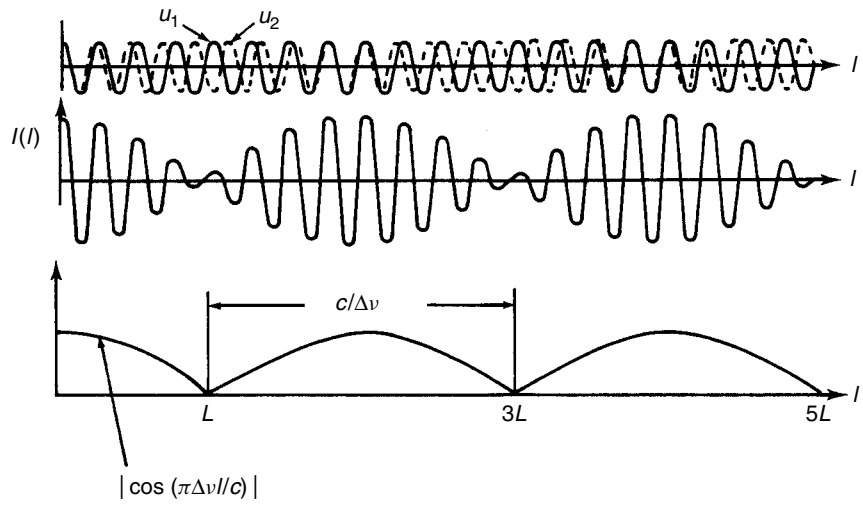

Figure 5.18 (a) Field amplitudes u_1 and u_2; (b) intensity distribution; and (c) temporal degree of coherence (visibility function) resulting from two resonator modes of a laser

5.5 HOLOGRAM RECORDING MEDIA

5.5.1 Silver Halide Emulsions

When making a hologram recording such as that illustrated in Figure 6.1, the photosensitive material must resolve the complicated intensity distribution resulting from the interference between the object and reference waves. The mean frequency in this pattern is given by the mean angle between these two waves. In practice this angle can be about 20° and greater. From Figure 3.5 we see that this angle corresponds to a spatial frequency of about 500 lines/mm. Normally a resolution of 1000 to 2000 lines/mm is desirable. This criterion is met by several silver halide emulsions, having a resolution of up to 5000 lines/mm. They also have a high sensitivity from about 1 to 10 $\mu J/cm^2$.

In the description of hologram recording in Section 6.2, we assume a linear relation between the amplitude transmittance t and the exposure E of a hologram, where E is the intensity times the exposure time, i.e. the energy density per unit area. This assumption is not strictly true. A typical $t-E$ curve for a film emulsion is shown in Figure 5.19. Another common transmission characteristic of film is the Hurter-Driffield curve, which is a density versus log E curve. The density D is defined as

$$D = \log \frac{1}{|t|^2}$$

Density is a common parameter for ordinary photography since the eye detects brightness differences on an approximately logarithmic scale. Photographic films are often characterized by the slope γ of the linear portion of the $D-\log E$ curve.

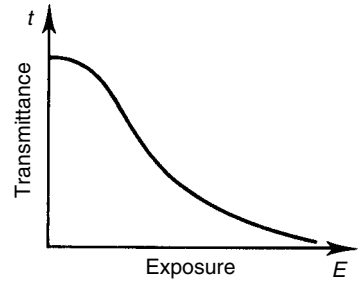

Figure 5.19 $t-E$ curve for holographic emulsion

In Section 6.5 we shall see that to obtain maximum diffraction efficiency of an amplitude hologram, the bias transmittance t_b should be equal to 0.5 which corresponds to $D = 0.6$. Because of the non-linearity of the $t-E$ curve it is advantageous to have t_b slightly lower (i.e. D slightly higher) than this value. In ordinary photography the bias density is centred at the linear portion of the $D-\log E$ curve which does not coincide with the linear portion of the $t-E$ curve. A properly recorded amplitude hologram therefore looks underexposed compared to a photograph.

Since the previous edition of this book, the use of silver halide emulsions in holography has gone down drastically. The holographic emulsions from Kodak, Agfa and Ilford are no longer on the market.

5.5.2 Thermoplastic Film

A thermoplastic, e.g. Staybelite Ester 10, is not photosensitive and must therefore be combined with a photoconductor in a film structure (Urbach and Meier 1966). The system is usually built up from a substrate of glass upon which is coated a conducting layer of, for example, tin oxide. On this is deposited a photconductor such as polyvinylcarbazole sensitized with trinitro-9-fluorenone, and on top of this is deposited a thermoplastic layer. The recording technique consists of a number of steps, beginning with establishing a uniform electrostatic charge on the surface of the thermoplastic with a corona discharge assembly. This charge is capacitively divided between the photoconductor and the thermoplastic layers, and upon subsequent exposure the interference pattern causes the photoconductor to discharge in a spatial pattern dependent on the exposure. However, this does not cause any variation in the charge on the thermoplastic. This is accomplished by recharging the surface uniformly, which increases the charge in the illuminated areas. The thermoplastic is then heated to the softening temperature, allowing electrostatic forces to deform it so that it becomes thinner at illuminated areas and thicker elsewhere. Cooling quickly to room temperature, the deformations are frozen in, resulting in a hologram with thickness variations, i.e. a phase hologram. Reheating the thermoplastic to a higher temperature tends to restore it to its original state. Thus the material has a write–erase recycling capability. This is a quite complicated procedure, but complete camera units are commercially available, giving a hologram ready for reconstruction within 5 s of the exposure.

A peculiar feature of the thermoplastic film is that it has a band-limited spatial frequency response centred at about 1000–2000 lines/mm. The sensitivity is between 10 and 100 $\mu J/cm^2$.

5.5.3 Photopolymer Materials

A photopolymer recording material consists of three parts: a photopolymerizable monomer, an initiator system and a polymer. When exposed, a part of the monomer is polymerized. This gives rise to diffusion of monomer molecules from the regions of high intensity to the regions of low intensity. The material is then exposed to light of uniform intensity until the remaining monomer is polymerized. A difference in the refractive index within the material is then obtained.

Photoplymer materials can be used for recording phase holograms, where applications in mass-production of display holograms and optical elements are of main interest. Companies such as AT&T Bell Laboratories, du Pont and Hughes have produced photopolymer materials for recording holograms. Advantages are a low noise level and its suitability for applying dry processing techniques. The sensitivity is about 10 mJ/cm^2.

Of the other materials for hologram recording, we mention dichromated gelatin, photoresist, photochromic materials and ferroelectric crystals.

