

CCD Technology Primer

1. CCD Sensors

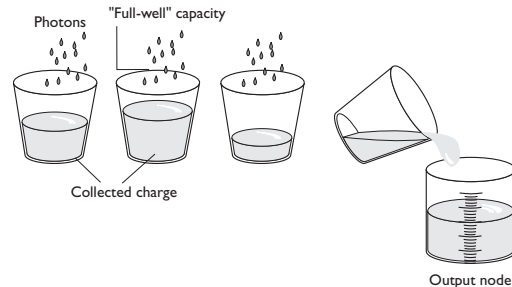
Charge Coupled Devices (CCD) have many applications, but the most important is in imaging. The basic operation of the sensor is to convert light into electrons. When light is incident on the active area of the image sensor it interacts with the atoms that make up the silicon crystal. The energy transmitted by the light (photons) is used to enable an electron to escape from the tight control of one atom to roam more freely about the device as a "conduction" electron, leaving behind an atom shy of one electron. The electron-deficient atom is often referred to as a hole. The more photons incident on the sensor, the more electron-hole pairs that are generated.

Since the photons must have sufficient energy to accomplish this, low energy photons (long wavelengths) are less easily detected and tend to pass further into the silicon crystal. High energy photons (short wavelengths) on the other hand are absorbed more closely to the surface of the sensor and may not reach the active part of the detector. Hence, there will be a spectrum over which the sensor will operate, falling off at short and long wavelengths. The number of electrons generated per photon is known as the "quantum efficiency", or QE.

The electrons can be separated from the holes within the silicon and collected (integrated) just as a bucket would collect rain water falling into it. The amount of charge (or water in this analogy) collected will depend on the light intensity, its spectrum and the integration time.

By setting out a line or an array of such "buckets" it is possible to build up a representation of the image incident on the sensor. The term "full-well" capacity describes the maximum charge that can be stored in each "bucket" before it spills over. Therefore, at regular intervals the "buckets" must be emptied and the amount of charge measured to determine the local light intensity. This is accomplished using a charge coupled device, or CCD register. A measuring device sits at the end of the row, known as the output node. The nearest "bucket" can be emptied into the measuring device to produce a signal that depends on the amount of stored charge. The empty "bucket" can be refilled from the next "bucket" along, and so on down the line. In this way the measuring device can be used to determine the amount of charge in all the "buckets" without having to move them from their positions. This ability to generate a serial data stream from a large number of "buckets" (or pixels) enables the light incident on the sensor to be sampled with high spatial resolution in a controlled and convenient manner.

Figure 1. CCD Pixel "Buckets"



The electrical equivalent of the storage "buckets" are MOS capacitors. A CCD register can be thought of as a series of such capacitors, arranged closely together across a silicon wafer so that charge can be moved from one capacitor to the next as efficiently as possible. The MOS capacitors are formed by depositing a highly conducting layer of polysilicon on top of an insulating layer of silicon dioxide ("oxide") covering the wafer substrate.

By applying a positive voltage to the polysilicon gate, with respect to the substrate, an electric field is generated within the structure. The substrate is lightly doped p-type and processed carefully to minimize the presence of contaminants that may give rise to defects. Any conduction electrons close to the surface will be attracted towards the gate, accumulating at the silicon/oxide interface, while the holes are pushed into the substrate. This forms a storage region, or "bucket" for electrons underneath the gate. The extent of the storage region is determined by the applied voltage, which can be turned on and off as required.

The operation of CCDs is complicated by thermal energy. At room temperature, the thermal energy of the silicon creates electron-hole pairs even in the dark. These electrons are sufficient to completely fill the storage region. However, if a positive voltage is first applied to the gate, there is an initial deficiency of electrons, providing an empty storage region for electrons generated by incident photons. Eventually the thermal generation of electron-hole pairs, the rate of which is very sensitive to temperature, will contaminate the signal. Therefore the storage region can only be used to integrate the signal charge for a limited time unless the thermal generation is reduced by cooling the sensor.

