IVAR '94 Tutorial
Image Acquisition and Display

Patrick Wambacq

Image processing and analysis has rapidly become a widely used tool in industry. It is used routinely in visual inspection and robot vision applications. Other applications include medical image analysis, remote sensing, security, image compression, ... Besides its use in many industrial applications, it is still an area of great research effort.

All these applications need hardware and software to deal with image data. Images need to be acquired, processed, displayed, etc., with the right equipment. This tutorial addresses the integration of these components to form a working system. The topics that are discussed in this text are:

- image sensors: which sensors exist and what are their characteristics;
- the video signal;
- displays: types and characteristics;

1 Overview of image sensors

Light sensors are based on two different underlying physical processes, giving rise to two classes of detectors: thermal and photon.

With thermal detectors, the incident power heats up the sensor to a temperature above the surrounding temperature, thereby responding equally to all wavelengths. This type of detector is not used in cameras that are normally found in image processing systems.

Photon detectors are devices in which incident photons cause electrons to be excited from the valence band to the conduction band thereby causing a measurable electrical effect. These detectors have a narrow wavelength response and normally have higher responsivities than thermal detectors.

Based on the principle of photon detectors, two families of cameras are actually used: vacuum tube cameras and solid state cameras.

1.1 Vacuum tube cameras

Briefly, a vacuum tube camera is a photoconductive device which employs a photosensitive sensor layer consisting of several million mosaic cells insulated from one another on a transparent metal film (refer to Figure 1.1). Each cell represents a small capacitor whose charge is a function of incident light. The sensor layer is scanned in a raster format with an electron beam over 625 lines in accordance with the television standard (discussed in chapter 3). This beam is deflected magnetically by a set of coils outside the tube bulb. The electron beam makes up charge lost through the incidence of light in individual mosaic cells and so generates the video signal at the sensor element. This video signal is simply a continuous analog signal proportional to the light intensity of the focused image. The camera electronics insert synchronization pulses to indicate scan lines, fields and frame ends (see chapter 3).

A distinction is made between the following camera types depending on the sensor element: the standard vidicon has a sensor element comprised of antimony sulphide (Sb_2S_3), the silicon diode vidicon has a sensor element made from silicon (Si), while the plumbicon has a sensor element made of lead oxide (PbO).

For most industrial applications, vacuum tube cameras are no longer used and are being replaced by solid-state cameras. The reasons are that solid-state cameras are more robust, are lighter, more sensitive, can achieve higher resolutions, and are often cheaper.

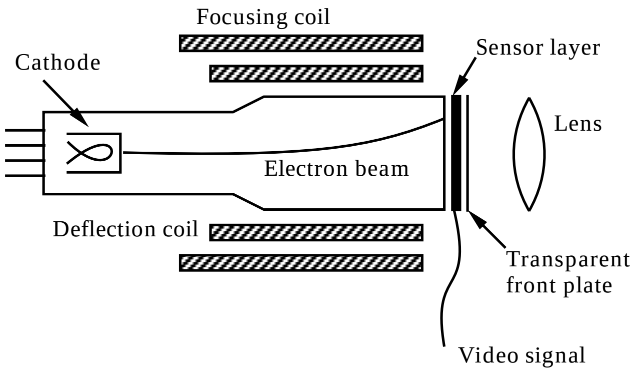

Figure 1.1: Pickup tube of a vidicon television camera.

1.2 Solid-state cameras

Most solid-state cameras are based on charge-coupled device (CCD) technology, though there are several variations on the theme. In order for the reader to be at least acquainted with the names of these devices, they are listed here:

- charge transfer devices (CTD)
- single transfer devices (STD)
- bucket brigade devices (BBD)
- charge coupled devices (CCD)
- charge injection devices (CID)
- surface charge coupled devices (SCCD)
- bulk charge coupled devices (BCCD)

The CCD technology is the most widespread, the basic structure of which is that of an analog shift register consisting of a series of closely spaced capacitors. Charge integration (accumulation) by the capacitors, photosites, caused by the photons comprising the incident light, provides the analog representation of light intensity. At the end of the integration period (exposure time) these charges are read out of the sensor.

CCD sensors most commonly use one of three addressing strategies: interline transfer, frame transfer, and column-row transfer.

The interline transfer CCD is organized into column pairs of devices. An imaging column of photosensors is adjacent to an opaque vertical shift register (see Figure 1.2). Charge accumulates in the imaging column until the end of the integration period, when it is transferred to the opaque column. The signal then shifts vertically into a horizontal shift register that represents the picture sequentially, line by line. The advantage

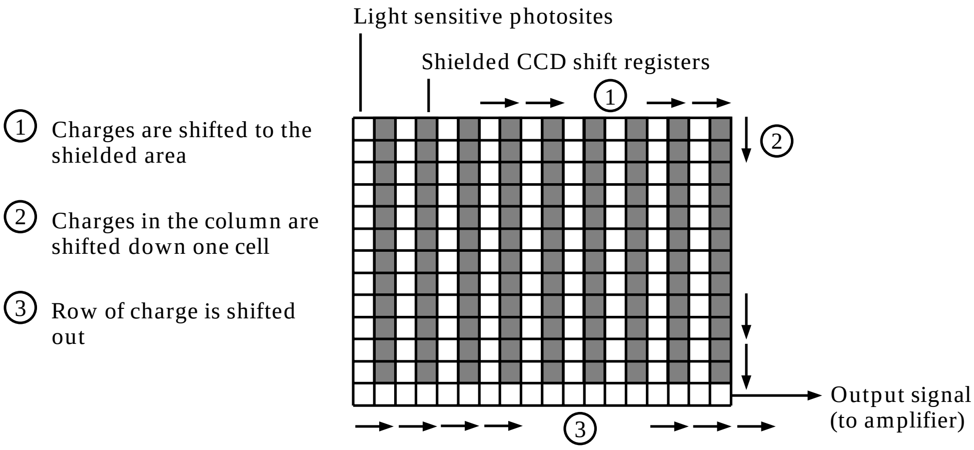

Figure 1.2: Interline transfer of charge in CCD sensors.

of the interline transfer is that the transfer time (to opaque storage) is short compared to the integration period. This is desirable because when transfer time approaches the integration time, solid-state sensors tend to exhibit a locally contained spreading of the image response, called smear. Thus, the interline transfer minimizes smear.

In the frame transfer organization (refer to Figure 1.3) the sensor consists of vertical columns of CCD shift registers divided into two zones. One zone, where charge accumulates during integration time, is photosensitive. When integration is complete, the whole array is transferred in parallel to the opaque storage area of the second zone.

A third type of solid-state sensor employs x-y addressing to transfer charge from the photosite to the output signal amplifier. The sensor elements are addressed by selecting individual column and row electrodes. Charge collected under the column electrode is transferred to the row electrode and amplified for output.

All devices discussed so far are called “area sensors”, i.e. the sensor consists of a two-dimensional array of photosites. However, there is another important class of solid-state sensor (and camera): this is the linear array sensor or line-scan sensor. In effect, these sensors are simply a one-dimensional array (row) of photosites, and use exactly the same technology as the two-dimensional array sensors. They differ in two important characteristics, however. Firstly, these sensors can have between 256 and 4096 photosites in a row and, hence, can achieve much greater resolution than state-of-the-art array cameras. Since they are inherently one-dimensional devices they can only take pictures of slices of a two-dimensional scene, and if a two-dimensional image is required, several such slices must be acquired. Thus, these sensors are best suited to the inspection applications in which the scene to be scanned is in continuous linear motion (or, indeed, where the camera itself can be translated). The second point to notice about these sensors is that the video signal that they generate does not correspond to any particular video standard and what is produced is essentially a time-varying analog

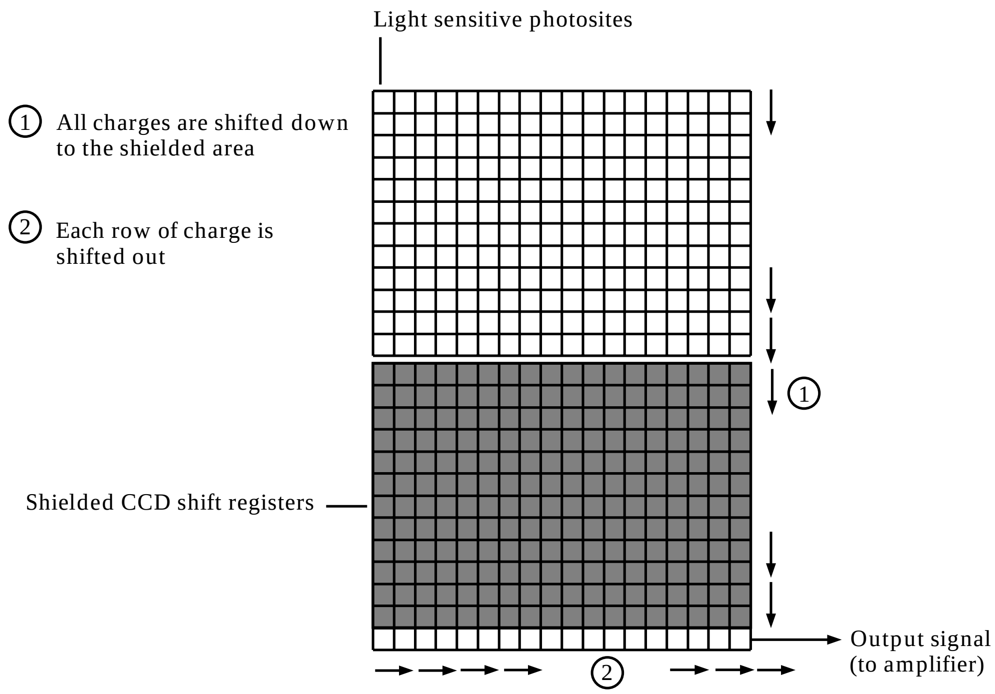

Figure 1.3: Frame transfer of charge in CCD sensors.

voltage which represents the incident light along the line of photosites. The repercussion of this characteristic is that most systems which use a line-scan camera tend to have custom-designed computer interfaces, although it is worth noting that matched linescan cameras and digitizers and, indeed, line-scan digitizers are now commercially available.

1.3 Colour cameras

In order to obtain colour images, three separate signals (red, green, blue) have to be generated. This can be done with vacuum tubes as well as with solid state sensors.

