TV & Video Engineer's Reference Book

EDITED BY

K.G. JACKSON

G.B. TOWNSEND

TV & Video Engineer's Reference Book

This page intentionally left blank

TV & Video Engineer's Reference Book

Edited by

K G Jackson

G B Townsend

With specialist contributors

Butterworth-Heinemann Ltd
Halley Court, Jordan Hill, Oxford OX2 8EJ

OXFORD LONDON GUILDFORD BOSTON
MUNICH NEW DELHI SINGAPORE SYDNEY
TOKYO TORONTO WELLINGTON

First published 1991

British Library Cataloguing in Publication Data
Television and video engineers reference book.
1. Television equipment 2. Video equipment
I. Jackson, K. G. (Kenneth George) *1930–* II.
Townsend, Boris
621.388

ISBN 0-7506-1021-2

Library of Congress Cataloging-in-Publication Data
Television and video engineers' reference book / edited
by Kenneth G. Jackson, Boris Townsend.
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0-7506-1021-2 :
1. Television–Handbooks, manuals, etc. I.
Jackson, Kenneth George. II. Townsend, Boris.
TK6642.T436 1990
621.388–dc20 90-2107

Typeset by Saxon Printing Ltd, Derby
Printed and bound by Hartnolls Ltd, Bodmin, Cornwall

Preface

'The compilers of this book would be wanting in courtesy if they did not expressly say what might otherwise be safely left to the reader's discernment.' So wrote the brothers Fowler in the preface to their famous work on the King's English. It is a precept which any editor could observe with advantage, though perhaps disciplining himself to avoid too slavish a repetition of the Contents List.

The most superficial examination of this volume will indicate that it is a comprehensive survey of television technology and that it is authoritative. Eminent engineers of international stature and from many countries have spared time from their development work to share their knowledge and experience with the rest of us.

It is, indeed, in the best traditions of science that research and teaching should go hand-in-hand; and many a student has discovered that the originator of new ideas was more comprehensible in his writings than the popularizing scribes who attempted to explain what the innovator was talking about.

Television is currently in turmoil. It has always been so and, like it or not, since it is engineering-led it will always be in turmoil. As electronics and materials science develop at an ever-increasing rate, so will television. In the United Kingdom, de-regulation is in vogue and the economics of programme making will press ever more heavily on equipment design, though the basic engineering principles will persist.

To say that television is international is a platitude – and is incorrect. It is nowhere near as international as film. Countries insist on using their own technical standards and their own language. Direct broadcasting from geostationary satellites across national frontiers is an everyday occurrence yet governments cannot agree on a world standard. Even in the European Economic Community the directives defining which scanning and coding standards should be used for broadcasting from space are flouted by entrepreneurs. But the inexorable march of engineering developments, based on the principles and factors detailed in this book, have their own logic. Conversion from one picture standard to another in real time has been solved by the engineers, and machine translation between spoken languages is a process already working in the laboratory. When this is coupled with higher definition, seeing and hearing at a distance will become an even greater pleasure and even more instructive.

The world is growing smaller, even for those of us who do not travel.