To prevent the electrons from simply spreading out across the sensor, the storage regions are laterally confined. Each row of a storage region is isolated from the others by a potential barrier using a highly p-doped trench under the edges of the gates. The higher doping level requires a higher gate voltage in order to form a storage region. In normal operation the gate voltage is not high enough, preventing charge from spilling sideways out of the lower doped storage regions. Formed by ion implantation, the highly p-doped

trench is known as the “channel stop”. Along the row of storage regions the charge packets are separated by using one or more capacitors biased to a low voltage to form an electrostatic barrier. Therefore, more than one capacitor (or gate) is required per pixel to store a single charge packet. By changing the voltage on the gates it is possible to alternate between the storage mode and barrier mode as required. The stored charge packets can then be moved along the CCD register in a controlled manner without interfering with each other by manipulating the gate voltages.

Buried Channel Devices

However, with this simple arrangement the signal charge is stored very close to the silicon/silicon-dioxide interface where the differences in material properties produce electronic defects that interact with the signal charge and degrade the sensor performance. This problem can be nearly eliminated by moving the storage region deeper into the silicon and away from the oxide. This is accomplished by incorporating a thin n-doped layer of silicon underneath the oxide.

This buried layer is fully depleted under normal operation and moves the temporary charge storage region deeper into the substrate and away from the interface to form what is known as a “buried channel” device. Nevertheless, if the storage regions are allowed to fill up, the electrons can still come fairly close to the surface. The best performance is a compromise, avoiding full wells to reduce interaction with the interface defects and avoiding nearly empty wells to maximize the signal-noise ratio.

Charge Transfer Efficiency

The maximum transfer rate at which the CCD can be operated is limited by the charge transfer efficiency (CTE), the fraction of the charge packet successfully transferred from one storage gate to the next. Because of bulk silicon defects

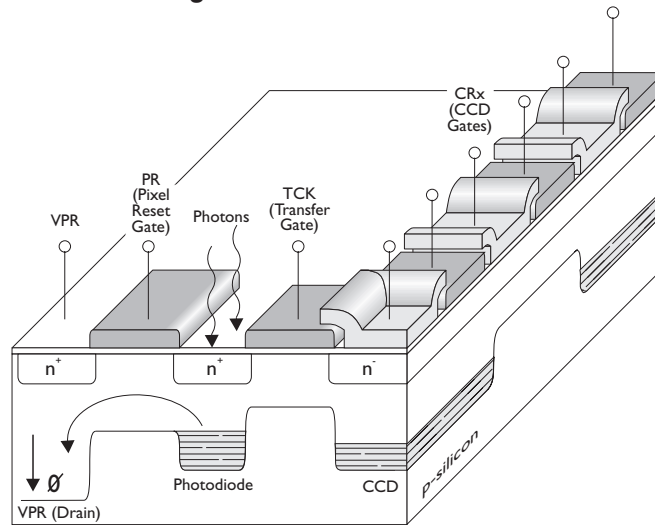
and the finite time interval required for the electrons to move, there is always a small amount of charge left behind. This charge will mix with that in the charge packets that follow and degrade the fidelity of the signal. Although the error is cumulative, growing with the number of pixels, it is greatly reduced in DALSA sensors through careful CCD design and manufacture.

Smear

Another source of error can arise when the stored charge packets are shifted through the imaging region CCD towards the output node. During this readout time, the photogeneration of electrons continues. This can contaminate charge packets that have originated from one part of the image with additional electrons from another. This effect is known as “smear”; it can be minimized by ensuring that the shift register is read out in a shorter time interval than used during the collection of signal electrons (the integration time). Hence, for a given integration time, low smear requires a high readout rate.

Another source of smear arises when electrons generated in the silicon are “captured” by a nearby storage gate rather than the storage gate upon which the light was incident. This is a statistical effect due to the random process of diffusion by the electrons towards the storage region. As longer wavelengths (e.g. red light) are absorbed more deeply in the sensor the smear tends to worsen. In specialized applications that require very low smear it is possible to greatly reduce this effect by providing the storage regions with separate p-doped regions in a lightly doped n-type substrate.

Figure 2. CCD Cross-Section



Antiblooming & Exposure Control

When the photogeneration rate is very high (e.g. a bright image) it may be possible to exhaust the storage capacity of the CCD well. Excess carriers can then spill over into adjacent pixels and degrade the accuracy of the signal. This is known as “blooming”; it can be minimized by decreasing the integration time. However, even this may not be sufficient and some means to drain off this excess charge has to be provided.