5.6 PHOTOELECTRIC DETECTORS

Optical detectors can be classified as in the block diagram of Figure 5.20. Here we classify photographic film, photopolymers, etc. as chemical detectors and they are described in Section 5.5. They do not give a signal output in the usual sense as do the other types, termed electronic detectors, which are divided into two branches: thermal and photon detectors. In thermal detectors, the absorption of light raises the temperature of the device and this in turn results in changes in some temperature-dependent parameter (e.g. electrical conductivity). Most thermal detectors are rather inefficient and quite slow, and because

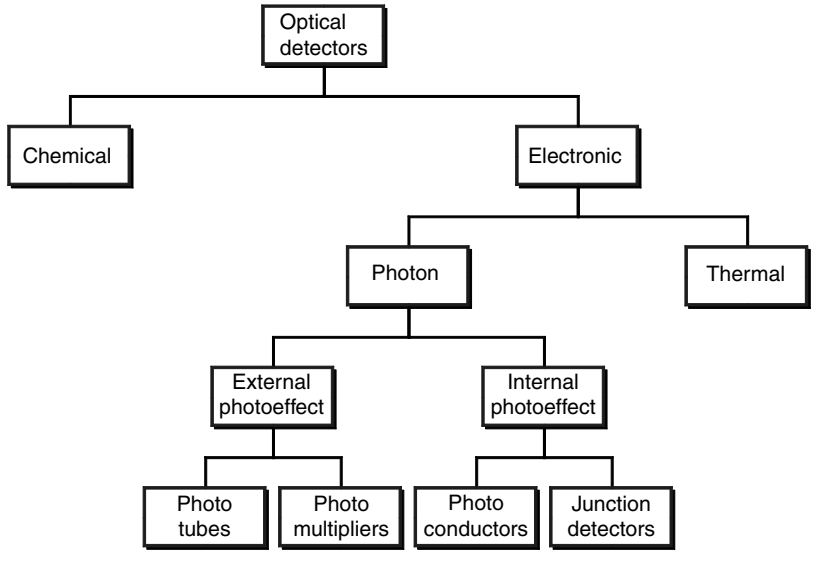

Figure 5.20 Optical detectors

of their relative unimportance in the field of optical metrology we name only some of the better known types – the thermocouple, the bolometer and pyroelectric detectors. The last type can be made with response times in the nanosecond region and with a wavelength response up to 100 μm. They have proved very useful as low cost, robust infrared detectors for fire detection and intruder alarms, for example.

The operation of photon (photoelectric) detectors is based on the photoeffect, in which the absorption of photons by some materials results directly in an electronic transition to a higher energy level and the generation of mobile charge carriers. Since the energy of a single photon is $E = h\nu = hc/\lambda$, photon detectors have a maximum wavelength beyond which they do not operate. A problem common to all photon detectors operated in the infrared is that the photon energy $h\nu$ becomes comparable with the average thermal energy ($\approx kT$) of the atoms in the detector itself. Therefore, most photon detectors operating above a wavelength of about 3 μm must be cooled to liquid nitrogen temperatures (77 K) or below.

The photoeffect takes two forms: external and internal. The former process involves photoelectric emission, in which the photo-generated electrons escape from the material (the photocathode) as free electrons with a maximum kinetic energy given by Einstein's photoelectric equation:

$$E_{\max} = h\nu - W$$

where the work function W is the energy difference between the vacuum and the Fermi levels of the material. Pure metals are rarely used as photocathodes since they have low quantum efficiencies ($\approx 0.1\%$) and high work functions ($W = 2.1$ eV for Cs) which makes them useful only in the visible and ultraviolet regions of the spectrum. However, semiconductors can operate with higher quantum efficiencies and lower work functions corresponding to wavelengths up to about 1.1 μm. Photoemissive devices usually take the form of vacuum tubes called phototubes. Electrons emitted from the photocathode travel to an electrode (the anode) which is kept at a higher electric potential. As a result, an electric current proportional to the photon flux incident on the cathode is created in the circuit. In a photomultiplier, the electrons are accelerated towards a series of electrodes (called dynodes) maintained at successively higher potentials. From the dynodes a cascade of electrons are emitted by secondary emission, resulting in an amplification by a factor as high as 10^7.

A microchannel plate consists of an array of channels (of internal diameter ≈ 10 μm) in a slab of insulating material (≈ 0.5 mm thick). Both faces of the plate are coated with thin metal films that act as electrodes, and the interior walls of each channel are made slightly conducting. Each channel thus acts like a miniature photomultiplier tube. On emerging from the channels, the electrons can generate light (and thereby an optical image) by striking a phosphor screen. The latter combination is called an image intensifier.

In the internal photoeffect, the photoexcited carriers (electrons and holes) remain within the material.

5.6.1 Photoconductors

Photoconductor detectors rely directly on the light-induced increase in the conductivity, an effect exhibited by almost all semiconductors (see Appendix E). The absorption of a photon results in the generation of a free electron excited from the valence band to

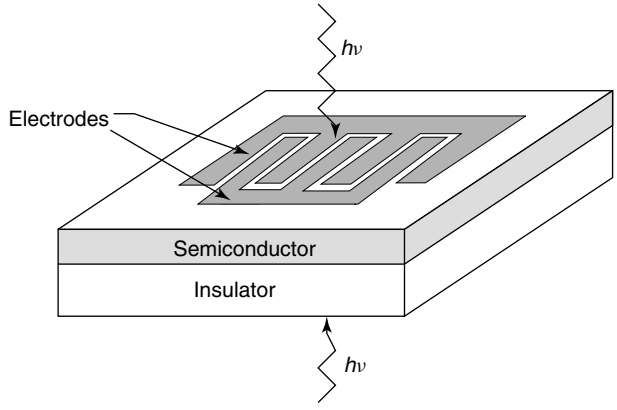

Figure 5.21 Photoconductor detector design

the conduction band, and a hole is generated in the valence band. An external voltage source connected to the material causes the electrons and holes to move, resulting in a detectable electric current. The detector operates by registering either the current (which is proportional to the photon flux) or the voltage drop across a series resistor. Unlike the quantum efficiency for the photoelectric effect, for example, the gain in a photoconductor may be larger than unity. The semiconducting material may take the form of a slab or a thin film. The contact electrodes are often placed on the same surface of the material in a geometry such as in Figure 5.21 to maximize the light transmission while minimizing the transit time. CdS and CdSe are both used for low cost visible radiation sensors in, for example, light meters for cameras. They usually have high gains ($10^3 - 10^4$) but poor response times ($\approx$50 ms). Other photoconductor materials for infrared detectors are PbS, InSb and HgCdTe.

5.6.2 Photodiodes

The photodiode detector is a p-n junction structure where photons absorbed in the depletion layer generate electrons and holes which are subject to the local electric field within that layer. Because of this field, the two carriers drift in opposite directions and an electric current is induced in the external circuit. Photodiodes have been fabricated from many of the semiconductor materials listed in Table E.2, as well as from ternary and quarternary compound semiconductors such as InGaAs and InGaAsP. Devices are often constructed in such a way that the light impinges normally on the p-n junction instead of parallel to it. A typical construction is seen in Figure 5.22. There are three classical modes of photodiode operation: open circuit (photovoltaic), short-circuit, and reverse biased (photoconductive). The usual i–V characteristic is seen in Figure E.2 (Appendix E). With increasing photon flux, the i–V characteristics move downwards as in Figure 5.23. In the photovoltaic mode, a voltage V_p is produced across the device that increases as a logarithmic function of the incident light irradiance. This mode is used, for example, in solar cells. In the photoconductive mode, a relatively large reverse bias ($\approx$10 V or more) is applied across the diode. Here the circuit current is directly proportional to the incident