In colour cameras with vacuum tubes, three separate tubes are used, each generating one of the three desired signals. A typical setup is shown in fig 1.4. A compact dichroic prism arrangement, situated between the lens and the tube faces, is used to split the beam into three. Colour filters are added to improve the approximation to the required spectral sensitivity curves, and neutral density filters may also be added to equalize the signals produced by white. After passing the filters, the three beams are converted to an electrical signal by the vacuum tubes. The images formed on the three tubes must be geometrically identical and in exact registration with respect to the electron scanning, otherwise poor definition and colour fringing will occur. Very pre-

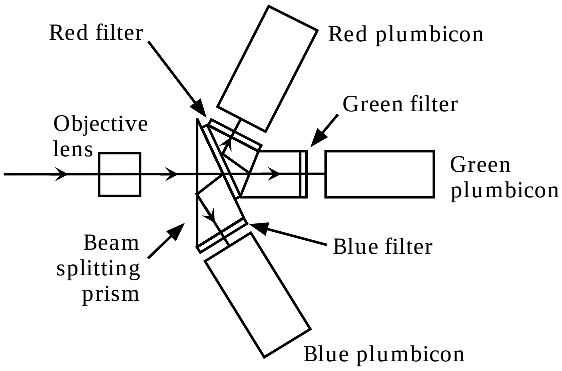

Figure 1.4: A colour camera using three plumbicons.

R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B

Figure 1.5: Example of a filter pattern for a CCD area sensor.

cise optical and electronic components are therefore required and the optics have to be mounted very rigidly. This makes these cameras fairly expensive, and therefore they are not normally used in industrial image processing applications. Their main use is broadcast television.

In the case of solid-state colour cameras, two approaches can be used. In a first approach, similar to using three vacuum tubes, three different CCD chips are used together with filters and a beam splitter. The three images must be registered exactly, but because CCD chips are much smaller than vacuum tubes, this is more easily obtained, and the resulting cameras are pretty compact.

Another approach is to have the three channels generated by a single CCD chip. In this case, three light-sensitive arrays of individual photosites, are merged onto a single device in an interleaved fashion. This has of course many advantages with respect to robustness and compactness. No expensive optics are required, and the colour filters are deposited on the photosites in a checkerboard pattern of red, green, and blue transparent material. One such pattern is shown in fig. 1.5. In the pattern shown in this figure, there are twice as many green areas as red or blue, and this is because the green signal contributes more to the luminance signal than the red or blue signals.

2 Characteristics and properties of image sensors

The most important properties of the image sensors will now be examined with respect to their industrial application. In particular the following characteristics will be considered in detail: resolution (definition), geometrical faults, sensitivity, transfer linearity, lag, spectral sensitivity, blooming, and noise. The classification of these characteristics makes it possible for the reader to understand the concepts appearing in the various data sheets and to assess their importance. At the same time the quantitative data which are scattered throughout the literature serve as a useful guideline.

2.1 Resolution and Modulation Transfer Function (MTF)

The resolution of an image sensor can be defined as the number of picture elements which can be discriminated. The resolution is limited by the number of light sensitive elements in the sensor since this defines the frequency with which the optical image is sampled. In fact, the number of light sensitive elements places an upper bound on the camera MTF (by the Nyquist sampling theorem), and it does not describe the camera's actual performance when observing high spatial frequencies.

Tube camera resolution is a function of the electron-beam diameter relative to the area of the photoconductive layer. Tube camera resolution easily outstrips the limitations imposed by the CCIR standard.

As discussed in chapter 3, the line frequency for the CCIR standard video signal is 15625 Hz. In addition, the nominal bandwidth of the signal, as specified in the standard, is 5 MHz, meaning that a signal can transmit a video image with five million periodic variations in the signal (brightness) levels. This results in an absolute maximum of $5 \times 10^6 \div 15625 = 320$ periodic (or sinusoidal) variations per line, that is, the maximum spatial frequency which can be faithfully represented by a video signal is 320 cycles per line.

The major theoretical tool for describing an optical system's performance is the Modulation Transfer Function (MTF), a measure of its spatial frequency response. A camera's response is the ratio of the contrast in the camera output to the contrast in the image. The function is usually normalized to make the MTF equal to one at low spatial frequencies. A camera's MTF decreases at high spatial frequencies because the pixels average and because their output is sampled. The detailed nature of the MTF curve allows prediction of a camera's ability to detect small features. Fig. 2.1 explains

Image sensor	Sensitivity μA / lm	Gamma
Sb_2S_3 vidicon	40 – 1200	0.6
PbO plumbicon	300 – 400	0.95
Silicon vidicon	4000	1
CCD camera	3500	1

Table 2.1: Sensitivity and γ for image sensors.

graphically the notion of MTF.

The pixel impulse response is another measure of resolution, containing the same information as the MTF, since the knowledge of one of both, allows the calculation of the other.

2.2 Geometrical faults

For television cameras with electron beam scanning, deviations in the constancy of vertical and horizontal deflection show up as faults in the geometrical assignment of the picture content. Standard industrial cameras are not designed as measuring cameras but to generate a picture for human examination. They therefore exhibit relatively large geometrical faults. For a standard industrial television camera this is usually $\pm 1\%$ to $\pm 2\%$ of the picture frame. With cheap cameras this fault can easily be larger.

With CCD television cameras there are no geometrical faults due to electron beam scanning; any geometric distortion is due to the lens.

2.3 Sensitivity and transfer linearity

The input signal of an image sensor is a brightness distribution. The output signal is a current or voltage proportional to this brightness. The sensitivity is defined as the ratio of the output magnitude to the input magnitude (dimension e.g. μA / lumen). Quite generally the following applies:

$$\text{Output magnitude} = (\text{Input magnitude})^{\gamma}$$

where gamma, the exponent of the transfer function, is given by

$$\frac{log(\text{output magnitude})}{log(\text{input magnitude})}$$

For a linear detector $\gamma = 1$. The exponent for different image sensor systems is specified in table 2.1 together with the sensitivity. These are typical values.

As a rule the standard Sb_2S_3 vidicon has roughly the same sensitivity as the plumbicon. The semiconductor cameras are however far superior to both camera types in terms of sensitivity.

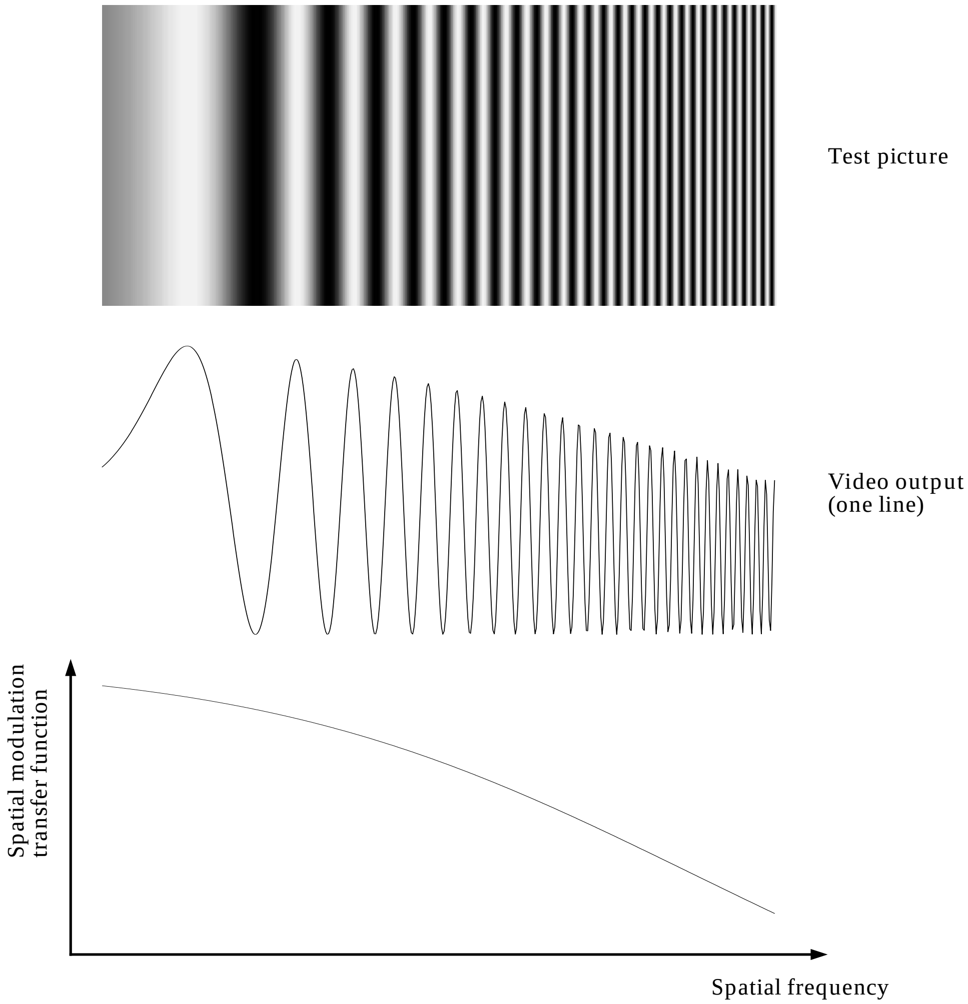

Figure 2.1: Illustration of the resolution by the MTF curve.

Image sensor	Lag
Sb_2S_3 vidicon	20%
PbO plumbicon	2%–5%
Silicon vidicon	8%
CCD camera	1%

Table 2.2: Lag for image sensors (elapsed time = 60ms).

2.4 Lag

Lag is defined as the percentage of the signal current at a certain point of the target after the illumination has been switched off. Typical values for an elapsed time of 60 ms are shown in table 2.2.

The lag means a limitation in the practical industrial application of television cameras in terms of the permissible speed of movement of an object under consideration. This case is most frequently encountered with parts being transported on a conveyor belt.

If in the simplest case the monitor shows up an impermissible movement blur due to the lag effect it is possible to illuminate the object briefly with a flash. In this way the lag is specifically exploited in order to store the picture after the flash for the duration of the picture processing. In the same way it is possible to carry out the short-term scanning of the object by means of a rotating aperture in front of the camera lens.

The problem of movement blur is absent if a sensor line is used for the moving object. With the help of the sensor line the object is very rapidly scanned in the coordinate vertical to the movement with a frequency which ranges from 100 KHz to a few MHz. Because of the movement of the object itself under the sensor line there is also no need for spatial scanning by the sensor. Thus a one-dimensional line is sufficient to process a two-dimensional picture.

2.5 Spectral sensitivity

The spectral sensitivity of an image sensor system is defined as the variation of the output as a function of the wavelength of the incident light. Normally the spectral sensitivity is of special interest in the visible region. However it can be just as important to give preference to a spectral region outside the visible region. An example is the observation of red-hot parts when especially high sensitivity in the near infra-red is an advantage. Fig. 2.2 shows the spectral sensitivity for different image sensor systems as well as for the human eye.

It can be seen that the semiconductor cameras are distinguished by good sensitivity in the invisible near infrared region.

