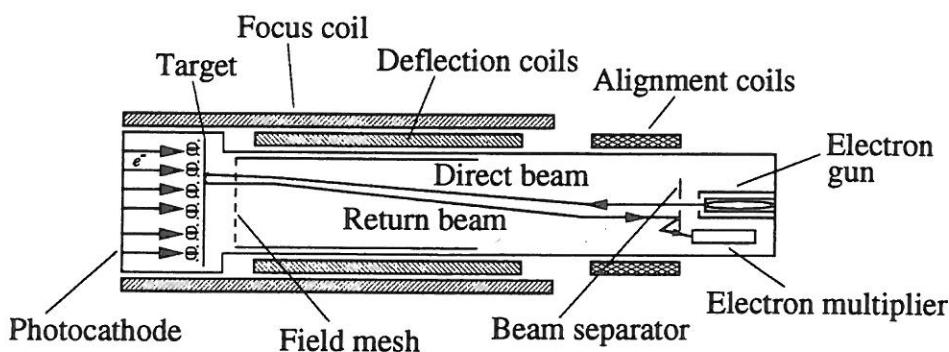


Figure 6.15: Principal components of the image orthicon return-beam tube.
(Adapted from Eccles et al. (1983).)

6.6.3 Secondary electron tubes

This term has generally come to mean a particular type of direct-beam tube having a target composed of low density potassium chloride. The advantage of this design is low dark current (up to 100 times less than any other target).

6.6.4 Vidicon

The term vidicon generally refers to a class of camera tubes in which a photoconductive target serves as both imaging medium and charge storage medium, figure 6.16. The use of one target for two separate functions leads to considerable disadvantages at low light-level imaging: severe image lag at low light levels and a gamma of less than unity.

6.6.5 Silicon target vidicon and SIT

The special feature of this tube is the target, shown in figure 6.17a. The target consists of a slice of n-type silicon into which p-type islands are diffused at $20\mu m$ intervals. This yields an improved response to red light but increases ten-fold the dark current.

The inclusion of an image intensifier section in front of the target has greatly extended the use of the silicon vidicon by permitting the detection of single photons. This combination is known as Silicon intensified target vidicon (SIT) and is shown in figure 6.17b.

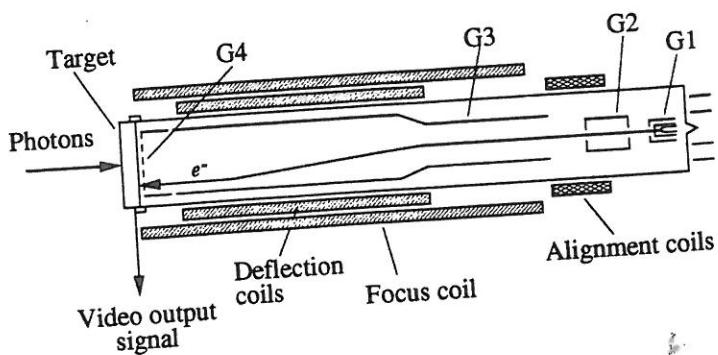


Figure 6.16: Essential components of the direct-beam vidicon camera tube.
(Adapted from Eccles et al. (1983).)

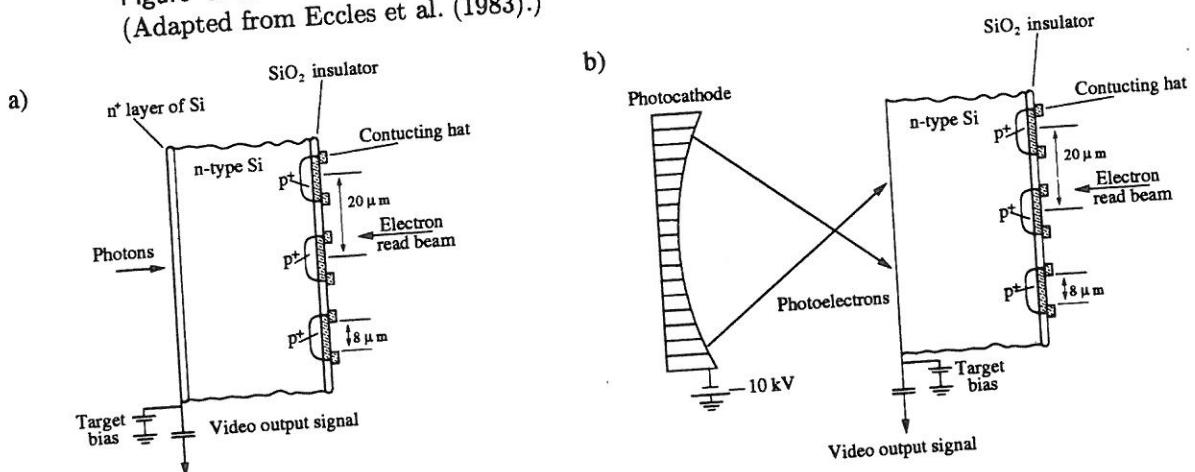


Figure 6.17: (a) The silicon diode array target. (b) The basic arrangement of the photocathode and silicon target in the SIT camera tube. (Adapted from Eccles et al. (1983).)

6.6.6 Intensified camera tubes

Intensification means here the addition of one or more stages of optical amplification before a particular camera tube so as to increase its sensitivity greatly by reducing the combined effects of dark current and read-out noise. Electrostatically focused image intensifiers are both used, coupled to the following camera tube with lenses and fiber optics. Popular examples of the latter are the fully encapsulated electrostatically focused commercial ISIT (Intensifier Silicon-Intensifier Target tube) and ISEC (Intensified Secondary Electron Conduction) tubes. A typical ISIT tube is illustrated in figure 6.18. ISIT tubes were often used in high time resolution auroral imagers from the mid 1980s, but they are now superseded by intensified CCD devices. Figure 6.19 compares spectral sensitivity of the tubes described in this section.