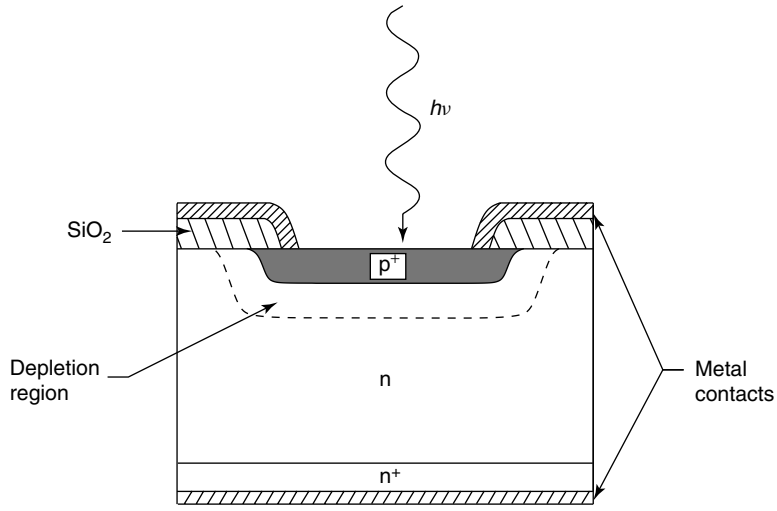

Figure 5.22 Typical silicon photodiode structure for photoconductive operation

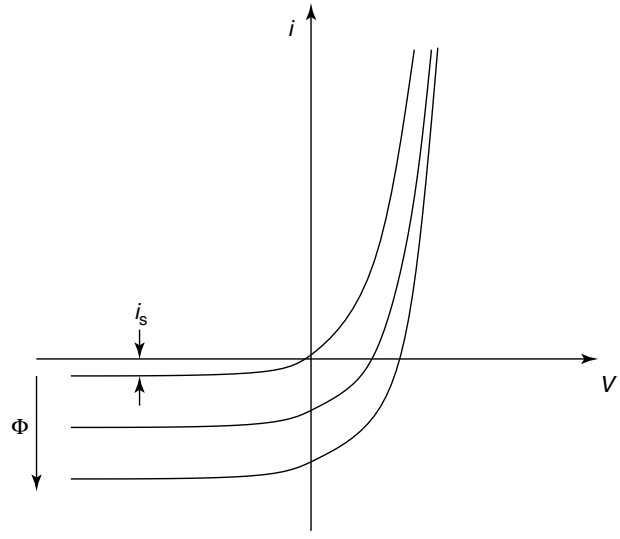

Figure 5.23 Current–voltage characteristics of a p-n junction under various levels of illumination

light irradiance. Other advantages of the photoconductive mode are faster response, better stability and greater dynamic range.

A strong reverse bias increases the width of the depletion layer, resulting in a larger photosensitive area, reduced junction capacitance and improved response time. A structure that results in a good long-wavelength response with modest reverse bias levels is the so-called pin (or PIN) structure. This is a p-n junction with an intrinsic layer sandwiched between the p and n layers. Here only a few volts of reverse bias are needed to cause the depletion layer to extend all the way through the n region.

By replacing the p-type (or n-type) layer in the p-n junction by a thin metallic film, we get a metal–semiconductor photodiode (also called a Schottky-barrier photodiode). There are a number of reasons why Schottky-barrier photodiodes are useful:

(1) Not all semiconductors can be prepared in both p-type and n-type forms.

(2) In p-n junctions one gets a substantial surface recombination and thereby a reduced efficiency. The metal–semiconductor junction has a depletion layer present immediately at the surface, thus eliminating this effect.

(3) The low resistance of the metal decreases the RC time constant thereby increasing the speed. Response times in the picosecond regime ($\approx$100 GHz bandwidths) are readily available.

Of particular interest is the Schottky-barrier photodiode of PtSi on p-type Si which is sensitive to wavelengths from the near ultraviolet to about 6 μm in the infrared (it must be cooled to 77 K). When used as elements in a CCD (see Section 5.7) one gets a device with multispectral imaging capabilities.

Finally it should be mentioned that with sufficiently large reverse bias, the electrons and holes may acquire sufficient energy to liberate more electrons and holes. Devices in which this internal amplification process occurs are known as avalanche photodiodes (APDs).

5.7 THE CCD CAMERA

5.7.1 Operating Principles

Until the mid 1960s, electronic devices for the pick-up of optical images were in the form of vacuum-type camera tubes. During the 1960s solid-state arrays with individual photoconductor elements connected to X–Y conductors and sequentially activated by voltages from thin-film shift registers were developed. The resulting images were however severely limited by response non-uniformities and other form of spatial noise associated with the X–Y readout techniques.

Workers at the Bell Telephone Laboratories (Boyle and Smith 1970) presented a new semiconductor device concept based on the manipulation of charge packets rather than the modulation of electric currents. Below we give a brief description of this concept. For a more thorough description, the article by Barbe and Campana (1977) is recommended.

A CCD is essentially a series of metal oxide semiconductor (MOS) capacitors. Figure 5.24 shows a simplified sketch of one of the capacitors. A semiconductor substrate of p-type silicon is covered with a thin layer of insulating silicon oxide which insulates the Si substrate from the metal electrode. When a positive voltage is applied between the electrode and the Si substrate, the minority carriers (holes in p-type Si) will be repelled from the interface between semiconductor and insulator, creating a region free of mobile carriers directly underneath the electrode. This region is known as the depletion region and has a thickness of a few micrometres. The metal electrodes (usually made of polycrystalline silicon) are transparent for wavelengths larger than about 400 nm.

If an incident photon has an energy larger than the bandgap in Si, it can create an electron-hole pair in the semiconductor. When this creation occurs in or near the depletion region, the photon generated electron is attracted towards the potential well which is

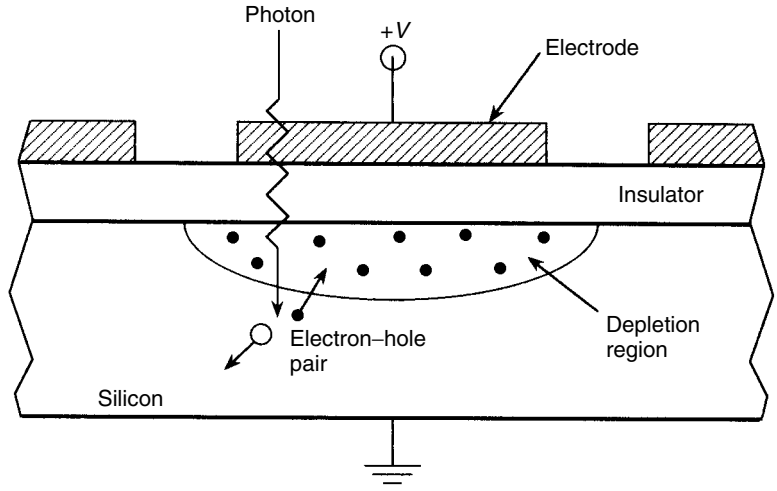

Figure 5.24 Simplified sketch of an MOS capacitor

formed under the positive charged electrode. In this way, a charge packet is formed consisting of photon-electrons which were created in the vicinity of a specific electrode.