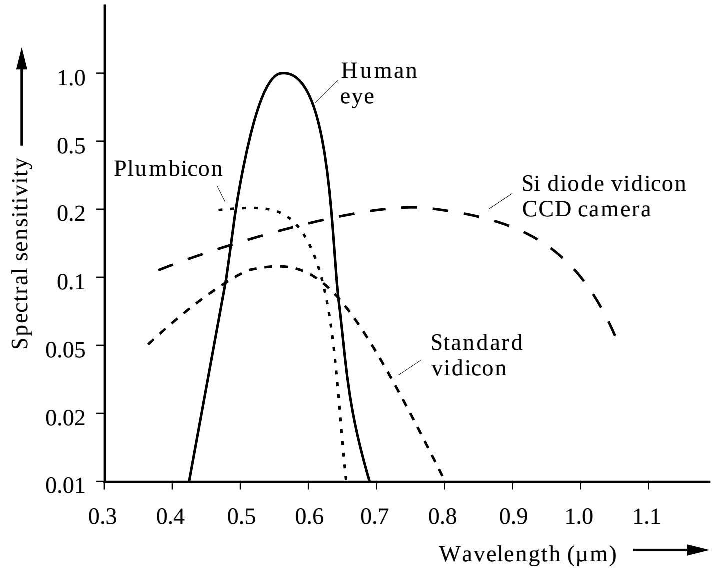

Figure 2.2: Spectral sensitivity of different image sensors in comparison to the human eye.

2.6 Blooming

If the target of a television camera is subjected to intensive brightness then the excess charge carriers spread into neighboring zones and introduce the bright area there. This effect is called blooming.

It is especially important that the target is not destroyed by overexposure. Semiconductor cameras have distinct advantages here. Whereas with standard vidicon and plumbicon a local overexposure destroys the affected point of the target, semiconductor cameras have extreme resistance to exposure.

2.7 Noise

There are a number of effects lumped under the term “noise”; they all degrade the accuracy of extracting scene luminance from pixel buffer values. It is important to select cameras with known noise characteristics and to screen and characterize cameras for their particular noise signatures. In the following discussion we will mainly concentrate on the noise phenomenon in solid-state cameras.

One cause of spatial noise is “bad pixels”, those with transfer functions so far outside the average as to make their information unusable in most machine vision applications. Since no useful information results from these pixels, the goal is to minimize their number and to note their locations so that the image processing software will not be misled by their output values.

Another type of spatial noise is fixed-pattern offset non-uniformity. This noise shows up as a fixed-pattern noise which is added to each pixel's output. A third type of spatial noise is fixed-pattern gain variation. It is caused by some pixels being more sensitive than others.

In addition to the above mentioned spatial noise, there is also temporal noise, frame-to-frame image variations within the same pixel, caused among others by statistical fluctuations in photon arrival, and electronic noise in the camera and lighting.

Noise is defined quantitatively by the signal-to-noise ratio (S / N), i.e. the ratio of the amplitude of the wanted signal to the average amplitude of the interference. For television cameras the signal-to-noise ratio is defined as the peak-to-peak video signal divided by the effective amplitude of the noise.

Following general practice in telecommunications and expressing the ratio of electrical variables of the same unit logarithmically, the level is obtained in Decibels (abbreviated to dB). The following reference points can serve as practical guidelines. A signal-to-noise ratio of 20 dB means that the picture quality is very bad — there is inadequate detail and resolution. For satisfactory picture quality S / N must be 40 dB and more.

The noise level puts a limit on the useful number of quantization levels that are obtained by an analog to digital conversion in a frame grabber. Generally, these frame grabbers are equipped with 8 bit converters, so that 256 different levels can be distinguished. If we want the noise level not to decrease the actual number of usable levels, the S / N ratio should be better than $20\log_{10}(256) = 48.2\text{dB}$.

3 The video signal

3.1 The CCIR standard

Most video sensors applied to vision have output signals which obey either the American EIA standard or the mostly European CCIR-625 standard. In this text, the given figures apply to the CCIR-625 standard, unless otherwise noted. The standards apply equally to cameras and displays and originate from television broadcasting. In order to provide the desired viewing quality, the physiological characteristics of the eye have to be considered. In particular, the discrete nature, both spatial and temporal, of displays should be invisible to the viewer. This has led to an image field with a 4:3 aspect ratio, consisting of 625 lines, 575 of which lie in the visible area. In order to avoid flicker, a sufficiently high image frequency has to be used, in the order of 50 images per second (depending on image contrast and brightness). At the time the standard was proposed, it was not economically feasible with the available technology to transmit 625 lines every 1 / 50 th of a second. Therefore, images are split into even and odd fields, containing the even and odd numbered lines respectively. Instead of transmitting all lines one after another from top to bottom, first the odd numbered lines are transmitted, followed by the even numbered lines, see also Figure 3.1. This results in a field frequency of 50 Hz, allowing an image frequency of only 25 Hz. This is called interlacing.

Since there are 625 lines in an image that is transmitted at a frequency of 25 Hz, the line frequency equals $625 \times 25 = 15625$ Hz. This corresponds to a time interval of 64μs, 12μs of which is blanked and needed to allow the electron beam to fly back from the right edge to the left edge of the image. Similarly, at the end of each field, a flyback time interval is required of 1.6ms, corresponding to 25 lines. Since there are two fields, the total number of lines during vertical blanking is 50, accounting for the difference between the total number and the visible number of lines.

Assuming that the resolutions in vertical and horizontal direction are the same (meaning that pixels are square), the number of pixels on a single line equals

$$\frac{4}{3} \times 575 = 767$$

The time interval devoted to scanning one pixel equals

$$\frac{52\mu s}{767} = 0.0684\mu s$$

and therefore the highest possible frequency in the video signal (i.e. one cycle per two adjacent pixels) amounts to $2 \times 0.068\mu s = 0.136\mu s$ or 7.4 MHz. This is however the-

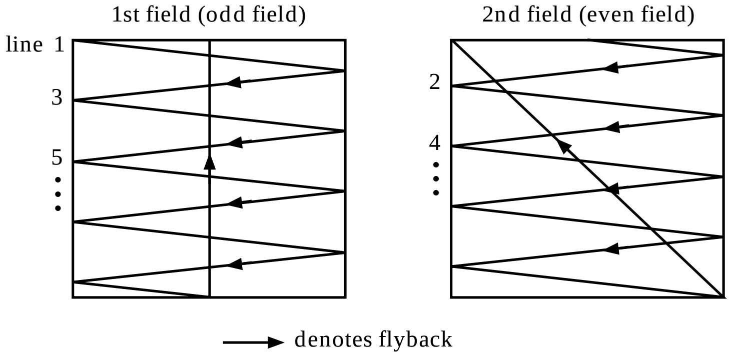

Figure 3.1: Interlacing: the division of the raster in two fields.

ory; in practice, the resolution is not that good (caused by the finite scanning beam diameter), and in the CCIR standard the video signal bandwidth is limited to 5 MHz.

The video signal is shown in figure 3.2. At the top of the figure, a single video line is shown. It consists of the actual luminance or intensity information part, and of a synchronization and blanking part. In the luminance part, the voltage level ranges between the black level and the white level, corresponding to a totally black or maximally white pixel on the screen respectively. This luminance part lasts for 52 μs. The remaining part, lasting for 12 μs, is devoted to synchronization and blanking. The voltage level is always below the black level, so that this part of the signal is invisible. It incorporates the time needed for the electron beam to fly back from the right to the left. In order to obtain a stationary image on a TV monitor, in synchronization with the camera and with the same raster, sync pulses are generated at the camera side and transmitted. These sync pulses indicate the beginning of one line and the end of another.

The bottom of figure 3.2 shows the vertical synchronization and blanking interval. There are several more detailed features that are not shown in the figure, but they would lead us to far and were left out for the sake of clarity. The signal marked odd field shows how the first or odd field ends with only half of a normal video line, followed by a vertical sync pulse and blanking interval, followed by the beginning of the second or even field with only half of a normal video line. The signal marked even field shows how the even field ends with a full video line, followed by a vertical sync and blanking interval, followed by the beginning of the odd field with a full video line. During the vertical sync pulse, which lasts for 160 μs, horizontal sync pulses are still present to prevent loss of synchronization (not shown in the figure).

3.2 The colour video signal

Since colour images consist of three components, the most simple way of dealing with a colour image is to transmit three separate signals, one for each of the signals red,

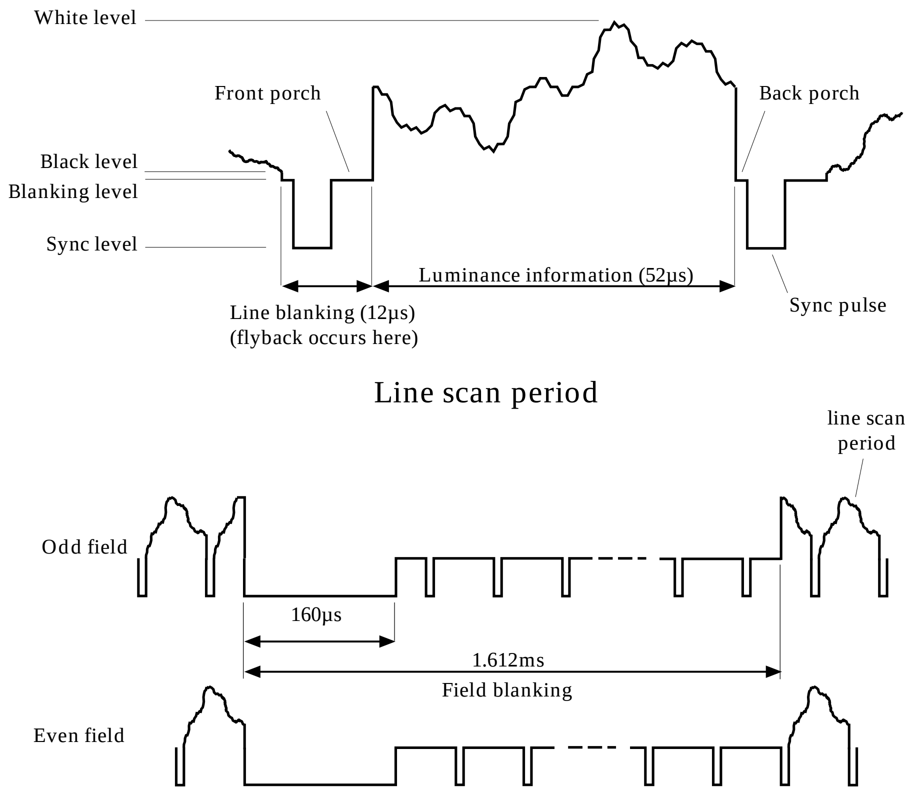

Figure 3.2: The CCIR video signal.

green and blue. This is in fact what happens in many closed-circuit installations. The red, green and blue camera signals are not converted into one signal, but instead directly coupled to monitors and frame grabbers using three separate wires. The signals themselves look then exactly as discussed in the previous section. By convention, synchronization is mostly derived from the green video channel, although in principle, the other channels are also suited for that purpose.

When colour images have to be broadcasted or recorded, the colour information is mostly encoded using a standard encoding such as PAL, SECAM or NTSC.