6.7 Solid state imaging detectors

Since the mid 1960s there has been an attractive alternative to the electron beam detectors just described, namely the Charge Coupled Device (CCD) and the closely related Charge Injection Device (CID). These are solid state imaging detectors and the general structure of such detectors is outlined in section 3.2. At the time being (1999), the devices have al-

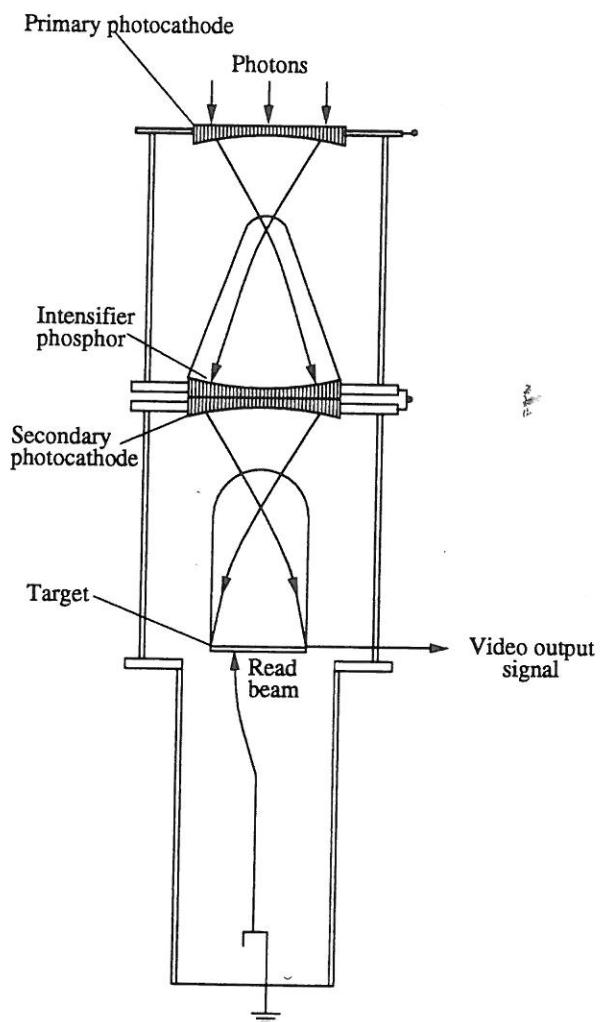


Figure 6.18: The basic arrangement of electrodes in an ISIT or ISEC camera tube.
(Adapted from Eccles et al. (1983).)

most completely replaced the electron beam detectors as well as the photographic emulsion in most scientific applications. These solid state detectors offer the obvious advantages of greatly reduced size, weight and power consumption together with scientifically more important advantages of high quantum efficiency and low read-out noise. The CCDs are also very difficult to destroy by overexposure.

6.8 The charge coupled device (CCD)

The CCD is by far the most important of these devices and will be described in considerable detail before a brief look is taken at the other imaging detectors. The discussion will be somewhat biased towards the state of the art CCDs used in the ALIS system (described in section 6.11.4).

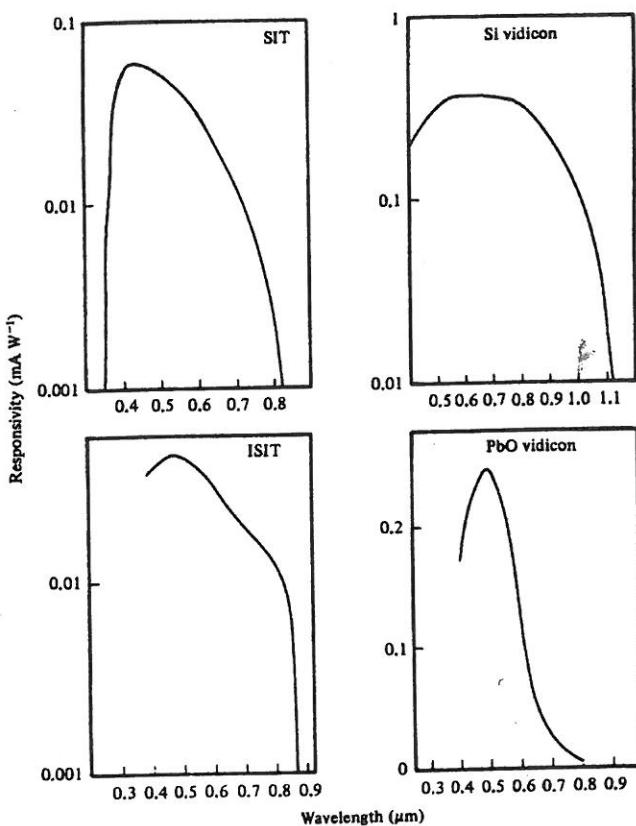


Figure 6.19: Spectral response curves for four important camera tubes. (Adapted from Eccles et al. (1983).)

6.8.1 Charge storage in MOS capacitors

The principal component of a CCD is a charge storage capacitor which is very similar to that found in a metal oxide semiconductor (MOS) transistor. The basic construction of the MOS capacitor is shown in figure 6.20. A silicon dioxide (SiO_2) insulating layer is grown onto a p-type doped silicon substrate. The insulating layer is etched away to a depth of about $0.1\mu m$ and a metal electrode called a gate is deposited. If the gate is biased positively with respect to the substrate, the majority carriers, or holes, are repelled from the $Si - SiO_2$ junction and a depletion layer forms. As the gate voltage V_G is increased the depletion layer moves deeper into the substrate. Once V_G exceeds a couple of volts, the $Si - SiO_2$ junction becomes sufficiently positive with respect to the bulk substrate that any free electrons present are attracted to the junction and form an inversion layer. In an MOS transistor, this inversion layer is the channel through which electrons pass from the source to the drain. In an uncooled MOS capacitor, thermally produced electrons will rapidly flow into the junction. However in a cooled device this inversion layer will not form, unless free electrons are generated by photon bombardment.