Figure 5.25 shows how the charge packets are transferred from one electrode to the next by proper clocking of the potentials of the electrodes. In a device having a planar oxide and uniform substrate doping, at least three phases are required for unidirectional charge transfer, i.e. a barrier is maintained behind the charge packet while a deeper well is formed in front of the packet. The clocking diagrams are shown in Figure 5.25(b). At $t = t_1$, charge resides in the wells under the ϕ_1 electrodes. At $t = t_2$, the potential on ϕ_2 is made positive forming wells under the ϕ_2 electrodes. Charge will then flow from the ϕ_1 wells into the ϕ_2 wells. At $t = t_3$, the potential on the ϕ_1 electrodes is reduced to a low value, and the remaining charge in the ϕ_1 wells will be pushed into the ϕ_2 wells. This sequence repeats with the result that the charge configuration moves from one cell to the next every clock period.

To allow the device to be clocked with two phases, potential barriers between electrodes must be built in. This is done either by forming alternate thin and thick oxide insulators, the so-called stepped oxide barrier, or the implanted barrier technique by non-uniform doping of the substrate.

There are two different ways in which CCDs are organized when applied as an imaging sensor. In the following description standard TV rates are assumed, i.e. 1/25 s, European standard (CCIR) or 1/30 s, American standard (RS-170) frame time. We also assume two-phase clocking of the electrodes. Figure 5.26(a) shows the organization of the so-called frame-transfer structure. This sensor is divided into two identical areas, the image section and a masked storage section. A TV-frame is divided into two fields A and B, see Section 5.9. Field A is formed by collecting photoelectrons under the odd rows of electrodes for 1/50 s (1/60 s for RS-170). This charge configuration is shifted into the shielded storage register in a time that is short (several MHz) compared with the integration (exposure) time. Field A is then read out a line at a time while field B is being formed by collecting photoelectrons under the even electrodes.

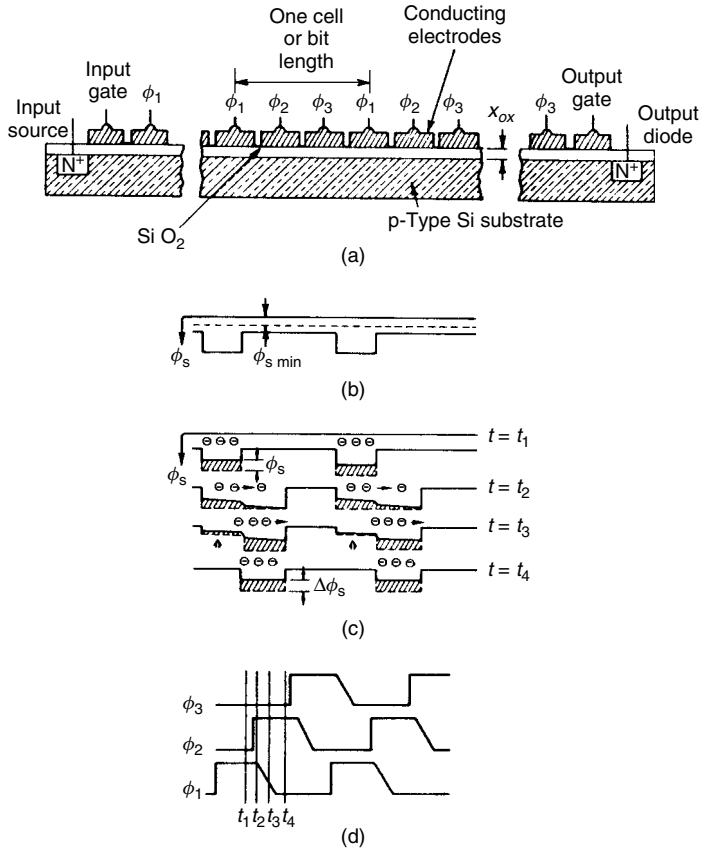

Figure 5.25 Three-phase CCD: (a) cross-sectional view showing input section, transfer section, and output section; a primitive electrode structure having unprotected gaps is shown for simplicity; (b) surface-potential profile showing potential wells under the ϕ_1 electrodes; (c) surface-potential profiles showing progression of charge transfer during one clock period; and (d) clocking waveforms used to drive the CCD during transfer. (From Barbe 1975. Reproduced by permission of IEEE)

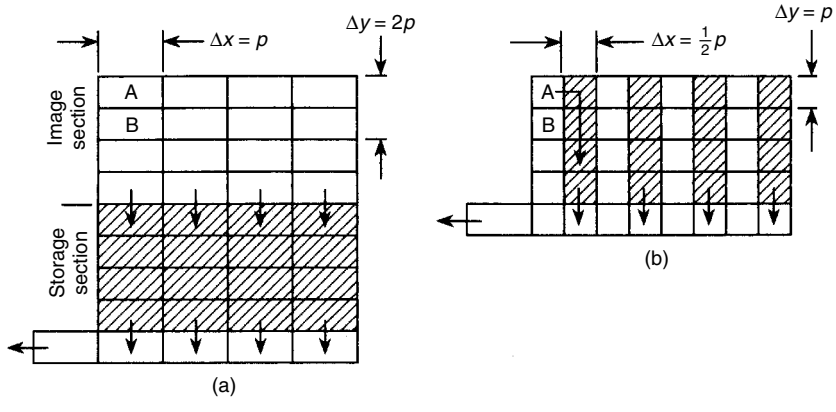

Figure 5.26 Schematic diagrams showing the (a) frame-transfer and (b) inter-line transfer array organizations

Figure 5.26(b) shows the organization of the so-called interline transfer structure. Here the shielded vertical readout registers are interdigitated with the photosensitive column. Potential wells are formed in the photosensitive regions by applying voltages to the vertical polycrystalline silicon (polysilicon) stripes. The horizontal polysilicon stripes are used to clock the vertical shielded register. Because the integrating cells and shift-out cells are separate, the effective integration time for both fields A and B is 1/25 s. The operation is as follows. After collecting photoelectrons in field A for 1/25 s, the charge configuration is shifted into the shielded registers and down, a line at a time, into the horizontal output register. When field A has been completely read out (1/50 s), field B is shifted into the shielded registers and out. Note that the effective integration time for the interline transfer structure is twice that of the frame-transfer structure because the integration in the interline transfer structure is performed in sites separate from the transport registers.

5.7.2 Responsitivity

A very important parameter for an imaging sensor used for optical metrology is its responsitivity R, particularly the responsitivity as a function of the spatial frequency $R(f)$ of the imaged scene. This can be formulated as

$$R(f) = R(0) \prod_i (MTF)_i \tag{5.44}$$

where $R(0)$ is the responsitivity at zero spatial frequency and the other term is the product of all of the modulation transfer function (MTF, see Section 4.6.2) factors that affect the frequency response of the chip. These factors are: (1) the loss of frequency response due to the geometry of the integrating cell (MTF_{integ}), (2) the loss of frequency response due to transfer inefficiency ($MTF_{transfer}$), and (3) the loss of frequency response due to the diffusion of charge between photon absorption and photoelectron collection (MTF_{diff}). We here will consider three chip designs: The front-illuminated interline transfer CCD (FIIT), the front illuminated frame transfer CCD (FIFT) and the back-illuminated frame transfer CCD (BIFT). For FIIT approximately one-half of the chip area is photosensitive because the other half is occupied by the vertical transport registers. When used in standard TV interlaced mode, however, the integration time for FIFT and BIFT is half of that for FIIT. The efficiency with which photons are absorbed and the resulting photoelectrons are collected in the integrating cells are twice as high for BIFT than for both FIIT and FIFT when averaged over the 0.4–1.0 μm wavelength band. In conclusion therefore, R(0) is twice as high for BIFT than for FIIT and FIFT.