The problem of transmitting coloured images is that the signal has to be such that it can be rendered faultlessly by black-and-white equipment (compatibility), and also that a black-and-white only image can be viewed in perfect conditions on colour equipment. This can only be achieved if information about luminance and colour can be derived from the signal. A coloured image can be defined by its luminance and chromaticity. The colour is characterized by hue — defined by the dominating wavelength in the spectrum — and saturation as a measure for spectral purity, i.e. the intensity of the colour with respect to the colourless white. While luminance can be represented by an electrical signal in a straightforward manner, this is not the case for colour. There is no technical possibility for a sensor to react selectively on hue or saturation. Therefore Helmholtz' theory is used, which states that a coloured scene can be represented by three basis or primary colours, red, green and blue. Colour selective sensors generate colour signals R, G and B which are directly proportional with the amount of these three primaries present in the scene. By encoding these three signals, a luminance signal can be obtained, as well as a chrominance signal. This chrominance signal must not be regarded as an electrical representation of colour. It contains information about hue, saturation and intensity.

From theory, and from the knowledge of the primary colours and the eye sensitivity, it can be derived that the luminance signal (usually denoted by Y), is the following combination of R, G, and B:

$$Y = 0.30 \times R + 0.59 \times G + 0.11 \times B$$

In principle, the signals R, G, and B, used in the above formula, should be gamma corrected values (see section 4.3).

Since the luminance Y does not contain colour information, additional signals need to be transmitted, from which the RGB values can be derived, e.g.

$$R - Y, G - Y, B - Y$$

Only two of these colour difference signals, together with Y, are required to recover the RGB values. The difference signals that are selected are $R - Y$ and $B - Y$.

The luminance and chrominance signals are then combined as

$$Y + (B - Y) \cos 2\pi f_{SC} t + (R - Y) \sin 2\pi f_{SC} t$$

where it is clear that for the chrominance signal amplitude modulation is used, of a carrier with frequency f_{SC}. Black-and-white equipment would simply smooth out the carrier frequency and therefore basically "see" only the luminance Y. How this carrier frequency is selected will be explained next. But first notice that the amplitude

$$A = \sqrt{(B - Y)^2 + (R - Y)^2}$$

yields the saturation and the phase angle

$$\alpha = \arctan \frac{(R - Y)}{(B - Y)}$$

determines the hue (see fig. 3.3).

An essential aspect of colour television is the way in which the colour information coexists with the monochrome information in the frequency spectrum. This coexistence is accomplished through frequency interleaving of the harmonic structure of the colour information between the harmonic structure of the monochrome information, as explained below.

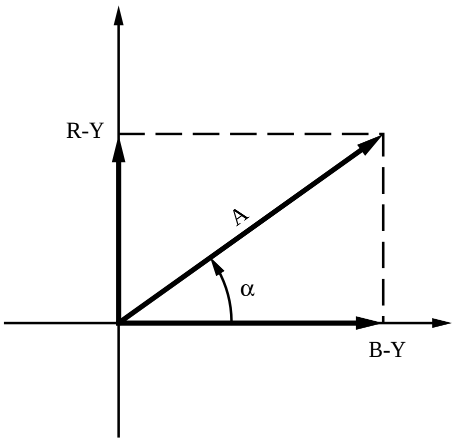

Figure 3.3: Colour represented by the colour difference signals.

Assume that some signal $s(t)$ is perfectly periodic and a single period repeats itself at a rate f_h Hz. The frequency spectrum of such a periodic signal will consist of lines at integer, or harmonic, multiples of f_h.

A luminance video signal is highly repetitive, since each scan line is very similar to each preceding line. The frequency spectrum of such a highly repetitive signal exhibits the harmonic structure typical of periodic signals. Hence, the spectrum consists of bursts of energy at harmonic multiples of the horizontal scanning frequency (roughly 15,625 Hz).

Of course, each scan line is not the same for a real television signal, even if the image itself is stationary. The scan lines for a stationary image change as the image is scanned from top to bottom and then scanned again and again at the field and frame rates. The frequency spectrum for such a "still" image has components at harmonic multiples of the scanning frequency f_h but each of these components has smaller lines above and below, in essence, small sidebands. The spacing between these smaller lines is the field and frame rate, or 50 and 25 Hz. If the image changes with respect to time, the major harmonic lines and their smaller brethren develop sidebands and become diffuse.

Thus, the frequency spectrum of a television signal consists of bursts of energy at harmonic multiples of the horizontal scanning rate. Each burst of energy itself exhibits a symmetric harmonic structure at multiples of the frame and field rates of 25 and 50 Hz, as shown in figure 3.4. Each burst of energy decays to zero after a dozen or so harmonic multiples of the frame rate which is well before the beginning of the next burst of energy centered around the next harmonic of the horizontal scanning frequency. What this means is that there is a fair amount of empty space in the frequency spectrum between the bursts of energy centered around the harmonics of the horizontal scanning frequency. The frequency spectrum looks like the periodically spaced teeth of a comb.

The chrominance signal likewise is fairly repetitive at the horizontal scanning fre-

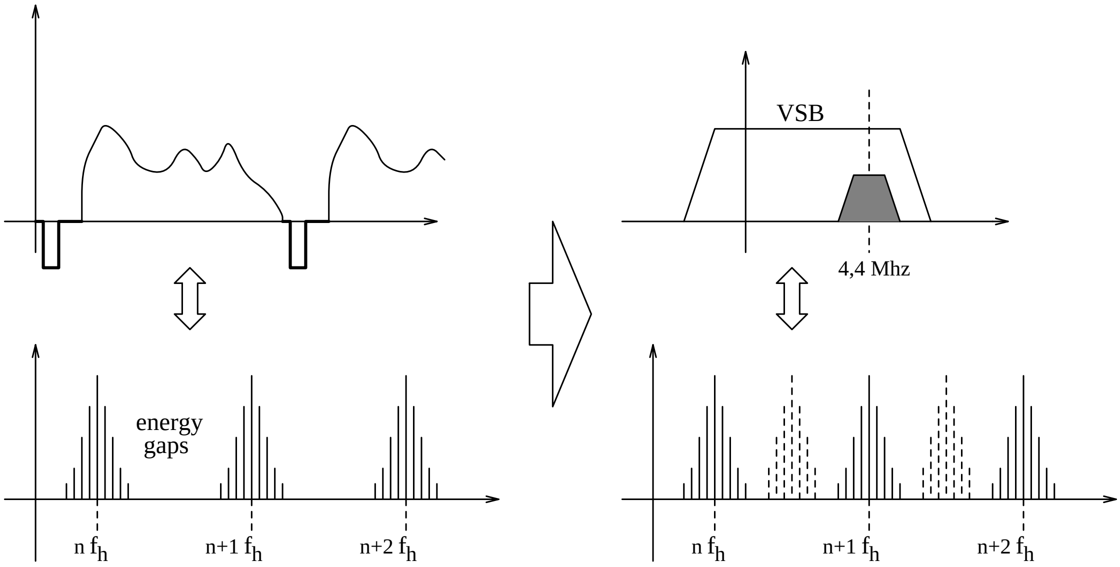

Figure 3.4: Spectrum of CVS and modulated colour subcarrier.

quency, and thus it, too, has a frequency spectrum that exhibits a harmonic structure very much like the spectrum of the luminance signal. Thus by offsetting the frequency spectrum of the chrominance signal by half the horizontal scanning frequency, its spectral bursts of energy fall exactly in between the spectral bursts of the luminance signal. In this way, the two signals share the same spectral space with no interference between them, as shown in figure 3.4. This technique of interleaving two signals in the frequency spectrum is called frequency interleaving.

This can be achieved by modulating the chrominance signal onto a colour subcarrier whose frequency f_{SC} is located between the frequency components of the composite video signal (CVS).

$$f_{SC} = (2n+1)\frac{f_h}{2}$$

Generally such a subcarrier can generate an interference pattern on a *b*&*w* picture. A sinewave for instance will produce a bright-dark pattern on the screen. The situation is clarified in fig. 3.5. If the subcarrier had been a multiple of the line frequency, one would observe bright and dark vertical stripes. However, due to the half-line offset, the phase of the colour subcarrier alternates by 180° from line to line. Since the number of lines is odd, bright and dark dots will coincide after two fields. The interference pattern thus would be compensated completely (i.e. over the whole display) over four fields. This compensation is not perfect due to the non-linearity of the picture tube and to inadequate capability of the human eye to integrate. The annoyance to the human eye can be reduced by selecting a subcarrier frequency as high as possible. This way the interference pattern would take a very fine structure. Since the subcarrier is modulated

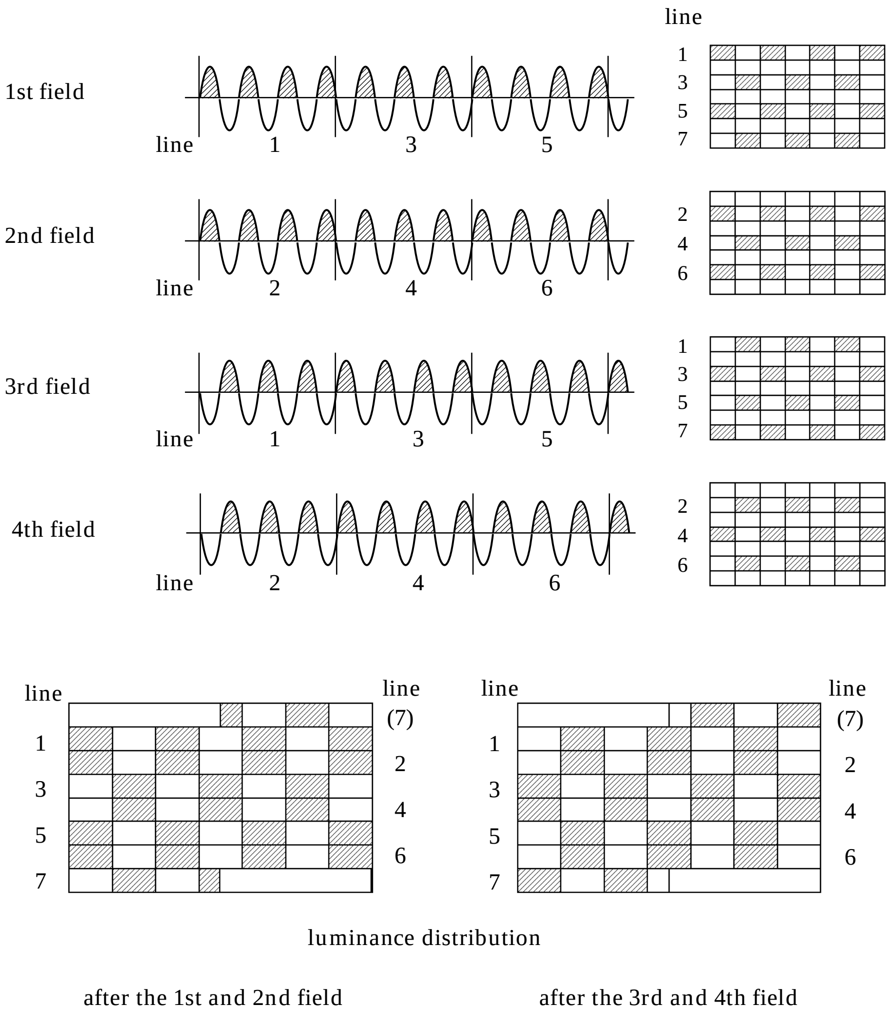

Figure 3.5: Interference pattern for different lines in the subsequent fields.