The depletion layer can also be thought of as a reservoir or potential well into which charge can be deposited and later retrieved. It must be stressed, however, that although the potential well concept is very useful, the electrons are in reality stored from the $Si - SiO_2$ interface downwards into the substrate and not from the bottom of a well upwards.

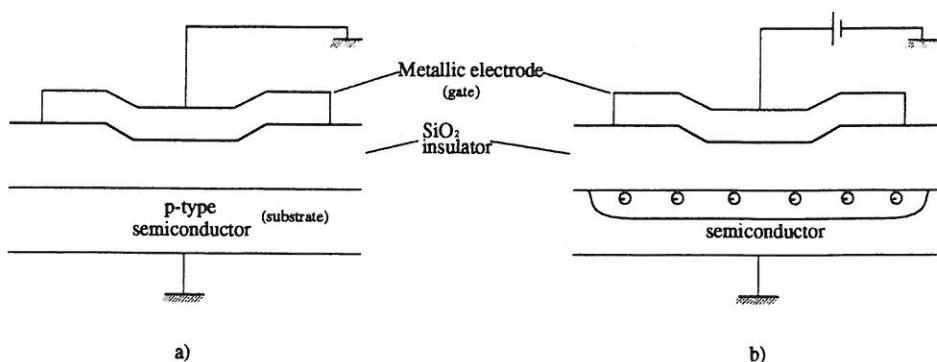


Figure 6.20: (a) Basic construction of a metal oxide semiconductor capacitor. (b) Formation of depletion layer under reverse biasing conditions. (Adapted from Eccles et al. (1983).)

6.8.2 Basic CCD operation

Although the concept of charge coupling is well defined, it has been implemented by semiconductor manufacturers in a variety of ways, and a thorough description of each of them falls far beyond the scope of this compendium. We will only briefly discuss the four most common architectures used with CCD imagers. The first is a simple linear shift register. An image is recorded by slowly scanning the scene vertically past the imager while rapidly reading out the shift register in the horizontal direction. The second architecture uses an area array as a full-frame imager. In this case the scene is imaged onto the entire device and integrated, and then with the shutter closed the entire array of pixels is read out. Third, to achieve the relatively high frame rates needed for commercial broadcast systems and to most efficiently use the available frame time, several manufacturers have used a frame-transfer structure. In this architecture the scene is imaged onto half of the array for an integration time (typically 1/30s). It is then rapidly transferred to the optically masked lower half of the array for read-out at a conventional TV data rate while the next field is being integrated in the imaging half of the array. Fourth, to employ "electronic shuttering" and to eliminate image smearing common to the frame-transfer CCD, an interline transfer system is used. This structure can be visualized as consisting of a parallel array of line sensors separated by a transfer electrode from opaque CCD read-out registers, all of which lead in parallel into a single output register. When the transfer gate is opened, the entire image from the line arrays shifts into read-out registers. After the gate is closed, a new image is integrated while the previous one is read out. For the interline transfer CCD the integrating duty cycle is essentially 100%.

Figure 6.21 presents two specific structures that represent well the diversity of approaches. In the three-phase CCD (figure 6.21a shows the cross-section) each pixel consists of three overlying gates that induce a potential profile to gather and confine charge. If these gate potentials are clocked in a coordinated fashion, the charge units can be made to physically travel through the silicon. The oxide and overlapping gates are thick enough to absorb short wavelength photons, so the silicon substrate is thinned to about 10 μm and illuminated from the back. Figure 6.21b shows the cross-section of a virtual phase CCD. In this device a four-step potential profile within each pixel is achieved with ion implantation, and a single overlying gate clocks two of these potential steps to effect charge transfer. Since there is only one gate layer that overlies half of each pixel, it is possible to achieve significant short

wavelength-response down to about 1800 Å with frontside illumination. Other CCD structures include two-phase and three-phase designs.

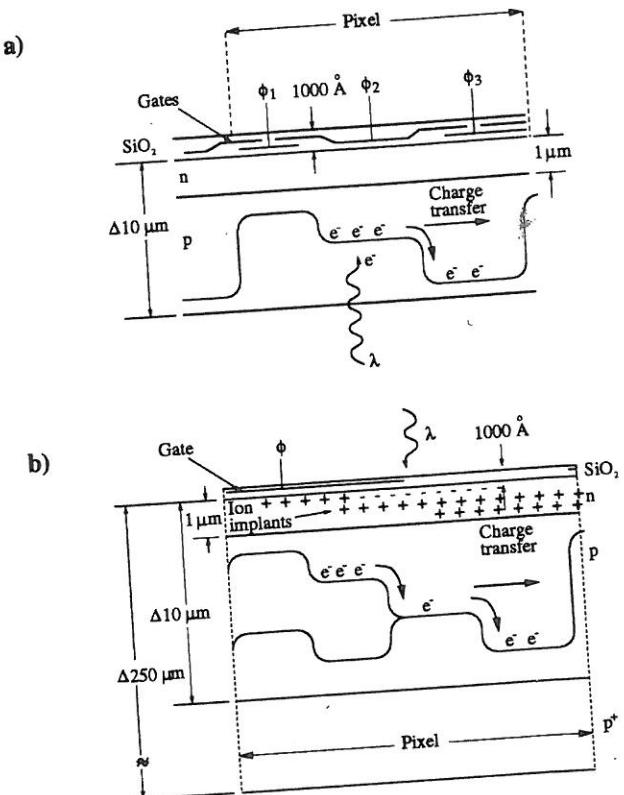


Figure 6.21: Cross-sections representing two different CCD technologies (viewed perpendicular to the channel stops): (a) three-phase and (b) virtual phase.