GBT
Wrea Green, England
1991

Contents

Part 5 · Distribution of Broad-band Signals

Part 6 · Direct Broadcasting by Satellite

Part 7 · TV Studios and Studio Equipment

Part 15 · Industrial, Commercial, Medical and Defence Applications of Television

Part 16 · Performance Measurements and Electromagnetic Compatibility

List of Contributors

I G Aizlewood
Managing Director, Continental Microwave Ltd

C P Arbuthnot BEng
British Telecom Research Laboratories

P Audemars MA
Senior Film & Video Editor, London Weekend Television

J L E Baldwin BSc, MInstP, FRTS, FSMPTE
Consultant

J Barron BA, MA (Cantab)
University of Cambridge

P G J Barten
Barten Consultancy

D J Bradshaw BSc(Eng), AMIEE
Design and Equipment Department, BBC

D Bryan
Formerly Technical Director, Michael Cox Electronics Ltd

J H Causebrook BSc, PhD, MIEE, CEng, AMIOP
Service Area Planning Section, IBA

C K P Clarke
Senior Engineer, BBC Research Department

P A Crozier-Cole
Head of Telemetry and Automation Section, IBA

K Davison
Manager Communications, Thames Television

C Dawkins BA (Cantab)
British Telecom Research Laboratories

C Debnam
Quality Assurance Manager, PAG Ltd

R Elen
Creative Technology Associates

S R Ely PhD, CEng, MIEE
Head of Carrier Systems Section, BBC Research Department

B Flowers MRTS
Head of Eurovision Control Centre, European Broadcasting Union

D Hardy
Design Manager, PAG Ltd

S Hirata
Senior Specialist, Toshiba Corp

R G Hunt DSc, FRPS, FRSA, MRTS
Professor of Physiological Optics, City University

Y Imahori
Chief Engineer, NHK

G A Johnson BSc, EEng, MIEE
Deputy Head Engineering Services, ITV Association

E A Jones BEng, AMIEE
Assistant Technical Support Manager, Philips Communications and Security

J Kelleher MIEE
Formerly Chief Engineer, Dynamic Technology Ltd

P Kemble CEng, MIEE, BSc
Principal Engineer, IBA

W H Klemmer
Broadcast Television Systems GmbH

K Komada
General Manager, Toshiba Corp

S J Lent CEng, MIEE
Engineering Research Department, BBC

S Lowe
Manager International Technical Training, Ampex Ltd

R G Manton BSc(Eng), PhD, CEng, MIEE
Transmission Engineering Department, BBC

C A Marshman BTech, CEng, MIEF
York Electronics Centre, University of York

J D Millward BSc, CEng, MIEE
Head of Research, Rank Cintel Ltd

P L Mothersole FEng, CEng, FIEE
VG Electronics Ltd

K C Quinton MBE, BSc, FEng, FIEE, FRTS
Formerly Director of Research, British Cable Services Ltd

A F Reekie AMIEE
Formerly Senior Engineer, European Broadcasting Union

F M Remley Jr FSMPTE, MBKSTS
Technical Director Broadcasting Services, University of Michigan

R S Roberts CEng, FIEE, SenMIEEE
Consultant Electronics Engineer

J T P Robinson
MVC Crow Ltd

J G Sawdy BSc, CEng, MIEE
Independent Broadcasting Authority

B L Smith
Chief Technical Writer, Thomson-CSF

P Sproxton
Alpha Image Ltd

R Stevens BSc, CEng, MIEE
Engineering Project Supervisor, Thames Television

L Strashun MSc, CEng, MIEE, MBKS, MRTS
Senior Manager, Sony Broadcast Ltd

J Summers
Formerly Lighting Director, BBC

M Talbot-Smith BSc, CPhys, MInstP
Formerly BBC Engineering Training Department

D G Thompson BSc
Philips Components

E C Thomson
Mullard Application Laboratory

A Todorović
Director, Televigija Beograd

G P Tozer BSc
Principal Lecturer, Sony Broadcast and Communications

E Trundle MSERT, MRTS, MISTC
Chief Engineer, RNF Services Ltd

L W Turner FIEE
Consultant Engineer

N Wassiczek
European Broadcasting Union

G T Waters
European Broadcasting Union

I M Waters CEng, MIEE
Product Manager Transmitters, Varian TVT Ltd

R Watson CEng, MIEE, MRTS, MBKS
Consultant

P W Wayne
Formerly Marketing Director, Vinten Broadcast Ltd

L E Weaver BSc, CEng, MIEE
Formerly Head of Measurements Laboratory, BBC Designs Department

J P Whiting MSc, CEng, FIEE
Head of Power Systems, IBA

R Wilson BSc
Continental Microwave

G W Wiskin BSc, CEng, MICE, MIStructE
Architectural and Civil Engineering Department, BBC

D Wood
European Broadcasting Union

J M Woodgate BSc(Eng), CEng, MIEE, MAES, MInstSCE
Electronics Design Consultant

Part 1
Basic Reference Material

This page intentionally left blank

R S Roberts C Eng, FIEE, Sen MIEEE
Consultant Electronics Engineer

1

Television Standards and Broadcasting Spectrum

Every colour television channel consists of three modulated carriers:

- The *vision* information, derived from a camera or other signal source, is used to amplitude modulate a carrier with the electrical equivalents of the basic 'black and white' variations that are encountered during transmission of the scene.
- A subcarrier, situated within the bandwidth of the vision modulated carrier, is itself modulated with information related to the *colour* information in the scene.
- A separate adjacent carrier is modulated with the *sound* information contained in the scene.