In modern CCD chips, $MTF_{transfer}$ and MTF_{diff} have negligible effects compared to MTF_{integ}. The latter is, however, a fundamental effect on the chip responsitivity and is related to the finite size of the integrating cells.

To calculate MTF_{integ} it is sufficient to find the response to a sinusoidal grating of frequency f

$$H_1 = H_0[1 + m \cos 2\pi f x] \tag{5.45}$$

When sensed by an array of cells of width Δx and inter-cell distance p, the output charge pattern becomes

$$H_2 = \frac{1}{p}\int_{x-\Delta x/2}^{x+\Delta x/2} H_1\,\mathrm{d}x = \frac{H_0}{p}\int_{x-\Delta x/2}^{x+\Delta x/2}[1 + m\cos 2\pi f x]\,\mathrm{d}x$$
$$= \frac{H_0\Delta x}{p}\left[1 + m\frac{\sin\pi\Delta x}{\pi f\Delta x}\cos 2\pi f x\right] \tag{5.46}$$

Therefore the MTF of the integrating process is (cf. Section 4.6.2, Equation (4.56))

$$MTF_{\text{integ}} = \left|\frac{\sin\pi f\Delta x}{\pi f\Delta x}\right| \tag{5.47}$$

By introducing the so-called Nyquist frequency $f_{\text{n}} = 1/2p$, this can be written

$$MTF_{\text{integ}} = \left|\frac{\sin\dfrac{\pi}{2}\dfrac{f}{f_{\text{n}}}\dfrac{\Delta x}{p}}{\dfrac{\pi}{2}\dfrac{f}{f_{\text{n}}}\dfrac{\Delta x}{p}}\right| \tag{5.48}$$

Figure 5.27 shows the responsitivity as a function of spatial frequency for FIIT, FIFT and BIFT in the horizontal and vertical directions.

5.8 SAMPLING

5.8.1 Ideal Sampling

Consider a one-dimensional function $f(x)$ which might represent e.g. the irradiance distribution along a TV-line on a CCD camera. To sample this function means to find the values of f at regular intervals, i.e. $f(np)$ where $n = 0, 1, 2, \ldots$ and p is a constant called the sampling period (Goodman 1968). This is equivalent to multiplying f by a comb function (Equation (B.16), Appendix B.2) to get the sampled function f_{s},

$$f_{\text{s}} = \text{comb}\left(\frac{x}{p}\right) f(x) \tag{5.49}$$

The spectrum F_{s} of f_{s} is given by its Fourier transform

$$F_{\text{s}}(f_x) = \mathscr{F}\left\{\text{comb}\left(\frac{x}{p}\right) f(x)\right\} = \mathscr{F}\left\{\text{comb}\left(\frac{x}{p}\right)\right\} \otimes F(f_x) \tag{5.50}$$

where the last equality follows from the convolution theorem and $F(f_x)$ is the spectrum of $f(x)$. Now we have that

$$\mathscr{F}\left\{\text{comb}\left(\frac{x}{p}\right)\right\} = \text{pcomb}(pf_x) = \sum_{n=-\infty}^{\infty}\delta\left(f_x - \frac{n}{p}\right) \tag{5.51}$$

It follows that the spectrum of the sampled function is given by

$$F_s(f_x) = \sum_{n=-\infty}^{\infty} F\left(f_x - \frac{n}{p}\right) \tag{5.52}$$

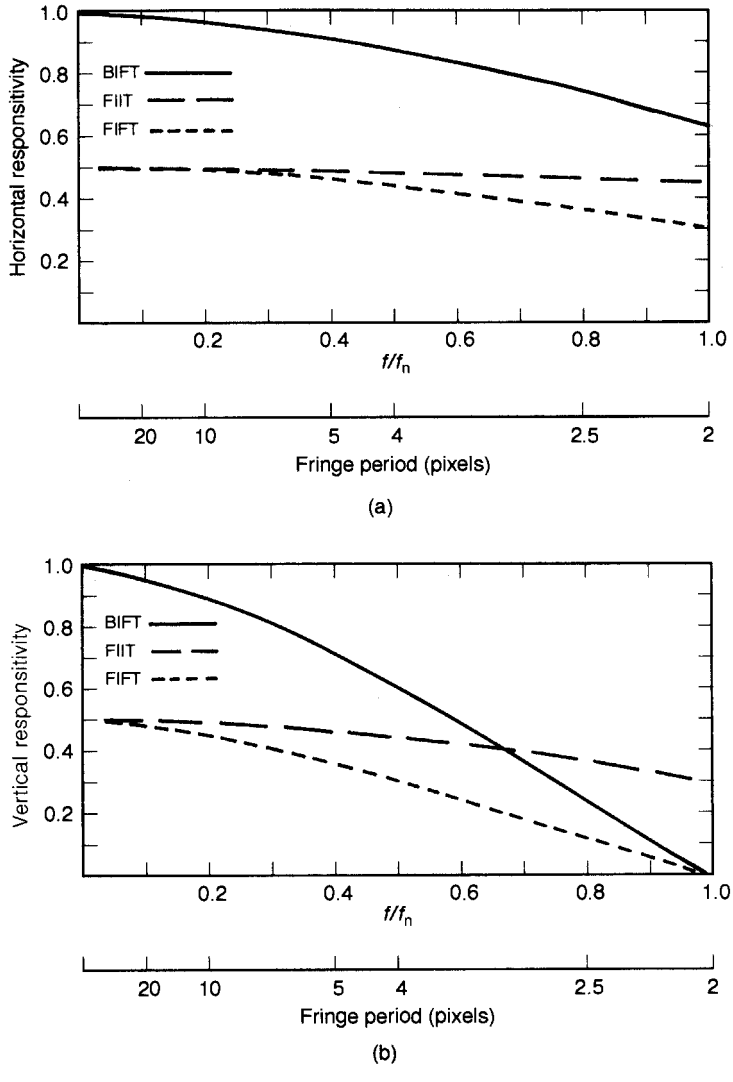

Figure 5.27 Responsitivity for back-illuminated FT (BIFT), front-illuminated FT (FIFT) and IT (FIIT) arrays: (a) versus normalized horizontal spatial frequency and fringe period and (b) versus normalized vertical spatial frequency and fringe period. The relations between Δx, Δy and p are as in Figure 5.26, i.e. standard video signal transfer is assumed

Evidently the spectrum of f_s can be found simply by erecting the spectrum of f about each point n/p along the f_x-axis as shown in Figure 5.28c).