6.8.3 The CCD as a scientific imaging detector

Traditionally CCD performance has been measured against broadcast industry standards, which fall considerably short of requirements for many scientific applications. To characterize the capabilities of the CCD for scientific applications numerous test tools and techniques have been developed (Janesick et al. 1987). In the following sections CCD operation has been divided into the following four largely independent functions:

1. Quantum efficiency (QE)—the absorption of a certain fraction of incident photons (or particles) and separation of electron-hole ($e-h$) pairs by the absorbed energy.
2. Charge-collection efficiency (CCE)—the movement of electrons into potential wells that correspond to individual picture elements (pixels).
3. Charge-transfer efficiency (CTE)—the sequential transfer of each charge packet through the pixel array and serial register to the read-out node.
4. Signal measurement—the precise measurement of the charge collected within each pixel, limited by the read-out noise floor of the CCD.

A schematic representation of the overall transfer function of a typical CCD camera is shown in figure 6.22. The camera can be described in terms of five transfer functions, three related to the CCD and two related to the external CCD signal processing circuitry. The input to the camera is given in units of incident photons P_I and the final output signal $S(DN)$ of the camera is achieved by encoding each pixel's signal into a digital number (DN), typically using 12 to 16 bits.

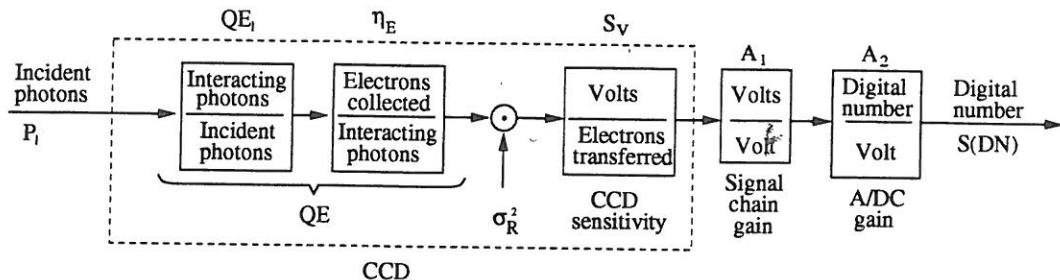


Figure 6.22: Block diagram of a CCD camera, showing individual transfer functions.

6.8.4 Quantum efficiency

The photosensitive volume of most CCDs is a layer of high-purity silicon (often epitaxial) bonded on one side by an oxide and gate structure and on the other side by a thick, low-resistivity substrate on which the device is fabricated. For photons with long absorption length (i.e. wavelengths less than 10\AA or greater than 6000\AA), the QE depends largely of the thickness of the photosensitive volume. Intermediate wavelengths have relatively short absorption lengths in silicon and silicon dioxide, and throughout this spectral region the QE depends largely on the transparency, reflectance and surface conditions of the layers that overlie the photosensitive volume. As one might expect, the QE performance of the CCD varies considerably, depending on the CCD technology employed.

Backside illumination

To obtain optimum QE performance and to exploit the advantages of multigate technology, the substrate of the CCD can be thinned to allow direct illumination of the epilayer. Following thinning the backside surface must be treated; otherwise the CCD will exhibit two undesired properties, namely instability in QE and loss of signal charge by recombination in a backside "dead" layer. The CCD detector used in the ALIS system is a thinned device with an extremely high quantum efficiency as shown in figure 6.23.

Frontside illumination

Frontside-illuminated multigate CCDs generally yield very poor QE response in the blue range due to the absorbing gate layers above the CCD potential wells (e.g. typically 5% at 4000\AA). By using implanted-gate technology frontside absorbing structure can be minimized obtaining good QE down to about 1800\AA , where absorption of SiO_2 becomes severe. The thick front-side illuminated CCD also shows instabilities in QE, specifically for wavelengths lower than 6500\AA .

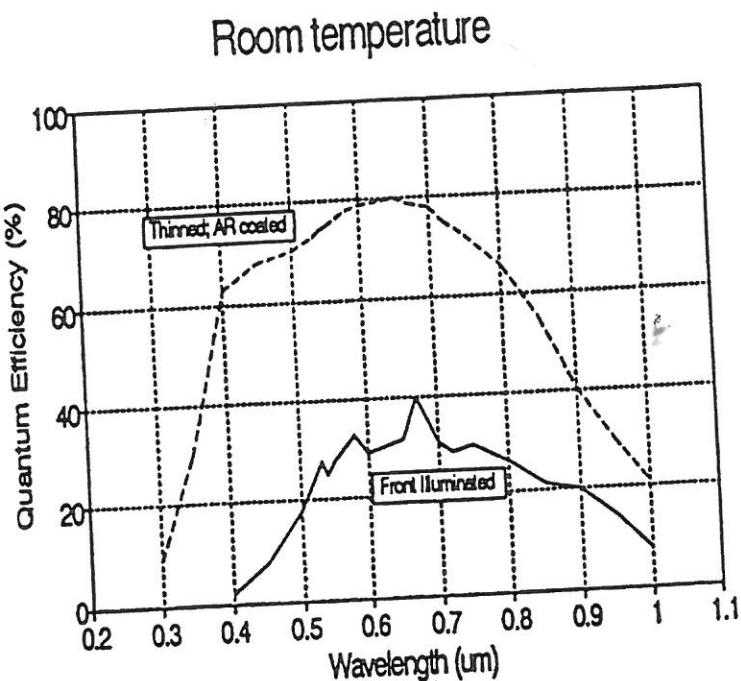


Figure 6.23: Typical quantum efficiency for the thinned anti-reflection coated Tk1024AB device used in the ALIS cameras (top curve).