The eye, as a visual communication system, 'sees' a large amount of detail simultaneously, by virtue of the fact that it has several million communication channels operating in parallel at any instant. The electrical signals that are generated by the millions of sensors in the eye are partly processed in the retina at the back of the eye, and further processed in the brain to provide the familiar human experience of normal vision. The mass of detail forming the visual scene consists of variations in light and shade, colour and, because we have two eyes, perspective.

Picture transmission, using electronic means to convey information of a scene, cannot be carried out as a simultaneous process embracing the total field of view. Any telecommunication system can process only a single item of information at a time, and hence the data relating to any visual scene must be analysed in such a way that the complete scene can be transmitted as separate items of electrical information. At the receiver, the individual bits of information are recovered and processed for display.

1.1 Scanning and aspect ratio

The visual scene is explored by examining the small areas of detail that are contained in it, a process known as *scanning*. When we read the page of a book, our eyes scan it line by line to extract the total visual information. Electronic scanning carries out a similar line by line scan process, the detail encountered being translated into voltage variations that can be used to modulate a radio transmitter. At the receiver, the received signals are demodulated and used to vary the beam current(s) of a display tube, the beam of which is sweeping in synchronism with the transmitter scanning beam.

A constraint of the electronic scanning system is the need to put a *frame* round the field of view to be transmitted. In the human seeing process the eye is quite unrestricted in its movements, and it roams freely over a very wide angular range which, with head and body movements, provides an unlimited field of view. In the electronic process, a finite limitation must be imposed by means of a frame, within which the picture can be analysed line by line.

The cinema industry has been in the business of presenting pictures as visual information for very many years, and has established a large number of basic principles. The newcomers to the art are the television engineers who did not 're-invent the wheel' but, very wisely, adopted many of the principles and standards that have evolved in film presentation. One of these concerns the shape of the frame. In the film industry, a standard rectangular shape with an *aspect ratio* (ar) of 4 (horizontal):3 (vertical) is the norm. This standard is still in general use for the main products of the film industry, despite the use of various 'wide screen' and other ratios. If a system has a standard aspect ratio at both the transmitter and the receiver, picture size is irrelevant. The relative dimensions of objects in the field will be correct.

The first television engineers concerned with the need to establish standards had no reason to depart from the 4:3 ar, particularly as it was realized that film would constitute a large proportion of programme material. These engineers were the team that created the standards for the world's first regular broadcast service of television programmes in 1936. This ratio has been adopted by all the systems that followed, and only recently has a change been considered in the 4:3 ratio.

Earlier experimental systems by Baird scanned the frame vertically with an aspect ratio of 1:2.

1.2 Still and moving pictures

A variety of picture transmission methods have been in use for many years. Still pictures have been communicated over telecommunication links by means of a *facsimile* (*fax*) system. The picture to be transmitted is wrapped round a drum, and scanned line by line as the drum is rotated and advanced with each turn. At the receiver, a photo-sensitive paper, wrapped round a similar drum, is synchronously rotated past a light beam which is modulated by the received signals. A high-quality 250 × 200 mm picture can be transmitted over a voice communication circuit in about 12 minutes. (Some modern fax scanners use linear flat scanning, similar to an office duplicator.)

The difference between scanning and transmission of a still or moving picture is one of time. Transmission of a still picture can take as long as we wish, but a moving field of view must be totally scanned in a time that is very short compared with the time being taken by any movements in the field of view. In other words, complete scanning of the moving picture must be so fast that we are concerned with what is, virtually, a still picture.