A scheme which gives greater flexibility uses a separate photodiode as a light sensor. An isolated n+ region can be formed in the p-substrate adjacent to a CCD pixel through standard processing. Signal charge is integrated by the n+ capacitors but cannot move between them. A transfer gate must be provided which can be biased to form a channel between the photodiode and a nearby CCD storage region. This gate can be common to all the photodiodes, enabling all the charge packets to be transferred to an adjacent CCD register in parallel.

Another gate on the other side of the photodiode, called the “pixel reset gate”, enables the charge to be transferred from the photodiodes to a common drain. This second gate provides a reset function since it can be used to empty the photodiodes of charge without affecting the signal in the CCD. This enables the integration time to be controlled independently of the CCD read out rate and thus provides electronic exposure control. During integration the reset gate bias can be adjusted to ensure that excess charge is drained off before a photodiode is completely filled, thus providing a degree of “lateral antiblooming”.

Fixed Pattern Noise & Photo Response Non Uniformity

When exposure control is not used, the potential of the photodiodes at the start of integration is determined by the transfer gate high level. However, nonuniformities in the transfer gate structure results in small differences in the potential underneath the gates. In addition, the finite resistance of the gate lines produces a voltage variation across the sensor which also produces a variation in the potential underneath the gates. This can produce small differences in

the potential of the photodiodes at the start of each integration period. Because it is a feature of the gate structure it is fixed, hence the term “fixed pattern noise” or FPN. This variation will not, however, influence the amount of signal charge transferred to the CCD, which depends only on the integration time and photogeneration rate.

However, when exposure control is implemented, the potential of the photodiode at the start of integration is now determined by the reset gate bias. If this bias is higher than it would have been if determined by the transfer gate, then an initial amount of charge is introduced. Even under zero illumination, this charge will be transferred to the CCD register by the transfer gate and appear as a signal. This is not in itself a problem since a dc signal is easily removed from the output signal. However, the nonuniformity in the bias underneath the transfer gates will produce a DC offset that varies from pixel to pixel increasing the FPN.

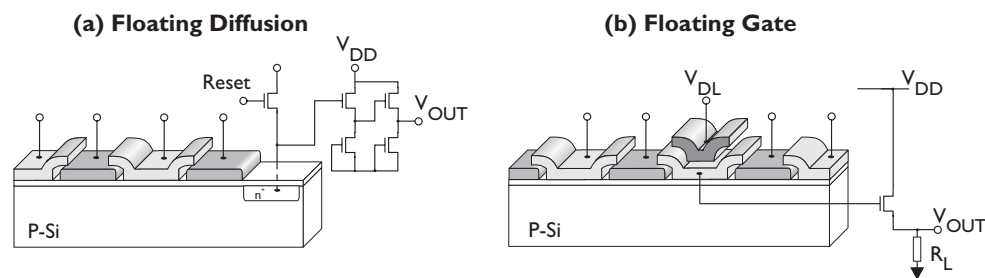
In applications requiring very high uniformity it is possible to compensate for the FPN by storing the signal generated under zero illumination and subtracting it from subsequent signals. This must be carried out in signal processing electronics within the imaging system.

Another consequence of small nonuniformities in the sensor is a variation in the charge produced by different pixels under identical illumination. This can arise, for example, when there are small variations in the light collecting area of the photodiodes. When necessary, this photo response non uniformity (PRNU) can be compensated for using a similar technique to that used for FPN.

Output Node & Reset Noise

The charge packets in the CCD register are too small to be transferred directly to the outside world without significantly degrading the signal-to-noise ratio. Instead, a small sensing capacitor is used to convert the signal charge to a voltage which can be amplified before transmission. In other words, the role of the output structure is to convert the small charge packets to a voltage. The performance of this stage is judged in terms of its charge conversion efficiency (CCE), the voltage generated per unit charge. The output structure achieves this by dumping the charge on a floating gate. The former is the most common, being generally

Figure 3. Output Structures



easier to fabricate and less temperature sensitive. However, it is necessary to reset the n^+ storage capacitor by draining the charge after each measurement. In the “bucket” analogy this is the same as having to empty the measuring jar at the end of the register.