6.8.5 Charge collection efficiency CCE

The ability of a sensor to faithfully and accurately record and reproduce the spatial information in a scene is an important measure. For the CCD this means that all of the charge generated by scene photons incident on a given pixel should be collected within that pixel. The most serious contribution to the degradation of spatial resolution in CCDs is charge diffusion. In the structure of the CCD there are essentially three regions in the charge-collection process: (1) the depletion region, which includes the charge-carrying channel and depleted bulk beneath it, (2) an undepleted, field-free, neutral-bulk below this, and (3) a region of high recombination that can be the surface of untreated back-side illuminated devices or low-lifetime materials for front-side illuminated devices.

6.8.6 Charge transfer efficiency CTE

Once charge is collected within a pixel, the next major task of the CCD is to transfer this packet to the output amplifier. The CTE is a measure of the device to transfer charge from one potential well to the next. This process for today's CCD is amazingly efficient. Typically for well-made buried-channel devices the CTE will be in ranges of 0.99999/transfers or greater for higher signal levels. This means that for a two-phase device of, for example, 1000 elements on a side, 96% of the charge will remain in the pixel farthest from the output after it has been transferred to the output (4000 transfers). The rest is lost through recombination or dribbles out as a deferred charge tail in the later, trailing pixels.

6.8.7 Read noise

One of the most important CCD parameters is the read noise. For CCD imagers there are essential three regimes of noise to be considered (figure 6.24). At high signal levels, the total noise is dominated by pixel-to-pixel sensitivity variations within the device. This is sometimes referred to as fixed-pattern noise. At the lowest illumination levels the device is limited by noise sources intrinsic to the device. This limiting noise is the read noise floor. Three sources may be classified as (1) trapping state noise, (2) reset noise, (3) background noise, (4) charge transfer noise, and (5) output amplifier noise. The last three are most important.

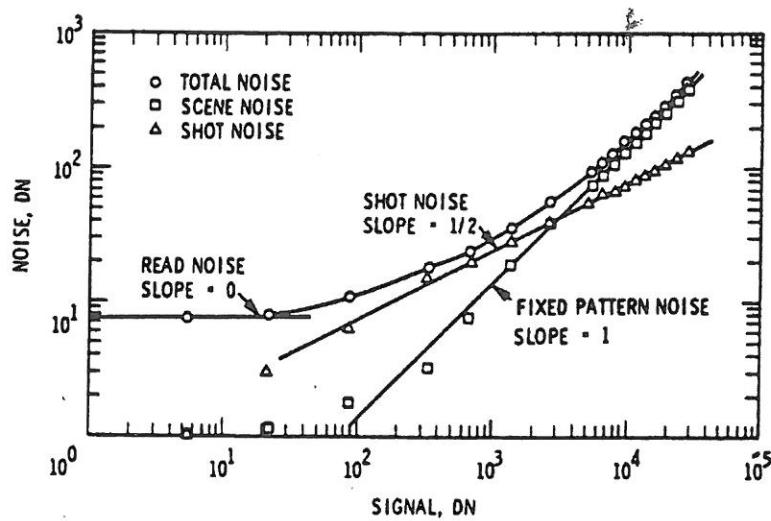


Figure 6.24: Typical plot of noise as a function of input signal for a 20x20 pixel subarray. The curve is divided into three regimes: read noise, signal shot noise, and pixel-to-pixel pattern noise.

6.8.8 Full well capacity

The full well capacity sets the physical limit to the dynamic range of a CCD device. It is the maximum amount of charge that each pixel can accommodate.

Parallel read-out devices

The overall read-out noise is proportional to the square root of the read-out rate. One way to improve the situation is to divide the chip into regions which are read-out in parallel. The Tk1024AB device used in the ALIS detectors is divided into 4 quadrants with parallel signal processing chains, providing a full frame read-out (1124×1024 pixels) in less than 2 seconds, keeping the overall read-out noise below 10 electrons (RMS).

6.8.9 Intensified CCD devices

As with intensified image tubes, the sensitivity of a CCD detector can be improved by installing an image intensifier in front of the CCD. However, it should be stressed that in doing

this, the inherent high quantum efficiency of the CCD (over 80% for some devices) is lost ², since the poor QE of the image intensifier (typically 20%) will set the total QE for the system. Intensified CCD devices are used mostly in older instruments, but also appear where a high time-resolution is desired at the cost of non-linearity, poor QE, etc.

There is a huge amount of literature on the subject of CCD devices. For the especially interested reader we recommend Janesick et al. (1987) and Eccles et al. (1983).

6.9 Charge injection devices (CID)

The device closest to the CCD in mode of operation is the charge injection device (figure 6.25). Each pixel consists of a pair of adjacent MOS capacitors, similar to those of a CCD. One of the two electrodes is held at high voltage so that a potential well is formed beneath it, capable of storing optically-generated charge while the other one is held at substrate potential. When the electrode potentials are interchanged the charge packet moves to the potential well now created under the other electrode, and its presence is detected by capacitive coupling to the electrode. This is a non-destructive process, so that a continuous monitoring of the integrated charge is possible. The CID has been used successfully, but will probably never seriously compete with the CCD due to a ten to twenty times higher read-out noise level.