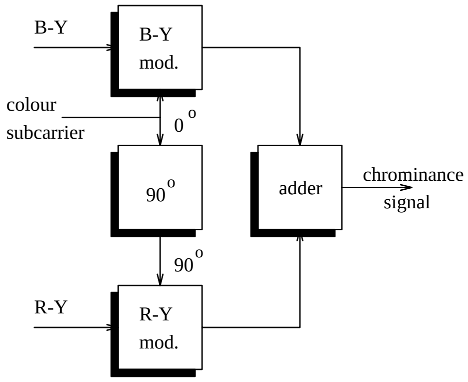

Figure 3.6: Quadrature modulation.

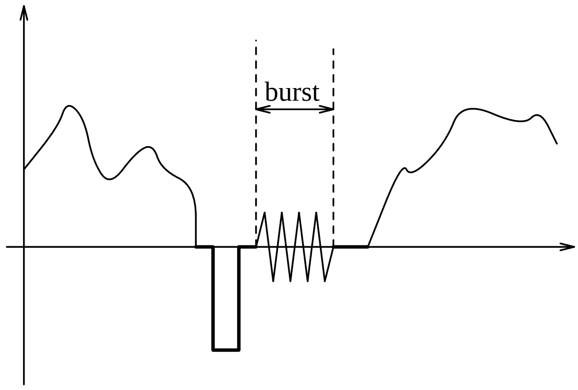

Figure 3.7: Burst signal in blanking interval.

by the colour difference signals, one has to provide a minimum spacing from the upper frequency limit of the CVS. The best compromise has been found to be about 4.4MHz.

The modulation method should permit the colour difference signals to be extracted separately at the receiver. In the NTSC and PAL methods (see later) a double amplitude modulation is used: the 0° component of the subcarrier is modulated by $B - Y$ and the 90° component by $R - Y$, the subcarrier itself being suppressed. In this way we obtain quadrature modulation (fig. 3.6).

To demodulate the chrominance signal, the unmodulated carrier of correct phase is required. Since the actual subcarrier isn't transmitted, it must be produced as a reference carrier at the receiver end. For synchronization, a reference signal is inserted into each line in the blanking interval. This colour sync signal or *burst* contains about ten oscillations of the subcarrier at the transmitter (fig. 3.7). In the NTSC system the burst phase is 180° compared with the reference phase of the subcarrier. In the receiver the burst is first separated from the chrominance signal, taken to a phase comparator which controls a reference oscillator. The control voltage becomes zero if the phase difference

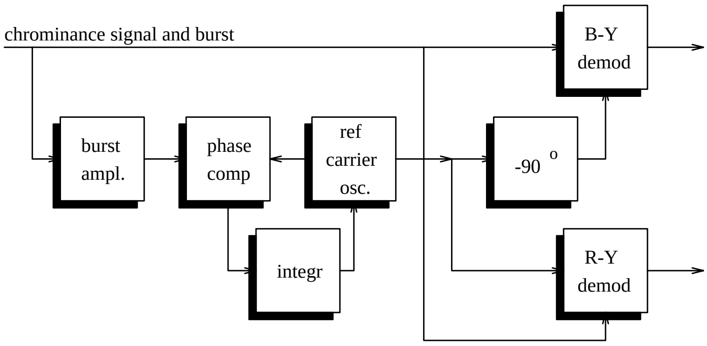

Figure 3.8: Synchronization at receiver.

between the reference carrier and burst is 90°. This 90° component is taken directly to the $R - Y$ demodulator, while the $B - Y$ demodulator gets it after a $-90°$ shift. The whole operation is summarized in fig. 3.8.

Investigations show that the human eye is much more sensitive to variations in brightness than in colour. Therefore, only the luminance signal has to be transmitted with full bandwidth (5 MHz), whereas the bandwidth of the colour-difference signal can be reduced (to 1.3 MHz).

When combining the chrominance signal with the CVS, one obtains the composite colour video signal (CCVS). This CCVS is modulated onto the RF vision carrier. The basic outlook of the CCVS is illustrated in fig. 3.9.

Note that the full level of the colour difference signals would cause overmodulation[1] for certain colour patterns, as is clearly visible for the standard colour-bar sequence. Overmodulation occurs in both directions, causing heavy interference. For this reason the amplitude of the chrominance signal is reduced. As a compromise between overmodulation and degradation of signal-to-noise ratio, it was decided to permit an overmodulation of 33%, bearing in mind that fully saturated colours hardly ever occur in practice. The reduction of the overmodulation can be ensured by multiplying the colour difference signals with different reduction factor:

→ 0.49 for the $B - Y$ signal

→ 0.88 for the $R - Y$ signal

The so-called reduced colour difference signals U and V are

$$U = 0.49(B - Y) = -0.15R - 0.29G + 0.44B$$

$$V = 0.88(R - Y) = 0.61R - 0.52G - 0.10B$$

[1]Overmodulation occurs when the amplitude of the modulating function surpasses that of the carrier. The resulting envelope distortion leads to incorrect demodulation.

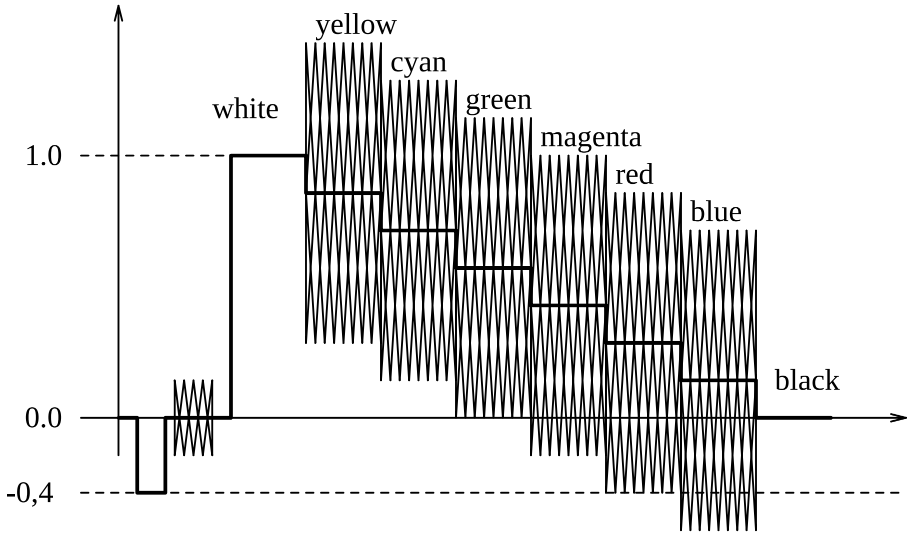

Figure 3.9: Composite colour video signal.

3.3 RF transmission of the CCVS

The NTSC, PAL and SECAM methods are the main ones for colour-TV transmission. They differ only with respect to the modulation of the colour subcarrier.

3.3.1 NTSC

The NTSC method is named after the National Television System Commitee. Its principle is explained by the sections above on modulation of the colour subcarrier and the CCVS. The original NTSC system, however, doesn't use the reduced colour difference signals U and V, but instead transmits two components I and Q. I corresponds to the axis for which the eye has maximum colour resolution, and Q the axis of minimum resolution, as shown in fig. 3.10. This results in a better transmission of colour transitions. The signal is then

$$Y(t) + I(t)\cos(2\pi f_{SC}t + \phi) + Q(t)\sin(2\pi f_{SC}t + \phi)$$

with $\phi = 33°$. The two signals are transmitted with different bandwidths, I with 1.3 MHz and Q with 0.5 MHz. An overview of the operations in the NTSC coder is shown in fig. 3.11.

The NTSC has one big disadvantage. Since the CCVS has to be produced by and transmitted through some network, distortion may occur due to non-linearities of the transmission characteristic. This distortion can be described by a *differential phase* (difference of phase shift through a network at two different points of the transfer characteristic) and a *differential gain* (similar definition). Both are illustrated in fig. 3.12. This distortion manifests itself as an incorrect hue on the television screen, a phenomenon

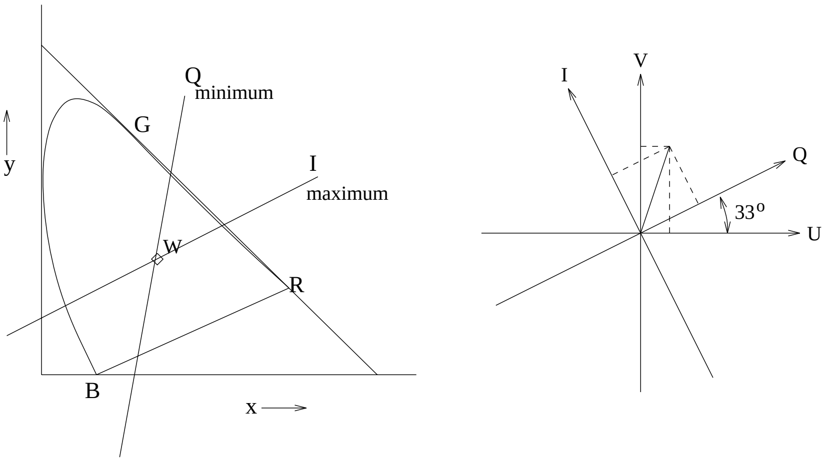

Figure 3.10: I and Q components in NTSC system.