The floating gate method, on the other hand, has the advantage of being non-destructive since the charge packet can be transferred back out again. This can be likened to the use of weighing scales to determine the amount of water stored in a particular “bucket”. Furthermore, the floating gate has the advantage that it does not have to be reset after each charge packet has passed through. This can be important because each reset operation is subject to thermal noise which appears as a pixel to pixel variation.

When required, reset noise can be virtually eliminated using correlated double sampling (CDS). The noise level on the output capacitor is determined immediately after each reset pulse by a fast sample and hold circuit and later subtracted from the signal. To maximize the overall system signal-noise ratio it is important to maximize the charge conversion efficiency, which means that the capacitance of the output node should be as small as possible. Additional steps can be taken to minimize this capacitance using feedback techniques to reduce undesirable parasitic capacitances.

2. Sensor Architectures

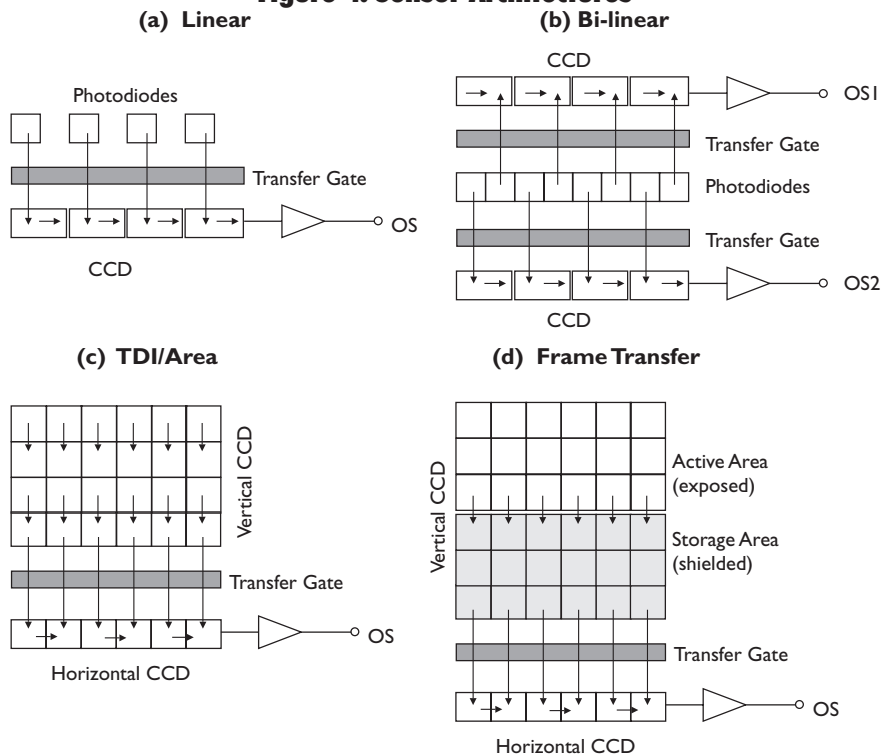
Line Scan Sensors

The simplest architecture is a linear sensor, consisting of a line of photodiodes adjacent to a single CCD readout register. In this arrangement the maximum pixel density is determined by the size of the CCD gates, of which there are more than one per pixel. Furthermore, large pixel counts require a relatively longer time to readout the data. To overcome these limitations it is possible to place additional CCD registers in parallel, with alternate photodiodes connected to different CCD registers. In this way the photodiodes can be packed more closely together, and the number of charge transfers is reduced by a factor equal to the number of CCD registers. For large pixel counts this allows higher readout rates. It is common to use at least 2 CCD registers to form a bi-linear sensor.

Area Array Sensors

For two dimensional images an area array is formed from a set of parallel light sensitive CCD registers, which by convention are assumed to be orientated up and down the page. At the lower edge of this array a single horizontal CCD register (HCCD) is used to combine the outputs from the vertical registers (VCCD) into a single output. This

Figure 4. Sensor Architectures



produces a compact and simple full frame area sensor. However, for large arrays the slow frame rate can result in a large amount of smear because the vertical registers remain exposed to the image and can only shift out one charge packet every time the horizontal register is emptied. To avoid this, the vertical CCD registers can be extended downward so as to double their length, with the lower half being shielded from the image to form a storage region. In this way it is possible to generate a signal from an image in the image region in the upper half and quickly transfer the signal down into the storage region. This is known as a frame transfer (FT) sensor. Often an additional transfer gate is positioned between the horizontal register and the vertical CCD registers to prevent charge transfer into the horizontal register while it is being emptied.