6.10 Digicon and Reticon arrays

The p-n and p-i-n silicon diodes are such linear, wide dynamic range, efficient and simple detectors that several attempts have been made to build diode arrays suited for scientific imaging. The Digicon is a device consisting of an image intensifier combined with an array of diodes where each diode has its own charge sensitive amplifier, etc. The QE is only around 5% due to the image intensifier which is much lower than for a CCD, but the time-resolution is excellent. The Digicon is well suited for real-time photon detections, and has its applications mainly in spectroscopy.

An alternative approach is to accept the increased read-out noise inherent in diode multiplexing but greatly increase the length of the array to some 1000 to 400 elements. This has been achieved by Reticon Corporation, whose diode array products are usually referred to as Reticon arrays. This detector type was used in the auroral height measuring system (a predecessor to ALIS) (Steen 1988).

6.10.1 Imaging photon detectors

An imaging photon detector (see figure 6.26) is a combination of a photocathode, converting the incident photons to photoelectrons, as described in section 6.5.1. The photoelectrons fall onto a set of micro channel plates, providing an internal gain. Finally the electrons hit a resistive anode, which gives the position of each photon hitting the detector. This detector is

²For those that see a contradiction in gaining sensitivity but losing QE consider the following example: Two people are sitting close to a highway. The first person picks up a stone with 80% probability each time he sees a car, the other person picks up 4 stones with 20% probability each time he sees a car. After 100 cars have passed, both will have approximately the same number of stones but the first one will be better at counting cars on a low-traffic road. The first one has a high quantum efficiency, while the second has higher sensitivity.

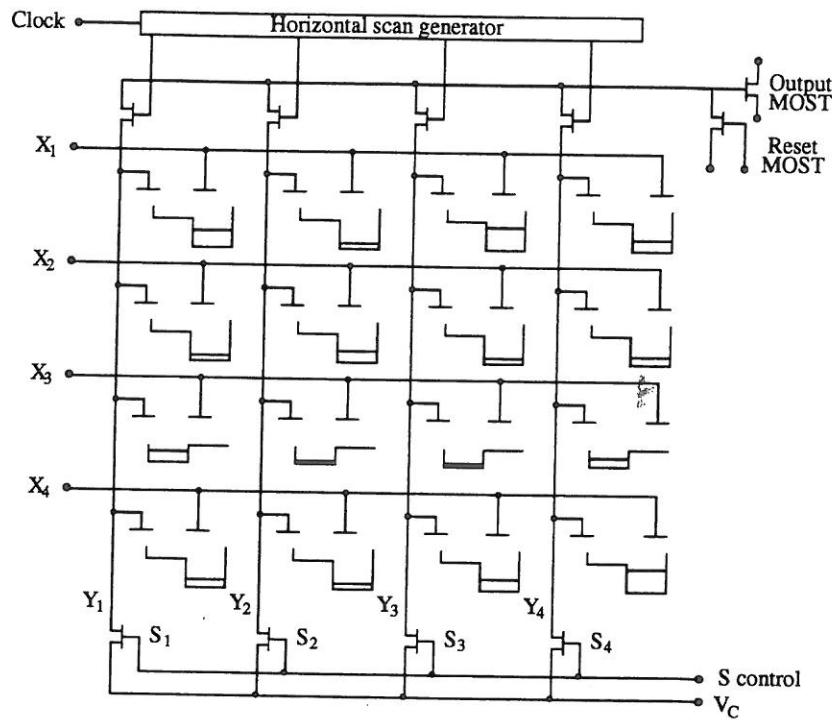


Figure 6.25: Organization and operation of a charge injection device. (Adapted from Eccles et al. (1983).)

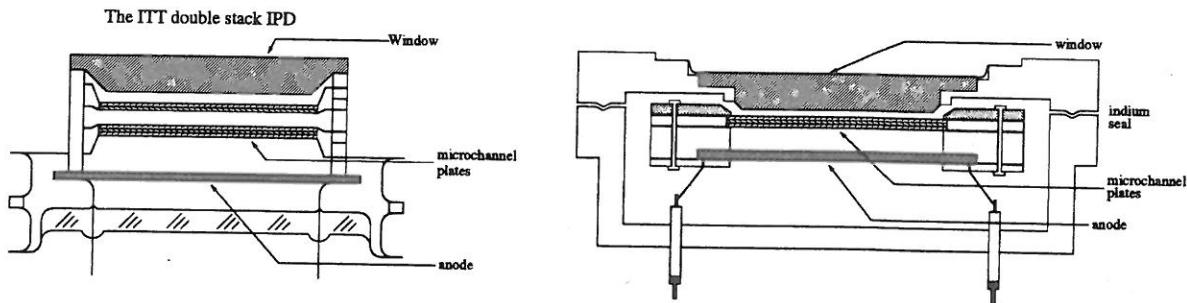


Figure 6.26: An imaging photon detector.

excellent for photon counting and was previously extensively used as detector in instruments such as neutral wind measurements with Fabry-Perrot interferometers. Unfortunately it is not manufactured any more.

6.10.2 Multi anode microchannel array detectors

Another similar detector type is the Multi Anode Microchannel Array (MAMA), where the resistive anode is replaced by an anode array (figure 6.27).

6.11 A brief survey of instruments

In this section, a very brief description of a number of common instruments is presented.