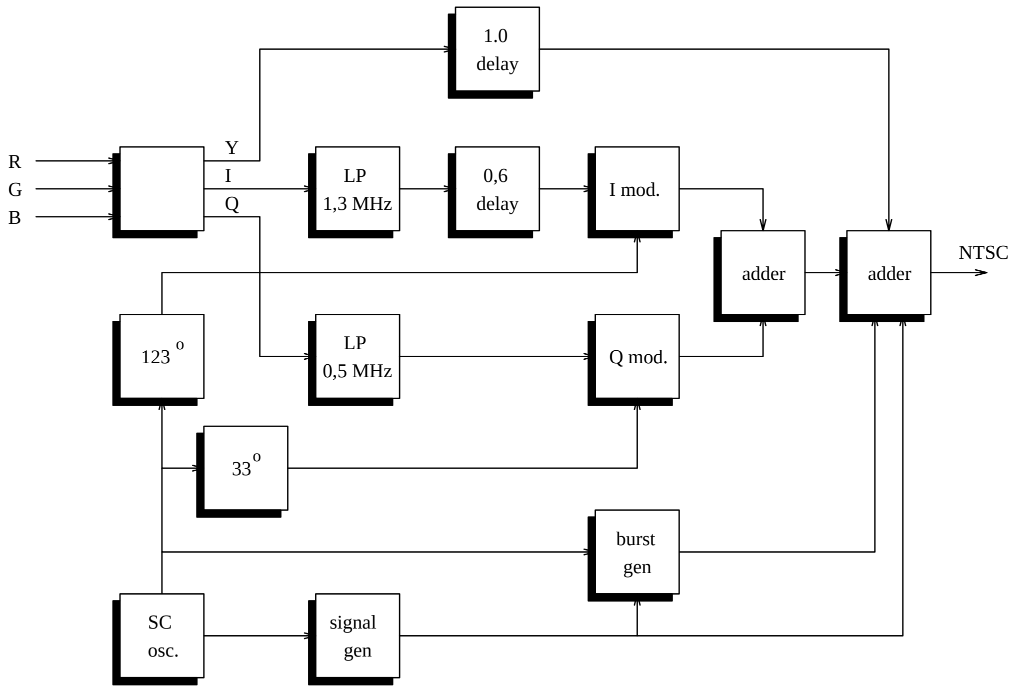

Figure 3.11: NTSC coder.

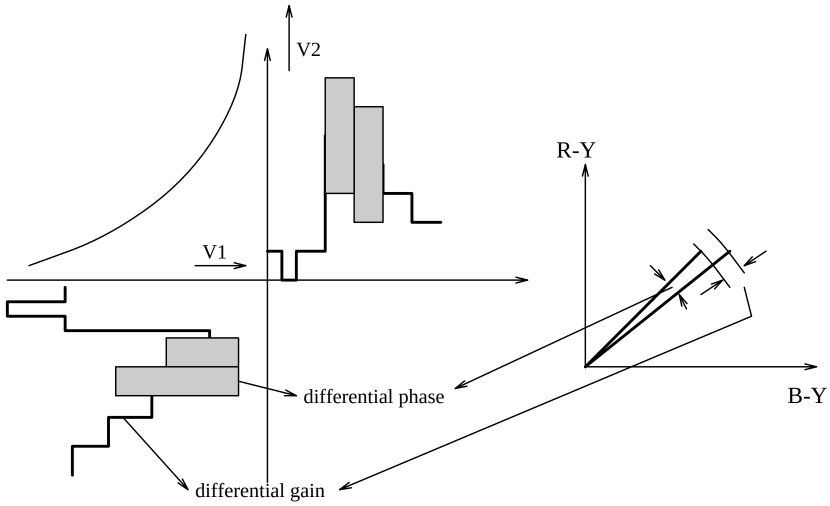

Figure 3.12: Differential phase and gain distortion.

to which the human eye is very sensitive [2]. Differential gain can be dealt with, by referring to the hue of a well known picture detail. However, differential phase is harder to control. The PAL and SECAM methods include the efforts to improve the NTSC method.

3.3.2 PAL

The effects of the differential phase error are considerably reduced by the PAL method. It is based on the concept that the phase error can be compensated by a phase error of opposite polarity. This can be realized by alternating the phase of one of the two chrominance signal components by 180^o from line to line. PAL thus stands for *Phase Alternation Line*. If the chrominance signal is delayed for the duration of a line and the delayed and undelayed signals are added, the two phase errors of opposite polarity will coincide and cancel. The principle is illustrated in fig. 3.13. This method of course assumes that the chromaticity doesn't change too much within two consecutive lines, which is satisfied in most cases. But even when it is not, the eye hardly perceives the falsification of the colours.

To detect the phase reversal at the receiver end, an additional identification is needed. The burst is therefore split into two components, one being transmitted at $180°$ and the other at $\pm 90°$ alternating from line to line. This yields the so-called swinging burst of $180° \pm 45°$. The actual burst phase can be recovered by averaging.

With PAL the reduced signals U and V are transmitted directly, with a bandwidth of 1.3 MHz. Compared to the NTSC coder, the $33°$ phase shift of the subcarrier com-

[2]This characteristic gave rise to the alternative explanation for the abbreviation NTSC: *Never The Same Color*

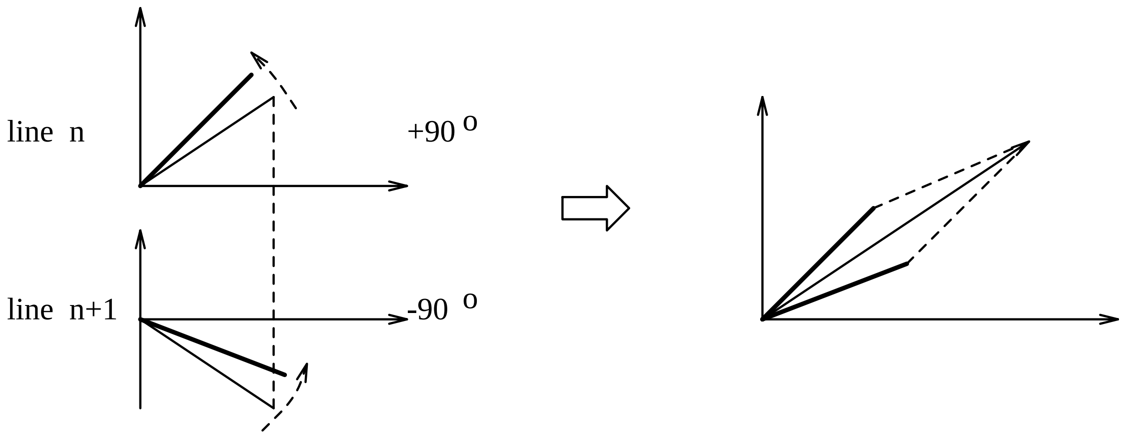

Figure 3.13: Compensation of phase error with PAL.

ponents is omitted, but the reversal of the subcarrier component $B-Y$ modulator and generation of the swinging burst are added.

Schematic representations of the PAL coder and decoder are shown in figures 3.14 and 3.15.

3.3.3 SECAM

Like the PAL method, the SECAM method assumes that the colour information doesn't vary essentially from line to line and that the human eye doesn't perceive any annoyance if the vertical colour resolution is reduced to a certain extent. Therefore, the colour difference signals needn't to be transmitted simultaneously, they can be sent separately in successive lines. In the receiver, the signal content of one line is stored for the duration of a line, an processed together with the signal of the next line. SECAM indeed stands for *séquentiel à mémoire*. Since the colour signals are transmitted separately, the type of modulation can be freely selected. It has been decided to use frequency modulation, since it isn't very sensitive to interference. Figure 3.16 summarizes the principles behind SECAM coders and decoders.

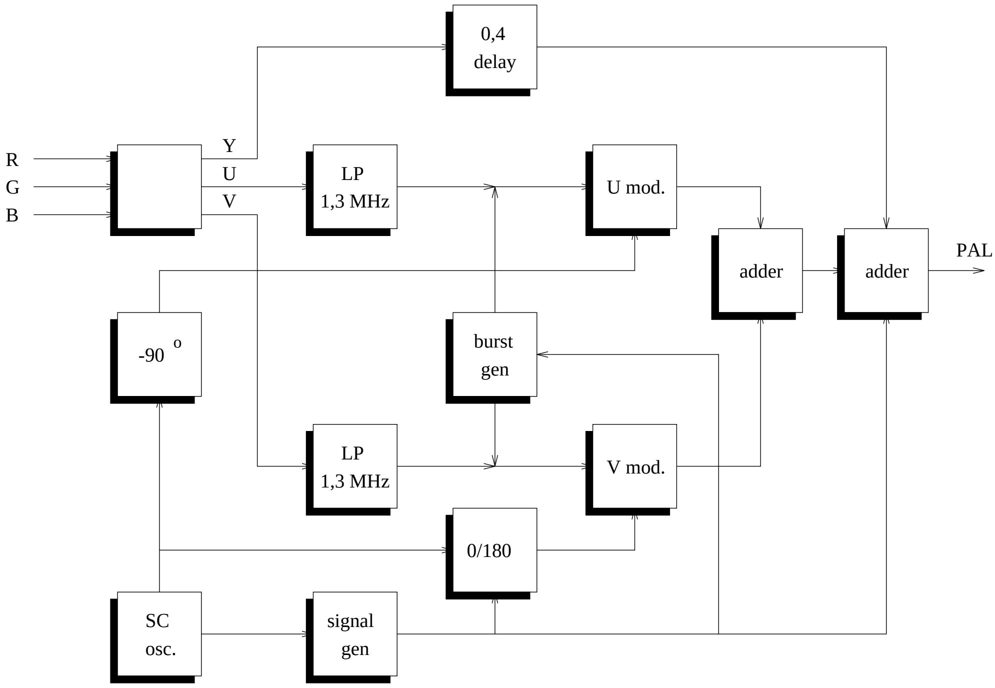

Figure 3.14: PAL coder.

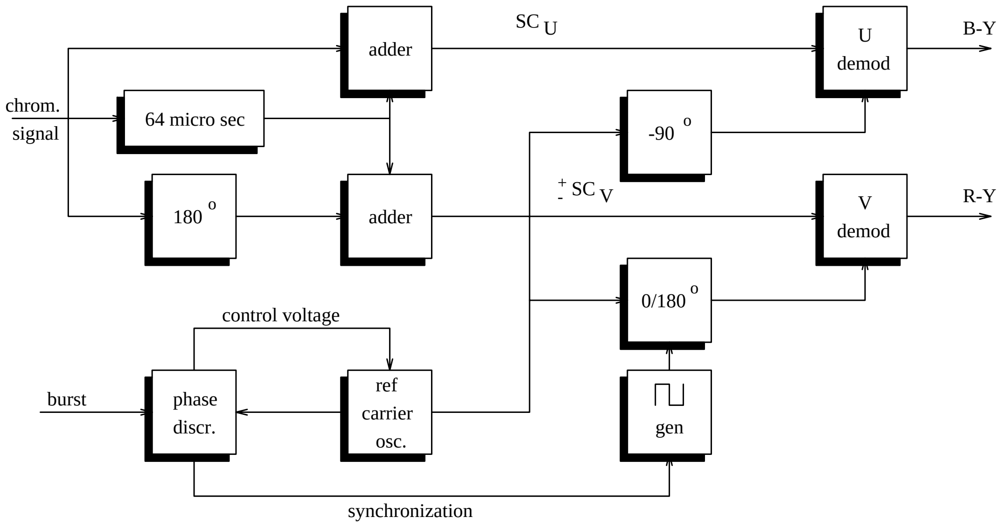

Figure 3.15: PAL decoder.