Now assume (as in Figure 5.28(a)) that the spectrum F of f vanishes outside some interval $[-W, W]$. A function whose transform has this property for any finite value of W is called a band-limited function. From Figure 5.28(c) we see that if

$$\frac{1}{p} \geq 2W \tag{5.53}$$

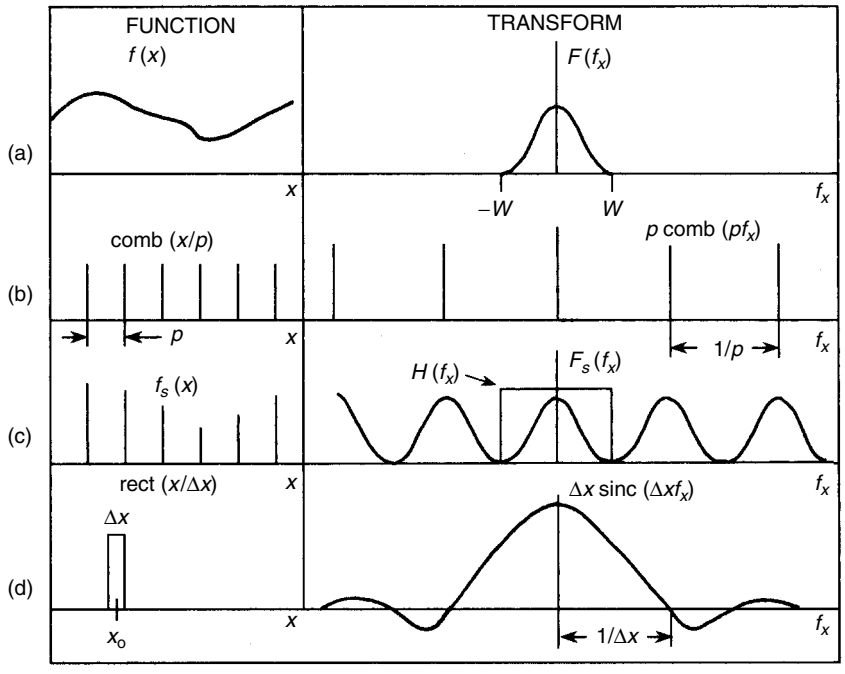

Figure 5.28 Sampling process in the x-domain and the spatial frequency domain

the spectra $F(f_x - n/p)$ constituting the spectrum of f_s do not overlap. If inequality (5.53) is fulfilled, we therefore can separate the $n = 0$ term of F_s from all the other terms by multiplying it by a filter function

$$H(f_x) = \text{rect}\left(\frac{f_x}{2W}\right) \tag{5.54}$$

We see therefore that F is recovered from F_s in that

$$F_s(f_x)\,\text{rect}\left(\frac{f_x}{2W}\right) \equiv F(f_x) \tag{5.55}$$

The inverse Fourier transform of Equation (5.55) yields

$$\begin{aligned}\mathscr{F}^{-1}\{F(f_x)\} = f(x) &= \mathscr{F}^{-1}\left\{F_s(f_x)\,\text{rect}\left(\frac{f_x}{2W}\right)\right\} \\ &= \mathscr{F}^{-1}\{F_s(f_x)\} \otimes \mathscr{F}^{-1}\left\{\text{rect}\left(\frac{f_x}{2W}\right)\right\} \\ &= f_s(x) \otimes 2W\,\text{sinc}\,(2Wx) = f(x)\text{comb}\left(\frac{x}{p}\right) \otimes 2W\,\text{sinc}\,(2Wx)\end{aligned} \tag{5.56}$$

Noting that

$$f(x)\text{comb}\left(\frac{x}{p}\right) = p\sum_{n=-\infty}^{\infty} f(np)\delta(x-np) \tag{5.57}$$

Equation (5.56) becomes

$$f(x) = 2pW\sum_{n=-\infty}^{\infty} f(np)\,\text{sinc}\,[2W(x-np)] \tag{5.58}$$

Finally, when the sampling interval p is taken to have its maximum allowable value $1/2W$, we have that

$$f(x) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2W}\right)\text{sinc}\,\left[2W\left(x-\frac{n}{2W}\right)\right] \tag{5.59}$$

Equation (5.59) represents a fundamental result which we refer to as the Whittaker–Shannon sampling theorem. It implies that exact recovery of a bandlimited function can be achieved from an appropriately spaced array of its sampled values, the recovery is accomplished by injecting, at each sample point, an interpolation function consisting of a sinc function.

It should be noted that other choices of the filter function $H(f_x)$ than that given in Equation (5.54) is possible as long as $H(f_x)$ passes the $n = 0$ term of F_s and excludes other terms. In fact it is a multitude of choices which will result in alternative sampling theorems.

5.8.2 Non-Ideal Sampling

The sampling of a function by discrete points is an idealized situation. In reality the sampling pulses always have finite width.

Consider Figure 5.28(d) with such a sampling pulse of width Δx centred at x_0. If this pulse represents the cell in a CCD chip, the irradiance $f(x)$ will be integrated over this cell. The charge at x_0 will therefore be given by

$$f_i(x_0) = \frac{1}{\Delta x}\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} f(x)\,\mathrm{d}x \tag{5.60}$$

If we introduce the rectangle function (Equation (B.13), Appendix B.2), this integral can be written as

$$f_i(x_0) = \frac{1}{\Delta x}\int_{-\infty}^{\infty} \text{rect}\,\left(\frac{x-x_0}{\Delta x}\right) f(x)\,\mathrm{d}x \tag{5.61}$$

Since rect $(-x)$ = rect (x) (it is symmetric), we have

$$f_i(x) = \frac{1}{\Delta x}\int_{-\infty}^{\infty} \text{rect}\,\left(\frac{x-\xi}{\Delta x}\right) f(\xi)\,\mathrm{d}\xi \tag{5.62}$$

where we also have changed variables. Equation (5.62) is recognized as a convolution integral, i.e.

$$f_i(x) = \frac{1}{\Delta x}\,\text{rect}\,\left(\frac{x}{\Delta x}\right) \otimes f(x) \tag{5.63}$$

This function is then sampled in the same way as for ideal sampling:

$$f_{sN} = f_i(x)\text{comb}\left(\frac{x}{p}\right) = \left[\frac{1}{\Delta x}\,\text{rect}\,\left(\frac{x}{\Delta x}\right) \otimes f(x)\right]\text{comb}\left(\frac{x}{p}\right) \tag{5.64}$$

The spectrum now becomes

$$\begin{aligned} F_{sN} &= \mathscr{F}\left\{\frac{1}{\Delta x}\,\text{rect}\,\left(\frac{x}{\Delta x}\right) \otimes f(x)\right\} \otimes \mathscr{F}\left\{\text{comb}\left(\frac{x}{p}\right)\right\} \\ &= \text{sinc}\,(\Delta x f_x)F(f_x) \otimes p\text{comb}(pf_x) = \sum_{n=-\infty}^{\infty} F_N\left(f_x - \frac{n}{p}\right) \end{aligned} \tag{5.65}$$

where

$$F_N = \text{sinc}\,(\Delta x f_x)F(f_x) \tag{5.66}$$

Apart from $F(f_x)$ being multiplied by a sinc function, this is the same result as for ideal sampling. This does not matter so much as long as f_x is well below $1/\Delta x$, the first zero of the sinc function: see Figure 5.28(d).