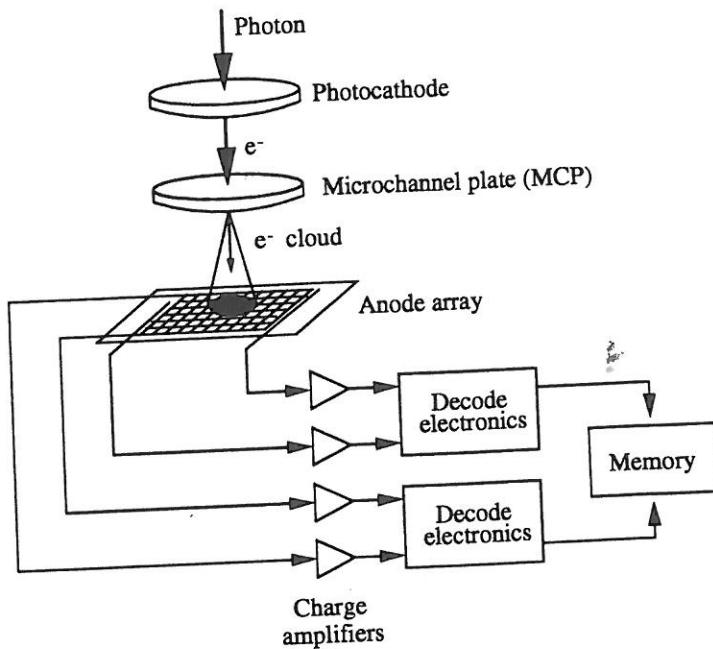


Figure 6.27: Schematic of the imaging MAMA detector system.

6.11.1 The photometer

The photometer is used to make line-of-sight measurements of mainly auroral emissions. Many different types of photometers exist. The classical auroral photometer consists of one or several detector units, each consisting of a photomultiplier tube (section 6.5.2) equipped with narrow field-of-view optics and an interference filter for the emission studied. The detector units are pointed towards a scanning mirror that usually makes meriodinal scans of the sky. Thus a time series of intensity as a function of time and zenith angle are obtained as output data. On spacecraft, the photometer usually consists only of the detector unit, and the orbital and spin motion of the spacecraft is used to obtain scans. A special case of photometers is the imaging photometer (Steen 1988) where the scanning mirror is replaced by a Reticon array (section 6.10). In this way the time-smearing caused by the motion of a scanning mirror is avoided. Modern photometers usually use solid-state detectors instead of a PMT.

6.11.2 The all-sky camera

All-sky cameras have been in wide use since the international geophysical year in 1957–1958 (Stoffregen 1962). The basic idea is to image the entire sky at a constant time interval. The primary use of the all-sky camera data was to study the occurrence and morphology of the aurora, but many other studies are also made (e.g. to check if the sky is clear enough to trust other optical measurements, or to image clouds, etc.).

Today auroral science is mostly made using other instruments than all-sky cameras, e.g. satellites, radars, photometers, and high performance digital imaging systems (e.g. ALIS section 6.11.4). However, it should be noted that most of these instruments do not provide a general overview of the sky. For this reason, the all-sky camera is still an important tool, providing rough information on the actual auroral situation. This information is often used

to validate other measurements.

The present camera in Kiruna is an old modified FMI-77 camera (Hypönen et al. 1974) (originally designed by Finnish Meteorological Institute) which makes recordings on 16 mm color film (Kodak Ektachrome 7251). It will be replaced by a digital system in a near future.

Some of the key requirements for a new all-sky camera system are listed below:

- Retain the color information.
- Capable of observing in daylight.
- Fully remote-controlled.
- Data directly accessible through the World Wide Web.

Description of the modified FMI-77 camera used in Kiruna

The sky is imaged onto a convex mirror (figure 6.28), which is photographed by the 16 mm film camera located above the mirror. This design results in a loss of information at the Zenith. The film camera is controlled by electronics, which exposes the camera at a selected interval, (typically once every minute) synchronizes the exposures with UTC, (using GPS) and advances the film, etc. Typical exposure time is 4 seconds, but every eighth exposure is four times longer, in order to reveal faint auroras. When the camera is not in use, or during heavy snowfall, a cover automatically closes and protects the mirror. An electrical heater keeps the mirror free of moisture. The first two digits of the time display should be ignored, then follow: Year, (two digits) UTC day, (3 digits) Hours, minutes and seconds. The betalights were used long ago for intensity estimations; this is nowadays done using more accurate digital imaging systems.

A new Finnish all-sky camera has been developed, that uses an intensified CCD detector combined with a filter wheel. Instead of a mirror a fish-eye lens is used (Syrjäsu 1997). Other types of all-sky cameras also exist.

6.11.3 Triangulation

Between 1700 and 1910 the height of the aurora was the subject of much discussion. Somebody believed that the mean height of the aurora was about 1000 km while others believed that it reached all the way down to the Earth's surface. In 1910-40 Carl Störmer and his assistant measured the height by triangulation using cameras (see figure 6.29) and found that most of the auroras were between 90 and 200 km (Egeland et al. 1996). The basic concept of triangulation is that the same object (in this case the aurora) is observed simultaneously from two different positions. The height of the object can then be calculated if the distance, d , between the two observers is known. Both observers measure the angle of the line of sight relative to the horizon, α and β , as sketched in figure 6.29. The height, h , is calculated from

$$h = \frac{\sin\alpha \times \sin\beta}{\sin(\beta - \alpha)} \times d. \quad (6.8)$$