Interline-Transfer & Fill Factor

In some applications it is desirable to provide a column of photodiodes for each vertical CCD register combined with a reset gate for lateral antiblooming and exposure control. With this arrangement the image can be integrated by the photodiodes, transferred into the vertical CCD registers and then down into the frame store. The disadvantage of this frame interline transfer (FIT) sensor is a reduction in the fraction of the light sensitive area in each pixel. This fraction is called the "aperture ratio" or "fill factor."

Color Filter Array

Color sensitive imaging requires the use of color filters or a prism. The compact solution is to incorporate a set of primary (or complementary) color filters onto the sensor array. There are a number of different schemes for organizing the layout of the colors and the recovery of the separate color signals.

Time Delay & Integration Sensors

Area sensors have another important application as time delay and integrating sensors (TDI) for linear imaging. In this technique the linear image is mechanically scanned from the top to the bottom of the sensor. The vertical CCD registers are clocked to ensure that the charge packets are transferred at the same rate and in the same direction as the image. This ensures that the signal charge building up in the CCD remains aligned under the same part of the image. In this way, the image signal can be integrated for much longer than is allowed by the temporary storage time of a single pixel. The sensitivity of the sensor is increased by a factor roughly equal to the number of "stages" in the vertical direction. Hence the TDI sensor can be used in low light levels, or used at very high scan speeds. It has the further advantage of averaging out any nonuniformities in each vertical column and gives an enhanced signal to noise ratio. At the bottom of the sensor a transfer gate is often used to control the movement of charge into the horizontal readout register.

3. Sensor Clocking Schemes

Four Phase Clocking

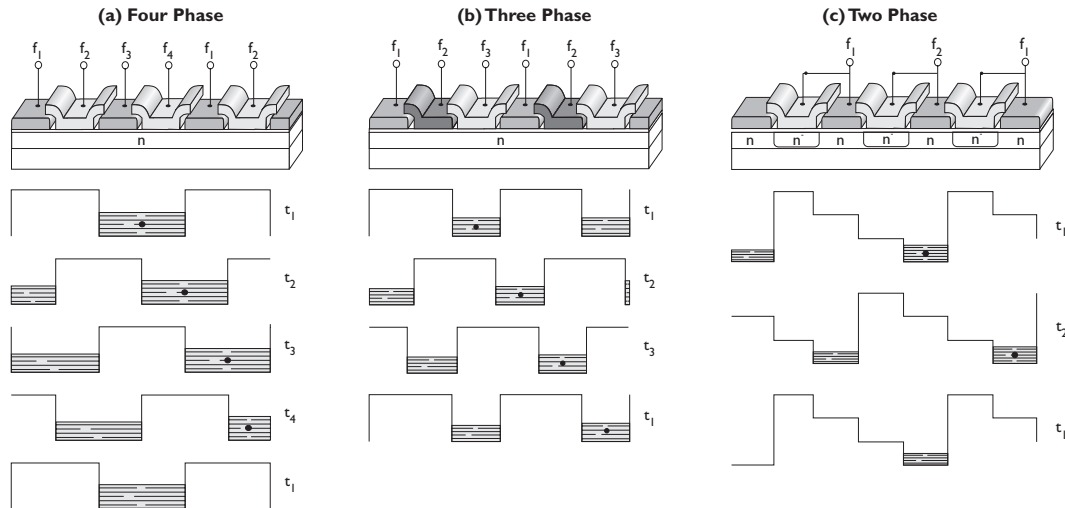
A common clocking technique is the 4-phase clocking system which uses 4 gates per pixel. At any given time, two gates act as barriers (no charge storage) and two provide charge storage.