One of the properties exhibited by the human eye is *persistence of vision*. When the image of a still picture is impressed on the eye, removal of the visual stimulus does not result in an immediate cessation of the signals passed to the brain. An exponential lag takes place with a relatively long time for a total decay of the image. The cinema exploits this effect by presenting to the eye a succession of still pictures (or *frames*) one after the other, each frame differing from the previous one only by the change in position of any moving objects in the field of view. The presentation of one frame after another must not allow time for the image decay to become obvious and, provided that the presentation is sufficiently rapid and not too bright, an illusion of continuous movement is maintained.

The still pictures are projected in succession onto a screen. A frame is drawn into position with light cut off by means of a rotating shutter. As the shutter opens, the frame is stationary and the projected image illuminates the screen. The shutter cuts off the light, the next frame is drawn into position, the light is re-exposed through the frame, and so on in a continuous sequence. A great deal of early work with film showed that, for most people, a projection rate as low as 10–12 frames/second is adequate to present a complete illusion of movement.

However, at this projection rate another property of the eye becomes significant. The eye is extremely sensitive to the interruption of light at this rate, and the viewer would be very aware of *flicker*. As a consequence, the standard rate adopted for projection was 16 frames/second, well above that necessary for the presentation of continuous movement and, with the low-level illuminants of those days, adequate to minimize any awareness of flicker.

The eye sensitivity to flicker is a function of picture brightness and interruption rate, the rate needing to be higher if brightness is increased. With improvements in projector lamps over the years, flicker became a problem. Raising the projection rate would reduce flicker, but would result in a larger quantity of expensive film being required. An ingenious solution of this problem became a standard feature of all film projection systems. The film is drawn into position with the light cut off by the shutter, as previously explained. The shutter exposes the light through the film, cuts it off, re-exposes the light through the same frame, cuts it off and only then is the next frame drawn into position for the next double exposure. The light is projected and cut off twice during each frame. For a picture projection rate of 16 frames/second, the interruption frequency is raised to 32 per second, and the visibility of flicker is very much reduced without doubling the length of film.

As time passed, better illuminants came into use, and flicker reappeared. This problem was solved along with another concerned with the sound track that films now required. Film was not passing through the projector system fast enough for good sound quality. The standard was changed to that in use today. The frame rate was raised from 16 to 24 frames/second. This raised the interruption frequency to 48, and increased the sound track length by 50 per cent.

1.3 Television picture frequency

The engineers who had the task of establishing standards for the first broadcast system (system A, see *Table 1.1*) adopted the aspect ratio of 4:3, but were concerned by the film projection rate of 24 frames/second. It was feared that, with a 50 Hz power supply frequency, any residual 50 Hz or 100 Hz power supply ripple in the receiver might modulate the beam current of the display tube with a sub-harmonic at 25 Hz which would produce a visible 'bar' across the picture. If film was being transmitted at the film standard rate of 24 frames/second, the difference frequency of 1 Hz would result in the bar sweeping down the picture once per second.

It was decided that the picture rate would be 25 pictures/second instead of the film rate of 24. It was considered that the effect on sound would not be serious. It was further considered that, in the event of interference from the power supply, a stationary bar across the picture would be less offensive than a bar sweeping down the picture at the 'beat' frequency of 1 per second.

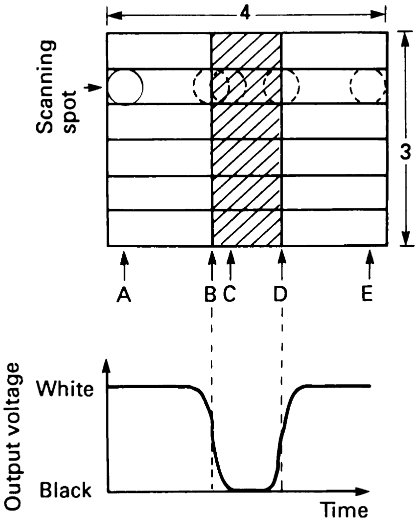


Figure 1.1 A six-line scan of a 4:3 field, showing aperture distortion

Figure 1.1 shows a simple six-line picture consisting of a black bar on a white background and the voltage output from the scanning system during a one-line scan. The scanning spot has a diameter equal to the width of one line, and sweeps across the picture from right to left along line 1, returns to its starting point displaced vertically by one line, sweeps line 2, and so on down the field of view to line 6. It is then returned to the top of the picture for the second picture scan, and so on. The voltage output may be as shown with maximum voltage indicating peak white, and minimum voltage corresponding to black, termed *positive modulation*. In a system, the polarity may be inverted, so that minimum voltage equals white and maximum is black. This would be *negative modulation*, as exemplified in the current UK system I (see *Table 1.1*). This method of scanning is termed *sequential scanning*, and results in a very obvious flicker because the field rate and picture frequency are the same, i.e. 25 per second.