5.8.3 Aliasing

If inequality (5.53) is not fulfilled, the repeated spectra F_N will overlap each other as seen in Figure 5.29. Since natural scenes are not band limited, the spectra will always overlap unless $F(f_x)$ is prefiltered. Overlapping of the spectra causes frequencies higher than the Nyquist limit ($f_\text{n} = 2/p$) to appear in the passband ($-f_\text{n} \leq f_x \leq f_\text{n}$) as lower-frequency components – thus the term 'aliasing'. Thus, for example the frequency $1.5f_\text{n}$ in F_N for $n = 1$ would give a response in F_N for $n = 0$ at $0.5f_\text{n}$.

An example of aliasing is shown in Figure 5.30. Here vertical bars of different spacings are imaged onto a 100×100 element interline transfer CCD chip. The Nyquist frequency in this case was 12.3 cycles/mm. Thus only the top row of Figure 5.30 represents the true imagery. The remaining six views are moirè patterns produced by the interaction of the CCD structure and the bars of the test chart.

5.9 SIGNAL TRANSFER

Most electronic cameras are equipped with a video output signal. This is an analog signal containing the image data (Grob 1984). To guide the scanning beam of the TV-monitor, this video signal also contains some timing information, see Figure 5.31. The timing information is transmitted between each horizontal scan line and is called the horizontal

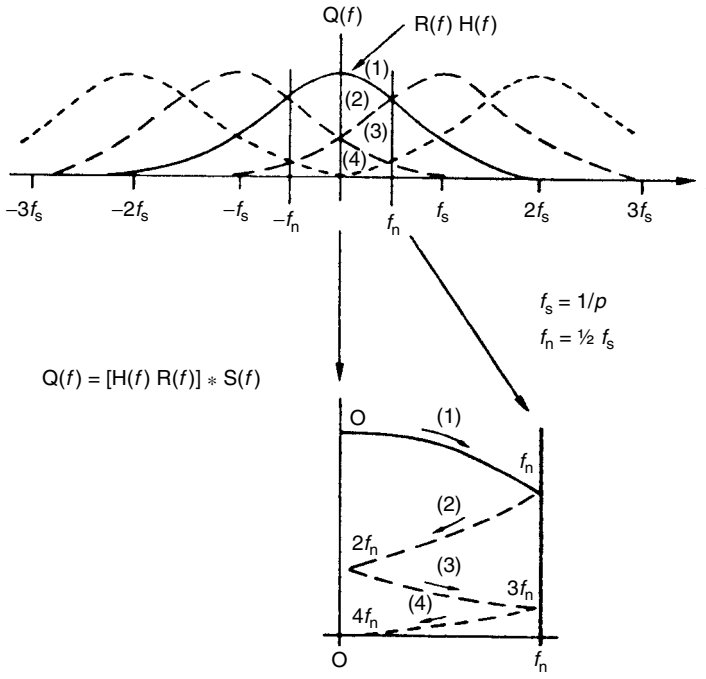

Figure 5.29 Spatial frequency domain analysis of the sampling process. Branch (1) is the desired response, and branches (2), (3) and (4) represent response to $f > f_n$ aliased into the passband. (From Barbe, D.F. and Campana, S.B., (1976) Aliasing and MTF effects in photosensor arrays, in P.G. Jespers, F.v.de Wiele and M.H. White (eds), Solid State Imaging, Noordhoff, Leiden. Reproduced by permission of D.F. Barbe)

blanking period. In this period the electron beam intensity is set to zero to avoid disturbing the picture during flyback to the next TV line. The horizontal blanking period consists of a horizontal (hsync) pulse which is used for synchronization and a back porch. In the AD converter of the frame grabber (see Section 10.2), this porch is used for adjusting the grey level to a known value (black). All video signals use interlaced raster technique. This means that the horizontal scan lines of a TV frame are divided into two fields. The even field consists of all the even-numbered lines in the frame, starting with line zero. The odd fields consists of the odd-numbered scan lines. This is done to avoid flicker on the TV screen. The even and odd fields are separated by a vertical blanking period.

A complete scanning pattern is shown in Figure 5.32, where the corresponding horizontal and vertical sawtooth waveforms illustrate odd-line interlaced scanning. A total of 21 lines in the frame is used for simplicity, instead of 525 (American standard). The 21 lines are interlaced with two fields per frame. Of the 10.5 lines in a field, we can assume that 1 line is scanned during vertical retrace to have a convenient vertical flyback time. So 9.5 lines are scanned during vertical trace in each field. Therefore in the entire frame, 19 lines are scanned during vertical trace, so 2 lines are lost in the vertical retrace lines. Starting at point A in the upper left corner, the beam scans the first line from left to right and retraces to the left to begin scanning the third line in the frame. Then the beam scans all succeeding odd lines until it reaches point B at the bottom, when vertical flyback

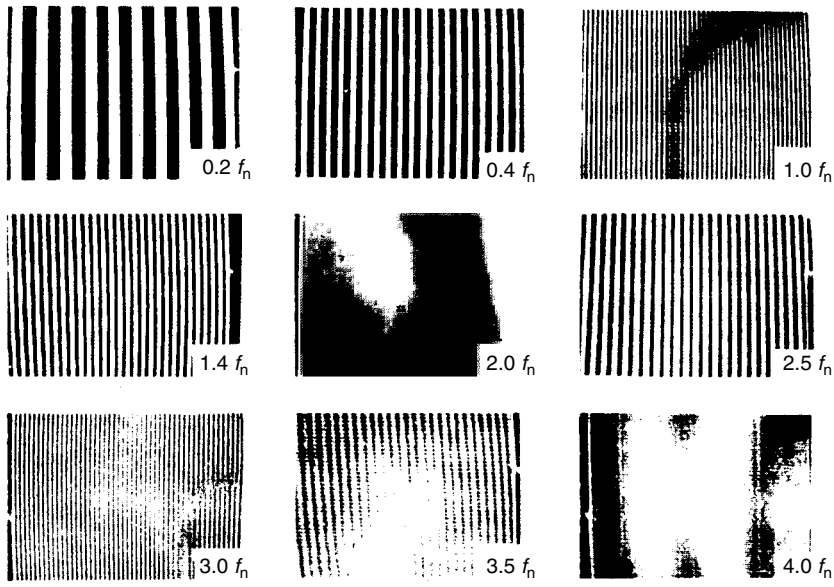

Figure 5.30 Bar-pattern imagery produced by the 100×100 element CCD array. The fundamental frequencies are given relative to the Nyquist frequency, f_n. (From Barbe, D.F. and Campana, S.B., (1976) Aliasing and MTF effects in photosensor arrays, in P.G. Jespers, F.v.de Wiele and M.H. White (eds), Solid State Imaging, Noordhoff, Leiden. Reproduced by permission of D.F. Barbe)