In normal operation the process begins when the right most storage gate is switched to a higher (positive) voltage. See Figure 5a. This widens the storage region towards the right hand barrier and allows the charge to spread out underneath the extra storage gate. At the same time, the left most storage gate is switched to a lower voltage. This narrows the storage region from the left hand side and forces the charge into the region underneath the remaining two storage gates. The net result is a movement of the signal charge to the right by one gate width. This sequence is repeated another three times to move the charge packet along to the next pixel. By clocking the gates continuously in this manner it is possible to shift the charge along the register from left to right. In the horizontal CCD or linear sensor this process is called readout, and the readout clocks are designated CR1, CR2, CR3 and CR4. The transfer gate between the horizontal and vertical registers is controlled by the transfer clock TCK, the photodiode reset gates are controlled by the pixel reset clock PR and the output capacitor reset clock is designated RST. In area array and TDI sensors the transfer of charge down the vertical CCD registers is controlled by the imaging clocks CI1, CI2, CI3 and CI4. The readout clocks are used to control the transfer of charge along the horizontal CCD to the output. During readout of the horizontal CCD it is possible to continue transfer of the image using CI1-CI4 in preparation for a quick transfer into the horizontal register when the transfer clock goes high. However, this can result in feedthrough of the CI clock edges into the video signal. This feedthrough can be avoided using burst mode clocking, whereby the CI clocks are held fixed until the horizontal CCD has been read out. Unless the speed penalty is unacceptable this scheme is recommended.

Three Phase Clocking

The spatial resolution can be improved using three phase clocking, which requires three gates per pixel. See Figure 5b. It differs from the four phase system by using only one storage gate and two gates acting as a barrier. The consequence is a reduced storage capability of the pixel by a third but the operation of the sensor is more robust with respect to the clock timing than is the case with a four phase clocking scheme. Furthermore, by using three independent polysilicon layers the three clock phases need not "share" the same polysilicon layer, thereby improving fabrication yield.

Figure 5. Clocking Schemes



Two Phase Clocking

Further simplification is possible by restricting the movement of the charge packets to one direction along the CCD using a two phase system. See Figure 5c. Four gates are needed for each pixel, with adjacent gates connected together in pairs, which in turn are connected to alternate clock lines. Of each pair of gates, one is designed with an increased n-type doping level underneath it. When a bias is applied to the pair of gates the potential underneath the gate with extra doping is more positive. This makes the charge storage area locally deeper, producing a step in the potential profile. In this way electrons are forced to move along in only one direction. By alternately pulsing the clock lines high it is possible to move the charge packets along the CCD in the direction determined by the position of the doping. A significant advantage of this scheme is the insensitivity to the clock phases, making it more robust than both the three and four phase schemes.

Binning

A special clocking arrangement can be used to combine charge packets from a number of neighboring pixels before sending them to the output amplifier. This reduces the spatial resolution of the sensor, but provides an improved signal-noise ratio in low light conditions by combining several small charge packets. As with the TDI sensor, the addition of many charge packets reduces the noise level by averaging out pixel to pixel variations. This technique is referred to as pixel binning.

4. Radiometry and Photo Responsivity

Radiometry is the science of measuring light, whether visible to the human eye or not. Photometry, a subset of radiometry, is defined in the same way except that the measurements refer to light as detected by the human eye. Photometry as a field of radiometry has its own set of units including Lux, Candela, Nit, Talbot and Lumens to name a few, while radiometry uses only standard SI units such as Joule, Watt, and meters. The conversion between the two sets of units is based upon a standard eye response curve known as the photopic curve. For low light level applications, a different standard eye is used, the scotopic (dark adapted eye) curve, which shows a different response and peaks at a different wavelength.

The photopic curve can be approximated by a 100% response between 500 and 600nm to avoid the need for computing integral calculations when performing conversions between photometric and radiometric units. If information on the illumination is given in photometric units it will be necessary to convert these to radiometric units to derive the camera response.

Irradiance

In order to determine the optimum camera set up for most applications it is useful to calculate the expected output from a camera given knowledge of the lighting conditions and the nature of the objects to be imaged. If the scene contains many different objects with different colors and reflective properties it is very difficult to calculate the nature of the reflected light. The amount of light falling on the scene is

known as the irradiance and is expressed as a power density.

For example, if all the light coming from a 100W source is uniformly incident on a 1 x 1m surface the irradiance is 100W/m². Some of this light will be reflected by the surface into one or many different directions. The light reflected from the surface and received at a point some distance away from it is known as the radiance, the power per unit area, per unit solid angle.