A scanning system was standardized which provided a similar effect to the double-shuttering used in film projection. Instead of scanning the lines in sequence, the picture field is scanned by one field using lines 1, 3 and 5, and a second scan fills the gaps by re-scanning the field using lines 2, 4 and 6. This is *interlaced scanning*, and constitutes two sweeps of each field for a complete picture field. This has the same effect as double-shuttering of film, and raises the flicker frequency to 50 per second.

The interruption rate imposes a limitation on the brightness level at which the display tube can be operated before flicker becomes visible. The relationship between flicker and brightness is expressed by the Ferry-Porter law:

$$f_c = F + 12.6 \log_{10} B$$

where f_c is the critical frequency below which flicker is observed, F is a constant related to viewing conditions, and B is the luminance of picture highlights.

Tests on the viewing conditions of television pictures have suggested a value of about 37 for F and, with $f_c = 50$ (as in all European and some other systems), a picture highlight value of about 10 foot-lamberts is obtained.

Television standards in the USA adopted the same general principles. The picture rate was related to a power supply frequency of 60 Hz, resulting in a picture rate of 30 per second and a light-interruption rate of 60 per second. This increase of interruption frequency, compared with the UK rate, results in a permissible increase in highlight value by 6.8 times.

The six-line system illustrated in *Figure 1.1* would have very poor picture quality. The picture has a sharp transition at the edges of the black bar, but the voltage output does not change instantly from the white value to the black value. At position A (*Figure 1.1*) the scanning spot 'sees' peak white. As the spot reaches the bar at B it 'sees' half white and half black, the output being a half of the peak value, as shown. The output only reaches the value due to black at position C. At D, the half value is derived as shown, and the remainder of the scan gives a white output. The resulting effect is termed *aperture distortion*; it prevents any small detail in the picture being reproduced accurately.

In any practical television system, the picture quality will depend on the ability of the system to reproduce at the display tube all the sharp edges and fine detail. This requires the scanning spot size to be reduced, with a consequent increase in the number of lines necessary for a complete scan of the field of view. It is shown in section *1.5* that the channel bandwidth is determined by the scanning spot size. The smaller the spot, the more lines that are required for a complete scan, and channel space is at a premium. This means that a 'standard' spot size has to be a compromise between the ability of the system to provide a picture quality that is acceptable, and the minimum demand for channel space. Such standards are quoted in terms of the number of lines that are required in the vertical dimension for complete scanning of the entire field of view.

1.4 The video signal

The video signal derived by the camera or other scanning device for a practical black and white system will consist of random voltages generated by the scanning of black, white and grey images during the line scan. A possible scan output is shown in *Figure 1.2*.

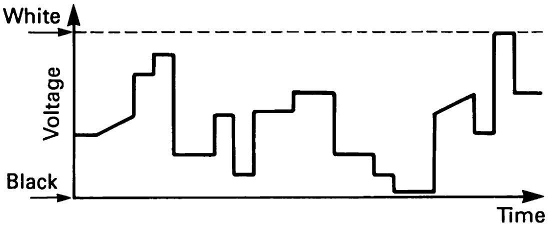


Figure 1.2 A possible video output from a camera during a one-line scan

Two important conclusions arise from consideration of this type of voltage waveform:

- Voltage variations will generally consist of 'step' changes from one value to another. Smooth transitions from white to black, or from black to white, will be rare.
- AC voltage variations are extremely unlikely, and their rare appearance might arise from a scan across regular bars, such as the black and white bars on a test card.