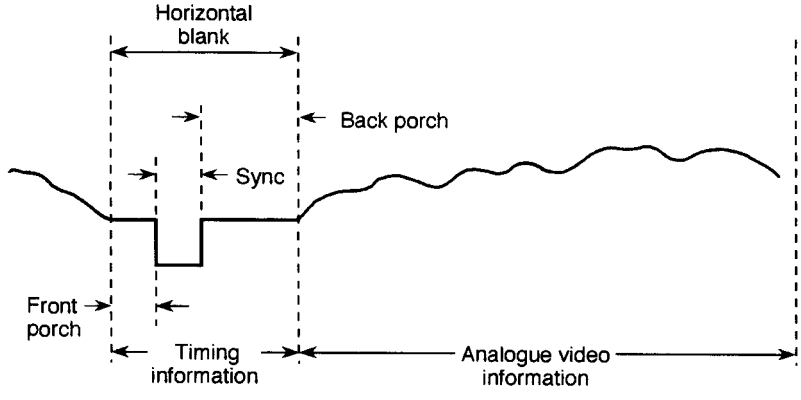

Figure 5.31 The analog video signal with timing information

begins. Note that this vertical retrace begins in the middle of a horizontal line. During this vertical retrace the scanning beam is brought to point C, which is separated from A by exactly one half-line so that the scanning of the second field can begin. In commercial TV broadcasting, some of the 'lost' lines in vertical retrace are digitally encoded to carry data for reproduction of full pages of alphanumeric characters for videotext. This requires a decoder at the receiver to gate out the specific lines and process the digital signal.

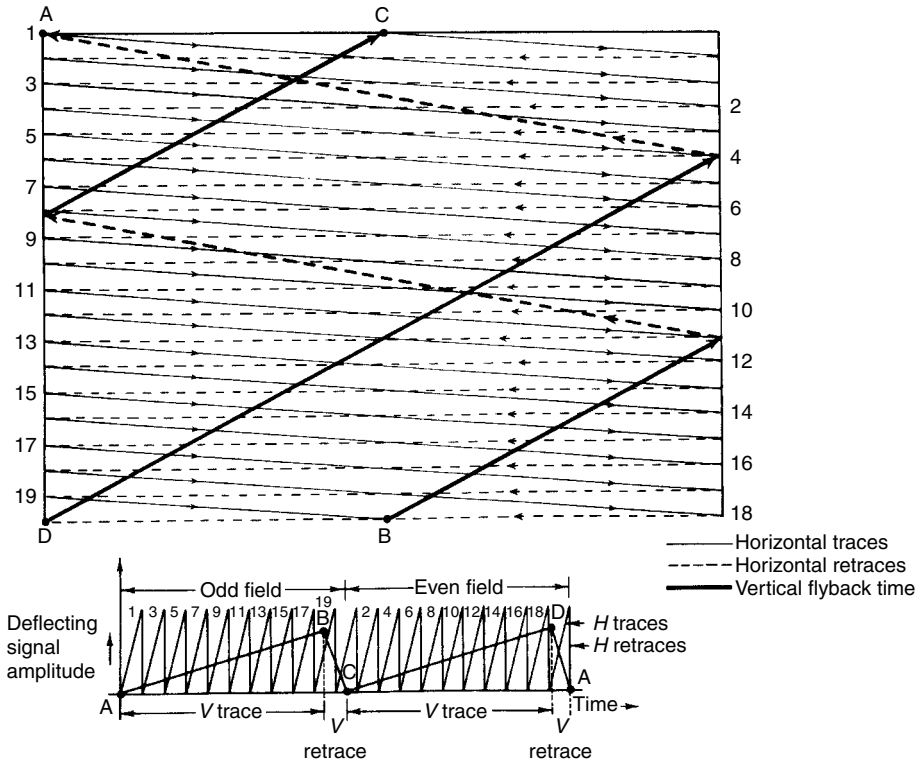

Figure 5.32 A sample scanning pattern for 21 interlaced lines per frame and $10\frac{1}{2}$ lines per field. The corresponding *H* and *V* sawtooth deflection waveforms are shown below pattern. Starting at point *A*, the scanning motion continues through *B*, *C* and *D* and back to *A* again. ((From Gro, B. (1984) Basic Television and Video Systems. (5th edn), McGraw-Hill, N.Y. Reproduced by permission of McGraw-Hill Inc., N.Y)

Table 5.3 Three different television standard video systems distinguished in acquisition speed, frame size and grey-value resolution

Pixel frequency	Slow scan <5 MHz	Standard video 7.5–15 MHz	HDTV >15 MHz
Frame size in pixels	1024 × 1024 4096 × 4096	256 × 256 512 × 512 780 × 540	1280 × 1024
Grey-value resolution	6–16 bit	6–10 bit	6–8 bit

Unfortunately, the video signal has several standards. Analogue input comes in three different forms, essentially distinguished by the corresponding 8-bit conversion frequency: slow-scan, standard video and high definition video (HDTV). Table 5.3 shows the different frequencies, frame grabber sizes and grey value resolutions.

The standard video norms for black and white are RS-170 (used in North and South America and Japan) and CCIR (used in Europe). A detailed overview of these systems is found in Table 5.4.

Table 5.4 Comparison of the black and white video systems RS-170 and CCIR

	RS-170	CCIR
Frame rate/field rate	30/60 Hz	25/50 Hz
Number of lines	525	625
Number of active lines	480	576
Field time	16 2/3 ms	20 ms
Time per line	63.49 μs	64 μs
Active line period	52.5 μs	52 μs
Nominal video bandwidth	4.5 MHz	5.5 MHz
Resolution	472	572
Aspect ratio	4 : 3	4 : 3

Standard video uses interlacing to avoid flicker in human perception of video images. This is not demanded in machine vision and video metrology, but will in many cases be a drawback, especially when monitoring high-speed phenomena. When using non-interlaced video, the resolution in the vertical direction of the frame transfer CCD cameras is also doubled, see Section 5.7.2. Specialized camera manufacturers such as EG & G, Fairchild and Dalsa offer matrix cameras with capabilities of up to 6000×6000 pixels for non-standard video transfer. The scan rate is driven by an external clock and can be selected by the operator. Such cameras are of course superior to standard video cameras with regard to resolution, but the amount of data and therefore the processing time increases dramatically, from $512 \times 512 \times 8 = 200$ kbyte to 288 Mbyte for a single image.

As mentioned in Section 10.2, for frame grabbers receiving composite video, line jitter is a problem especially when making measurements with sub-pixel accuracy. This problem is avoided when the signal is transferred digitally from the camera to the frame grabber and not via an analogue video signal. This is achieved by placing the A/D converter inside the camera and transferring the signal via a data cable. Such cameras are manufactured by Kodak (Videk Megaplus) and Cohu. Most frame grabbers today accept such signals.

PROBLEMS

5.1 Assume the Sun to be a 6000 K blackbody source and that its diameter subtends an angle $\alpha = 9.3$ mrad at the Earth.

(a) Find the wavelength λ_m corresponding to the maximum solar spectral radiant exitance M_λ.

(b) Find M_λ of the Sun's surface at this wavelength.

(c) What is the spectral radiance L?
The area of the solar disc dA is given by

$$dA = \pi \alpha^2 s^2 / 4$$

where s is the Earth-Sun distance.