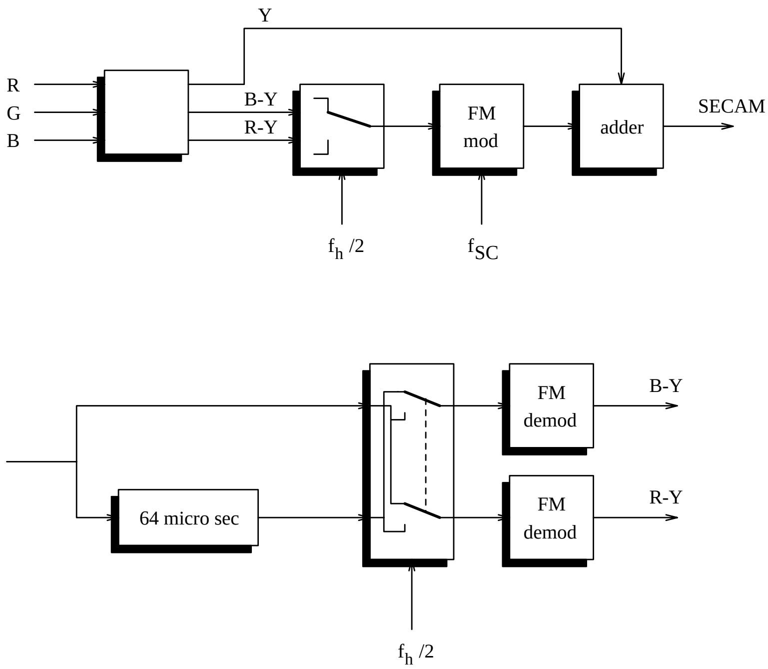

Figure 3.16: SECAM coder and decoder.

4 Displays

Many different technologies exist to convey visual information to users. Electronic displays can be divided into three descriptive groups — projection, off-screen and direct-view — and further categorized by the specific technology used to produce a visual image.

Projection-type displays offer an image consisting of light reflected off a display surface. By implication, there is an optical system to direct light to the reflective surface. This type of display is not used in industrial image processing applications and will therefore not be discussed further in this text.

Off-screen displays are also beyond the scope of this text. Examples are holographic three-dimensional displays using laser technologies or helmet-mounted systems used by fighter pilots.

The displays that are used in industrial image processing applications are all *direct-view displays*. By definition, the viewer sees an image directly on or near the display surface. Direct-view displays can be categorized not only by the technology that is used, but also by the application: alphanumeric displays, vector displays or photographic displays. Obviously, the category that we are interested in here, is the photographic display. The major technologies that are currently available for this application, are the following:

- cathode ray tube (CRT);
- electroluminescent display;
- liquid crystal display (LCD);
- light-emitting diode (LED) display;
- gas discharge or plasma display.

The electroluminescent, liquid crystal, LED and gas discharge displays are all flat panel displays (FPD's), whereas the CRT is a bulky device (as a consequence of the elegant electron-beam deflection technique — although references to flat CRT's exist; these are however not successful).

Of these, the CRT is still the most widely used. Even today it is capable of higher overall performance than any flat panel display yet demonstrated. In almost every category where FPD's are used, a CRT can perform better at a lower price. FPD's are being used only where CRT's cannot reasonably fit, for example, in the briefcase or pocket.

Of all above mentioned flat panel technologies, most recently the LCD has emerged as the clear leader. The production volume of high information content (HIC) LCD's is now more than 100 times greater than all other HIC FPD technologies combined. There are many reasons for the acceleration of LCD technology, including:

- highest immunity to ambient illumination;
- thinnest profile;
- lightest weight;
- lowest power requirement;
- color performance comparable to CRT;
- lowest cost, compared to other FPD technologies.

Negative limitations are diminishing and becoming more acceptable to consumers, including:

- very high cost compared to CRT's (more than ten times that of a comparable CRT);
- limited viewing angle (now greater than $\pm 45°$ in one axis with 10:1 contrast ratio);
- slow speed of response (now less than 200 ms in passive LCD's and 50 ms in active matrix LCD's);
- narrow temperature operating range (now as wide as $-30°$C to $+85°$C).

The LCD technology is still evolving. Several configurations are in high-volume production, such as the twisted nematic LCD, supertwisted nematic (STN) LCD, multiple row addressed STN LCD, metal-insulator-metal (MIM) active matrix LCD, split electrode STN LCD, thin-film transistor active matrix LCD, and others. These configurations cover a wide spectrum and all seem to be finding markets which best fit their individual performance / cost ratio and features.

It would be beyond the scope of this text to cover all these technologies in detail, especially since in the industrial image processing applications, the CRT is still most widely used. We will therefore discuss only the CRT in more detail here.

4.1 Cathode Ray Tubes

For reproducing brightness patterns from the electrical signals into which they were transformed by a camera, so-called *cathode ray tubes* (CRTs for short) are still the most widely used devices.

The global structure of a black-and-white CRT is shown in fig. 4.1. A CRT is a vacuum-filled glass tube, in which a heated cathode emits a continuous stream of electrons. These are formed into a beam by an aperture in a control grid surrounding the cathode. The electrons are then accelerated by one or more positive-potential grids and

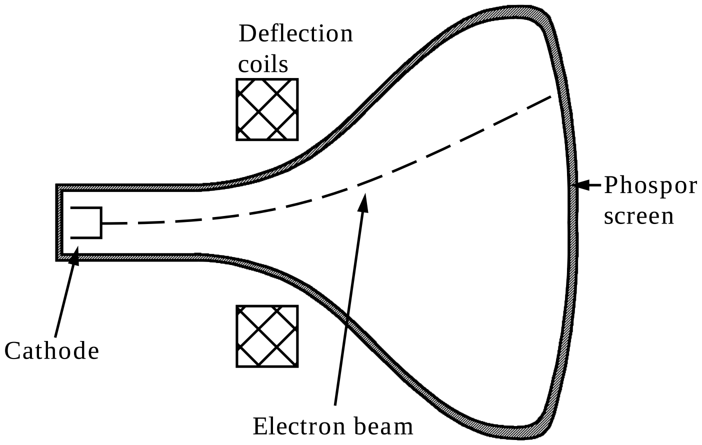

Figure 4.1: Basic design of a black-and-white CRT.

focused to a point at the display surface by an electrostatic or magnetic "lens". The focused beam is turned toward the top, bottom, or sides of the CRT screen by a deflection mechanism which again can be electrostatic or magnetic. In commercial monitors, mostly magnetic deflection and focusing is found, using deflection coils (as shown in the figure). The deflection is controlled by the sync pulses that are recovered from the incoming video signal. This guarantees that the lines and fields on the screen correspond to exactly the same lines and fields in the camera at the time the image was acquired. After deflection, the electrons travel in a straight line toward the faceplate in a uniform field created by a high positive potential on a conductive surface which covers the inside of the CRT "bottle". The front surface is coated with a thin layer of phosphor particles. When the electrons strike the phosphor particles, a fraction of the kinetic energy is absorbed, exciting electrons in the phosphor molecules to a higher energy state. A photon of visible light is emitted as each excited electron falls back to its natural energy state. The duration or *persistence* of the photon emission may last for milliseconds or seconds, depending on the type of phosphor (see table 4.1). The intensity of the electron beam is controlled by the voltage across the control electrode. This voltage is derived from the luminance part of the video signal.

4.2 Colour CRTs

In colour CRTs, three electron guns are incorporated in the same vacuum tube. This alleviates some of the disadvantages of using three different tubes, such as geometric registration. Electronic registration is still required, of course, but by having the three guns in the same tube, the same magnetic fields can be used for moving the three electron beams throughout the scanning sequence for each field of the picture. This is a considerable help, but there are many residual problems caused by the fact that, be-

Phosphor	Composition	Peak Wavelength Nanometers	Colour	Decay to 10%	Applications
P1	Zn_2SiO_4:Mn	525	YG	24ms	oscilloscopes; radar
P2	ZnS:Cu	543	YG	35-100μs	oscilloscopes
P4	ZnS:Ag				
	ZnS-CdS:Ag	460 / 560	W	25 / 60μs	television
P7	ZnS:Ag				
	ZnS-CdS:Cu	440 / 560	B / YG(W)	40-60μs / 0.3s	oscilloscopes; radar
P11	ZnS:Ag	460	B	25-80μs	photo recording
P16	$Ca_2Mg_2Si_2O_7$:Ce	385	UV	0.12μs	flying spot recorders; photo recording
P20	ZnS-CdS:Ag	560	YG	0.05-2ms	high efficiency phosphor
P22-B	ZnS:Ag	440	B	22μs	colour TV
P22-G	ZnS-CdS:Ag	535	YG	60μs	colour TV
P22-R	Y_2O_2S:Eu	625	R	1ms	colour TV
P22-G_{LP}	Zn_2SiO_4:Mn::As	525	YG	150ms	long persistence phosphor; colour graphics
P31	ZnS:Cu	522	G	40μs	high efficiency phosphor; oscilloscopes
P39	Zn_2SiO_4:Mn::As	525	YG	150ms	long persistence, low frame-rate displays
P42	ZnS:Cu				integrating phosphor;
	Zn_2SiO_4:Mn::As	520	YG	10ms	low frame-rate, high brightness displays
P43	Gd_2O_2S:Tb	544	YG	1ms	visual displays
P44	La_2O_2S:Tb	540	YG	1ms	visual displays
P45	Y_2O_2S:Tb	420 / 540	W	2ms	visual displays
P49	Zn_2SiO_4:Mn				penetration colour
	YVO_4:Eu	525 / 615	YG / RO	30 / 1.2ms	displays

Table 4.1: Characteristics of phosphors used in typical CRT applications.

cause the electron beams do not originate from the same place, they do not in fact scan the picture identically. Thus when the beams are scanning the corners of the picture they have further to travel to the screen than when they are scanning the center; hence if their convergence is correct for the center it will be too great for the corners, and so magnetic fields have to be altered during the scanning by means of special current waveforms applied to the deflection coils.

In a *shadow-mask tube*, the three beams meet a metal plate with about 400,000 holes in it, situated about 18 mm from the phosphors, before reaching the screen. The three phosphors are laid down as dots (see fig 4.2), and the geometry of the electron beam directions, the positions of the holes, and the positions of the dots, is such that all the red phosphor dots are irradiated only by the gun to which the red signal is applied, the green phosphor dots by the green signal gun, and the blue dots by the blue signal gun. To avoid moiré patterns, it is arranged for the lines of the picture and the lines of the dots to be more or less parallel.