Figure 1.3 illustrates another important feature that arises from this type of waveform. The two line scans each show the same signal voltage variation. In (a) a white bar is shown on a grey background, and in (b) the same output voltage variation shows a grey bar on a black background. The difference between the two identical signal variations is due to there being, in each, an average dc voltage component. The dc·level determines picture brightness.

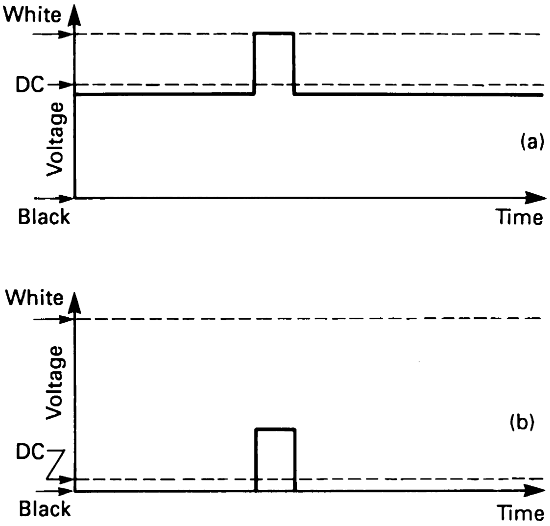


Figure 1.3 Similar video output signals are shown in (a) and (b), but with different dc levels

The video signals resulting from the scanning operation are processed and used to amplitude modulate the transmitter output. All standard broadcast transmitters use am, but fm is employed for certain links, and for satellite systems.

The principles of amplitude modulation are well known, but there are some important differences between sound and video as modulating signals. *Figure 1.4*(a) shows an alternating current variation that might be measured in the antenna system of a sound transmitter. Initially, the carrier is not modulated, then one cycle of an audio modulating tone is applied. The familiar features of this process are:

- The unmodulated carrier is radiated, and its mean level is constant with or without modulation being present.
- The carrier peak level is varied at the modulating frequency. The audio variation in carrier peak values during modulation is termed the *envelope*.
- An obvious limit exists in the modulating process whereby the carrier peak must not exceed twice the unmodulated level if distortion due to *clipping* is to be avoided.

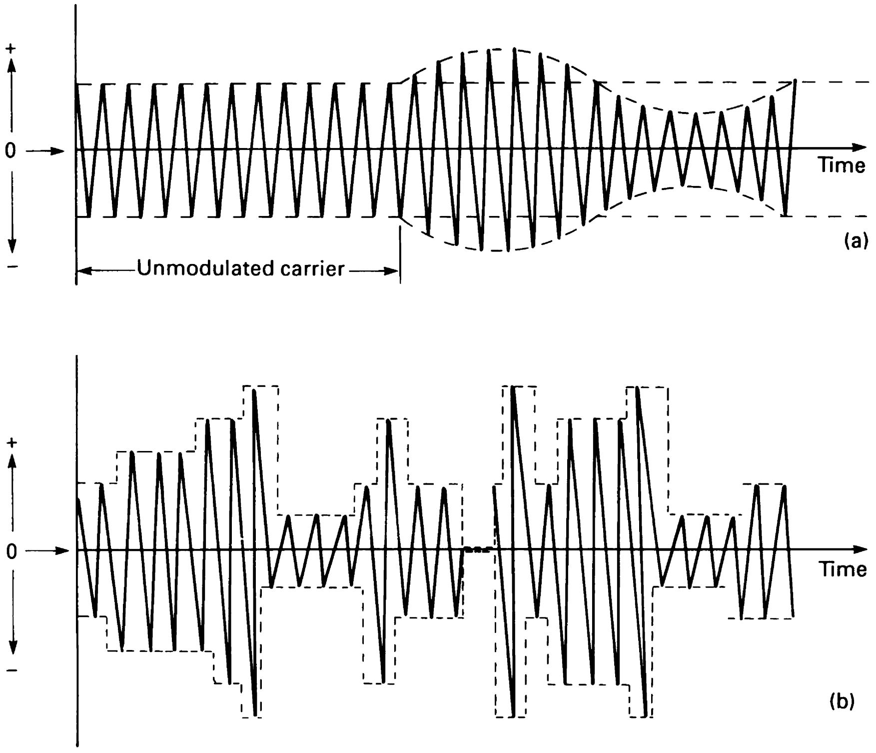


Figure 1.4 A carrier is shown amplitude modulated (a) by an audio tone and (b) by a video signal

Figure 1.4(b) shows a similar situation; it is amplitude modulation by a video signal similar to *Figure 1.2*. The envelope is now of a random character, and there is no mean carrier of constant level during modulation. When no modulating signal is present, no transmitter output is radiated.