A scene containing many different objects is difficult to characterize when their reflectance varies widely, particularly when some highlights in the scene are brighter than the rest. In order to accommodate the wide latitude of a complex scene it is possible to use an average value for the reflectance. In this way, the radiance is assumed to be equivalent to that which would have resulted from a "standard" surface. As with conventional photography, a reflectance of 18% represents a fair estimate. In applications where it is necessary to image a particular object this value is unlikely to be appropriate. The complex calculations required to determine the reflected light are difficult and therefore, at best provide only a rough guide to the camera response. Hence it is often easier to measure the average visible light intensity reflected from an object with a simple light meter, bearing in mind that the spectrum will depend on the nature of the illumination and the object. Nevertheless, if the irradiance of the camera is known, then the output of a camera system can be calculated by integrating the known spectral response function of the camera with the spectral illumination curve at the sensor.

Light Source

To determine the spectral curve of the illumination source it is often possible to model it using an equivalent black body, an ideal object which emits light in a smooth and

continuous spectrum determined only by its temperature. At low temperatures this light will be at relatively long wavelengths (about 13µm at 300K) and will appear to be perfectly black to the naked eye. At higher temperatures the peak of the spectrum shifts to shorter wavelengths and the intensity increases. Hence some hot objects appear to glow red, and at even higher temperatures they can appear to be white. The black body curve doesn't work for non-incandescent light sources. Fluorescent tubes, lasers etc. which have their own unique emission spectra are often difficult to work with. Nevertheless, the color balance of fluorescent light sources including computer monitors are sometimes characterized by an equivalent black body "color temperature" although they are often rich in light from the green part of the spectrum.

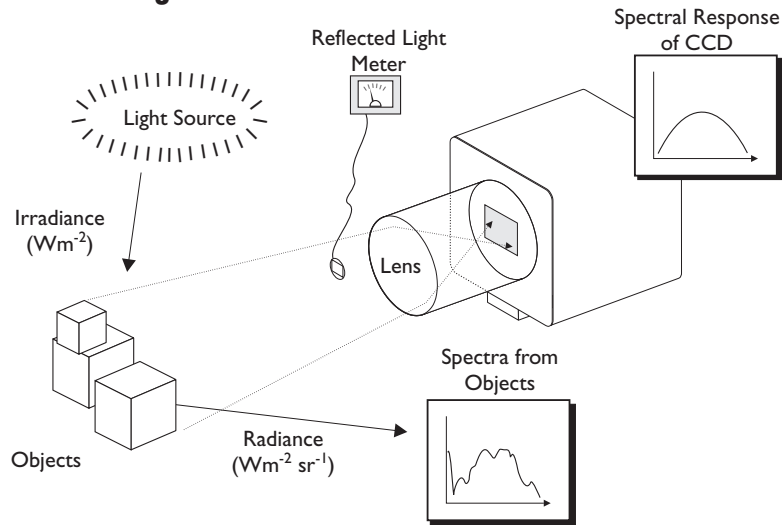
Camera Lens

The light reflected from the scene is collected by the camera and focused onto the CCD sensor by the lens. In addition to providing a sharply focused image the lens also increases the light intensity. Strictly speaking, the lens will absorb some of the light and will itself have a spectral absorption response, but for the purpose of estimation this can be ignored. The amount of light collected by the lens will also depend on its diameter. As a result the effective light gathering power of the lens is related to its f-number, the ratio of its focal length to diameter, or more accurately, directly proportional to the numerical aperture (NA), the reciprocal of 2 x f-number. Hence, it can be shown that the irradiance of the CCD sensor is:

$$\text{Irradiance} = \pi \times \text{object radiance} \times \left(\frac{\text{NA}}{(m + 1)} \right)^2,$$

where m = magnification

Figure 6. Radiance and Irradiance



Responsivity

The spectral responsivity curve of the sensor describes the way in which the sensor responds to a given quantity of light input at particular wavelengths. Over a broad range the sensor response is linear with respect to the exposure time (integration period). This curve should be combined with the spectral illumination curve to derive a sensor output for a given input light level. The combination of the two curves is carried out by integrating over the spectral response of the sensor. This is non-trivial, so it is usual to simplify the response curves by discretizing them into a number of simple rectangles. Despite the use of such spectral response calculations, it is recommended that any proposed set up be tested experimentally with the intended camera, particularly when using daylight illumination.