A three-gun tube in which the phosphor is laid down in stripes, is the *Trinitron*. In this tube the three electron beams lie in the same horizontal plane, and a metal plate with vertical slots in it is positioned so that the electrons from one beam can only reach stripes of red phosphor, those from another beam only stripes of green, and those from

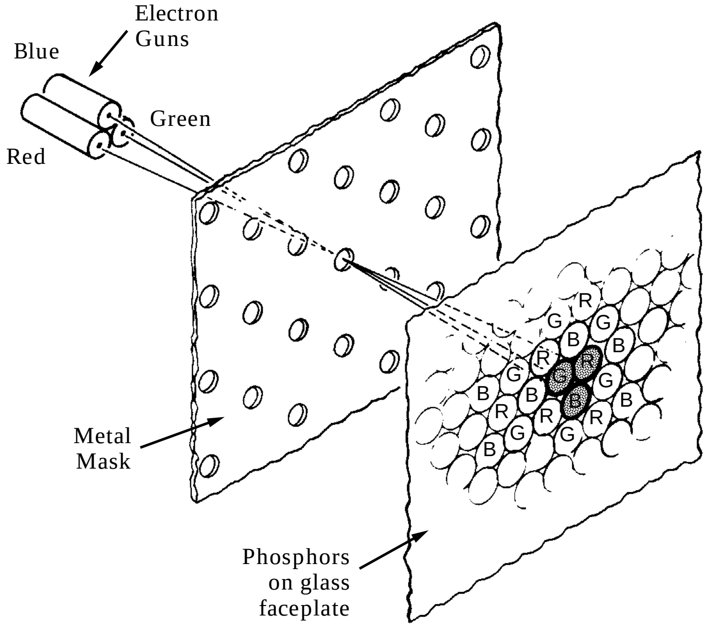

Figure 4.2: Shadow-mask CRT.

the third only stripes of blue (see fig. 4.3).

This tube has certain advantages over the shadow-mask tube. First, deflection of the three electron beams is easier because the gun construction enables the neck of the tube to be smaller. Secondly, the displayed picture emits twice as much light per unit area: this is because, for the same spot size, the beam current can be increased by a factor of 1.5, and because the stripes of phosphor cover 1.33 times as much area of the tube face-plate. Thirdly, adjusting the convergence to obtain registration of the three images is easier because the three beams are in a single plane.

In Trinitron and conventional shadow-mask tubes, it is necessary to provide *dynamic convergence correction*. This is required because stronger magnetic fields are needed to bring the three electron beams into coincidence around the center of the picture, than those required for the corners, which are further away and therefore have longer electron paths. As the three electron beams scan the picture, the amount of convergence is therefore adjusted dynamically according to their position in the scan.

A third tube, the *Precision In-line (PIL) tube*, is shown in fig. 4.4. The three electron guns are arranged parallel to one another in the same horizontal plane, as in the Trinitron tube; but, instead of providing dynamic convergence correction, a special deflection coil is cemented to the neck of the tube. This coil is designed to converge the three electron beams on to the shadow mask at all positions in the picture.

The conventional shadow-mask, Trinitron and self-converging types of tube are widely used in monitors and domestic receivers.

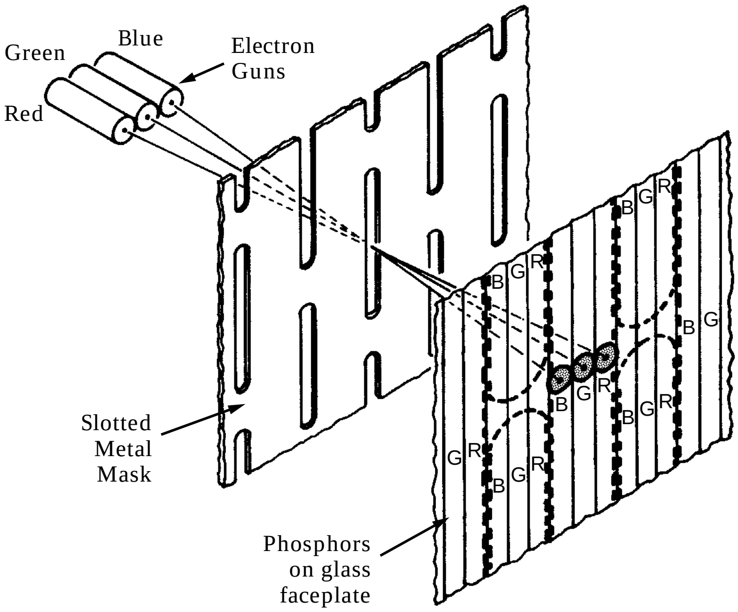

Figure 4.3: Trinitron CRT.

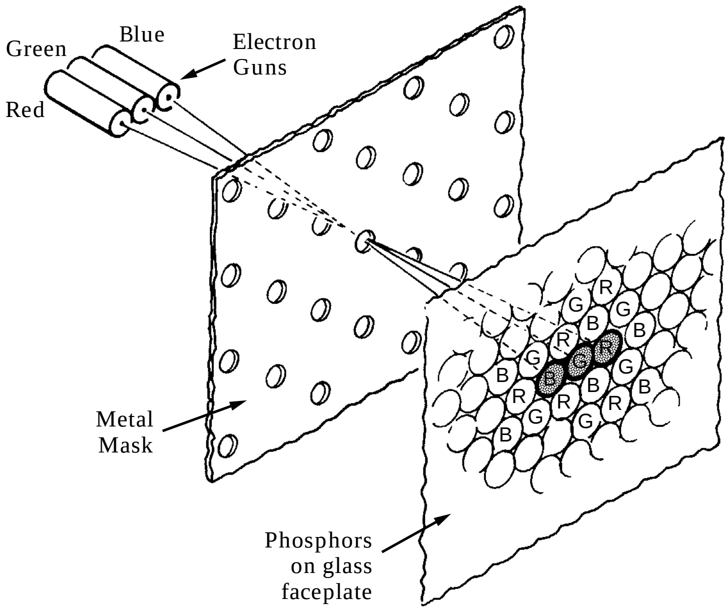

Figure 4.4: Precision in-line CRT.

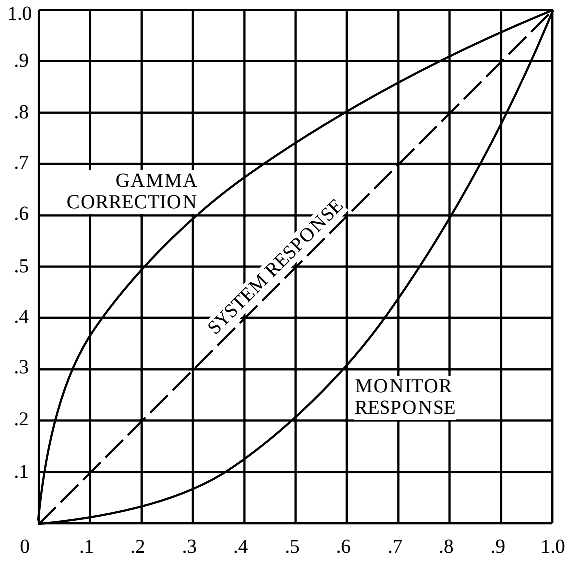

Figure 4.5: Monitor response and gamma correction.

4.3 Gamma correction

There is a direct relationship between the display signal and the voltage applied the monitor CRT. The luminous output of the phosphors on the CRT monitor screen is not, however, directly proportional to either of these values.

Figure 4.5 plots the luminous output of the monitor as a function of the display-signal voltage. Both are shown as fractions of their full-scale values. The monitor output is less than linear at low display signal values, more than linear at high display signal levels.

This relationship is described by the following function:

$$\text{Output} = (Input)^{\gamma}$$

From the point of view of signal-to-noise ratio, a high gamma is desirable because the darker portions of the picture, where noise is most obvious, tend to be reproduced nearly black. But while a monochrome picture which has a gamma of about 3 might be tolerable, a colour picture will exhibit severe colour distortion.

Suppose the electrical signals E_R, E_G, E_B are intended to produce a colour $r(R) + g(G) + b(B)$ on a linear display. If, instead, they are applied to a cube law display, the resulting colour is $r^3(R) + g^3(G) + b^3(B)$. For example, if $r = 1$, $g = 0.5$, $b = 0.5$, and unit quantities of R, G, and B result in a white (W), then the intended colour is equivalent to $0.5W + 0.5R$. But the displayed colour will be $r = 1$, $g = 0.125$, $b = 0.125$, or $0.125W + 0.875R$. Hence the luminance has decreased, and the saturation has increased. A simple means of correcting for this is to pre-distort the signals E_R, E_G,

E_B at the transmitter to $E_R^{1/\gamma}, E_G^{1/\gamma}, E_B^{1/\gamma}$, before they are being combined to form the luminance and chrominance signals.

This method gives distortion-less large-area reproduction. But the luminance carried by

$$E'_Y = 0.30E_R^{1/\gamma} + 0.59E_G^{1/\gamma} + 0.11E_B^{1/\gamma}$$

is $(E'_Y)^\gamma$. The ratio of luminance carried by E'_Y to the true luminance is:

$$\frac{(E'_Y)^\gamma}{E_Y} = \frac{(0.30E_R^{1/\gamma} + 0.59E_G^{1/\gamma} + 0.11E_B^{1/\gamma})^\gamma}{0.30E_R + 0.59E_G + 0.11E_B}$$

For the worst case (saturated blue) this ratio is:

$$\frac{0.11^\gamma E_B}{0.11E_B} = 0.11^{\gamma-1} = 0.11^{1.8} = 0.019$$

So in this case the luminance signal E_Y carries only 2% of the true luminance; hence small-area saturated colours are reproduced too dark. Also compatibility suffers, as a monochrome receiver will display too little luminance; however, in practice, the effect of the non-linearity of the cathode-ray tube characteristic on the dots produced by the chrominance signals increases the 2% quoted to about 5%. Thus monochrome errors are not too bad, and large area colour is correct. But as E'_Y does not carry all the luminance, the remainder must be carried by the sub-carrier modulation (see chapter 3), and hence the constant luminance principle is not obeyed, with the result that the subjective effect of noise and interference is increased.

A further point is that, as the sub-carrier modulation is severely limited in bandwidth, definition will suffer because the luminance content of the sub-carrier will also be limited in bandwidth. But for white and the more prevalent neutral shades the ratio will not be very much less than unity. Hence the above shortcomings are evident only for the higher saturations.

There are several alternative methods for gamma correction, but in general these involve additional complications at the receiver. For instance, if the luminance signal was composed *before* E_R, E_G, and E_B were pre-distorted then the above difficulties would not arise. But the recovery of the green signal is then much more complicated, requiring the signals first to be raised to the power γ, then mixed to obtain the green signal, then re-distorted to the power $1/\gamma$, before finally applying them to the tube.

The practice in industry is thus to correct for this "gamma" non-linearity at the signal source: the camera. Since cameras themselves behave linearly within certain limits, no gamma correction is needed for the camera itself (see also table 2.1 in chapter 2).

There is considerable disagreement on the most appropriate general-purpose value for gamma. Values up to 2.8 ($\pm$0.3) have been cited; however, in practice a factor of 2.2 is used. The result of this is that in the final display the gamma of the picture is increased over that of the original scene by a factor of 2.8/2.2, that is 1.27 times. This increase in displayed gamma is necessary in order to overcome the reduction in apparent contrast caused by the dim surround conditions in which television is normally viewed, but increases in purity and shifts in dominant wavelength occur.