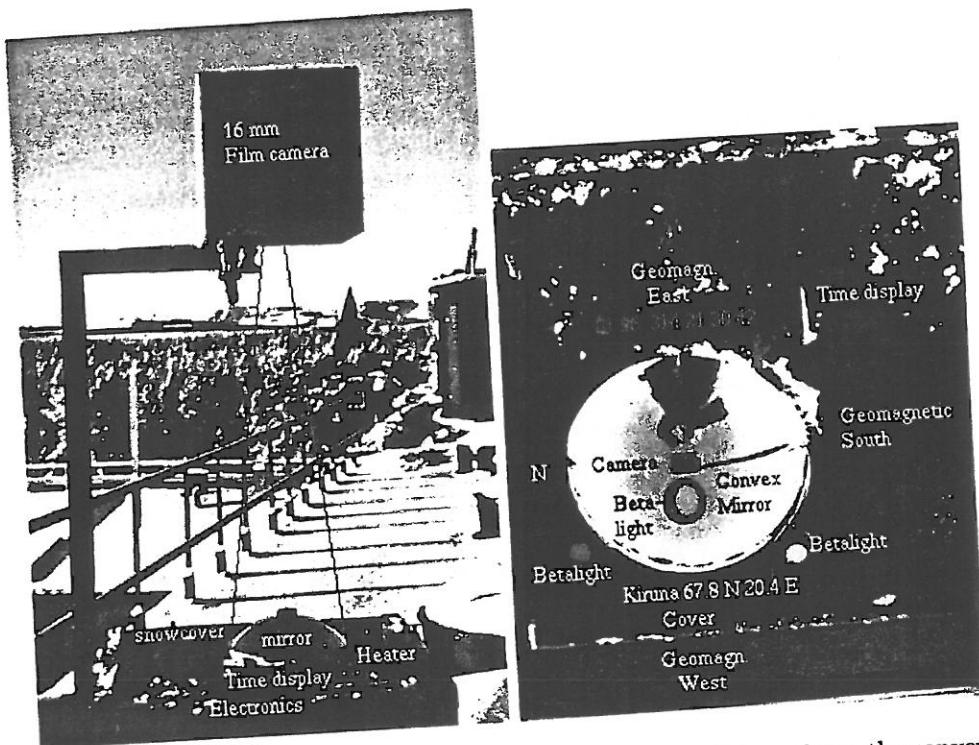


Figure 6.28: (a) Side view of the FMI-77 camera; (b) This picture shows the convex mirror viewed almost as the 16 mm film camera sees it (the betalight and camera are at the center of the mirror in a "real" image). The supporting rod for the camera is towards geomagnetic South.

6.11.4 ALIS

A new initiative to ground-based auroral imaging that gives 3-dimensional images is the multi-station facility ALIS (Auroral Large Imaging System, figure 6.30) located in Northern Sweden. The original idea was published in 1989 (Steen 1989) and recent descriptions of the system are found in (Steen et al. 1999, Brändström et al. 1999, Steen & Brändström 1993).

The multi-station concept

ALIS consists of a number of unmanned, remote-controlled stations (see figure 6.30). The design of a station resembles that of a satellite, so that all vital functions can be remote-controlled and monitored from the control center located in Kiruna. Secondary control centers can exist at any site with an Internet connection. All configuration and operation of the system is software-controlled using graphical or text-based user interfaces and a special experiment configuration language.

The ALIS camera

The ALIS cameras consist of a telecentric lens system consisting of a front lens field-of-view $54 \times 54^\circ$ a 6-position filter wheel, more optics and a CCD detector. The Peltier-cooled CCD chip used is a Tk1024AB thinned backside illuminated quad output-device, with a quantum efficiency of about 85 % and a maximum frame rate below 2s. The entire camera unit is

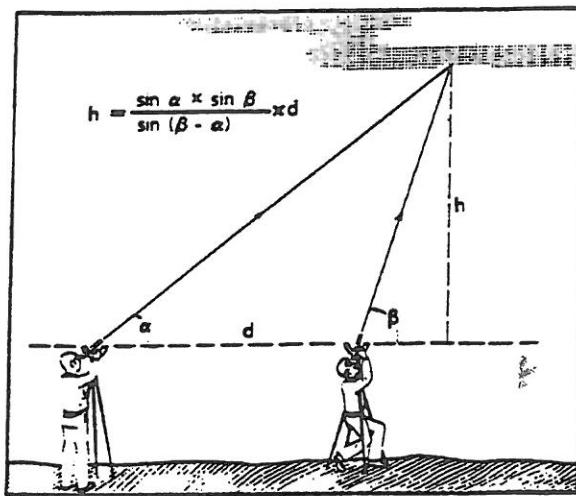


Figure 6.29: Two observers at a known distance from each other (d) measure simultaneously the direction (α and β) to a specific point in the aurora (the same stars are usually seen in the photographs). The height is then easily calculated from $h = \frac{\sin \alpha \times \sin \beta}{\sin(\beta - \alpha)} \times d$. (Adapted from Egeland et al. (1996).)

mounted in a precision camera-positioning system that makes it possible to image any part of the sky.

6.11.5 Fabry-Perrot interferometers

The Fabry-Perrot interferometer, FPI is a device that uses interference in thin-layers to determine the wavelength of the input signal (figure 6.32). This is used to determine the velocity of the neutral particles at around 250km altitude (the F-region neutral wind) by measuring the Doppler shift of various airglow (and auroral) emissions (line of sight measurements).

The majority of FPIs located around the world are pressure-scanned, and use photomultiplier detectors that only offer 1-dimensional scans at low time resolutions of several tens of minutes. A more advanced type of FPI uses Imaging Photon Detectors (IPDs) with very high light-gathering properties that provide 2-dimensional images and a time resolution that is at least an order of magnitude higher (figure 6.32). Many of these FPIs run continually, under computer control. Data can be accessed remotely and the aim is to be able to control the instruments remotely as well. The majority of other FPIs are run only on a campaign basis.

More reading about these devices is found in e.g. McWhirter (1993) or Vaughan (1989).

6.12 Optical absorption spectroscopy

Absorption spectroscopy using extraterrestrial light sources is a widely used technique for the measurement of atmospheric trace gases, in particular for remote sensing of stratospheric species. Gases like ozone, CFCs, NO_x , BrO_x , ClO_x , etc. (Brasseur & Solomon 1986) show characteristic absorption features in the UV, visible and infrared spectral ranges, which can be detected in spectra of solar, lunar or stellar light measured at the ground or from satellites.