1.5 Channel bandwidth

If a carrier is amplitude modulated with, for example, a 1 kHz tone, three frequencies are produced: the carrier, a frequency lower than the carrier by 1 kHz (a *lower side frequency*), and a frequency higher than the carrier by 1 kHz. If the modulating signal is a band of frequencies such as voice, music or video, a band of frequencies is generated each side of the carrier, termed *sidebands*, extending on each side of the carrier frequency to a limit determined by the highest frequency in the range of modulating frequencies.

There are several ways in which the highest frequency component in a modulating video signal may be determined, and thereby the channel bandwidth. One is indicated in *Figure 1.5* in which (a) is the top left-hand corner of a picture consisting of a regular pattern of alternate black and white squares. The sides of the squares are equal in length to the scanning spot diameter, and consequently the output signal generated will be a sine-wave, the frequency being the highest that will be generated at full amplitude. Any detail of smaller dimensions will not generate maximum output.

Consider now the six-line picture discussed in section *1.3*, where this picture is of the pattern shown in *Figure 1.5*, i.e. alternate squares of black and white, each with a width and height equal to the scanning spot diameter. The total number of squares in a 4:3 ar frame will be (6 × 6)4/3=48. If we transmit 25 complete pictures per second, the squares will be scanned in 1/25 s so that 48 × 25 = 1200 squares will be scanned in one second.

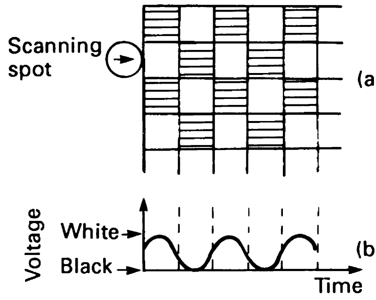


Figure 1.5 The smallest detail than can be resolved at full amplitude

The resulting ac cycle corresponds to the scan of one black and one white square. Thus the highest modulating frequency is 1200/2 = 600 Hz, and the two sidebands would require an overall rf channel bandwidth of 2 × 600 = 1200 Hz.

We have seen that the picture quality would be poor, and severely lacking in the resolution of edges and fine detail. We know that, to improve resolution, we must reduce the scanning spot diameter and increase the number of scanning lines appropriately. For example, if we halved the spot diameter of our *Figure 1.1* model, we would have to double the number of lines for a complete picture scan, and then the total number of squares would be 12 × 12 × $^4/_3$ = 192 and the number of squares scanned in 1 s would be 192 × 25 = 4800. This would result in a channel bandwidth of 4.8 kHz, four times that of the original six-line system. The bandwidth increases as the square of the number of lines.

A standard adopted for a good quality television system has to be a compromise between the need for acceptable definition of edges and fine detail in a picture and the overall bandwidth of the channel.

There are two historical examples of 'line standards' that are worth consideration. The first UK television broadcast system developed by Baird used an aspect ratio of 1:2 and 12½ pictures per second. Scanning was sequential and vertical. Thus, the number of squares would be (30 × 30) × 2 = 1800, and the number scanned in 1 s was 1800 × 12.5 = 22 500. The highest modulating frequency was thus 11 250 Hz.

The second example was the first ever 'high definition' broadcasts that commenced in 1936 in the UK with a picture frequency of 25 per second, interlaced scanning and an aspect ratio of 4:3. The compromise on definition and bandwidth was decided on the basis of the scanning spot being of such a size that 405 lines would be required to cover the picture area, and give acceptable picture quality. The highest video frequency is $405 \times 405 \times \frac{4}{3} \times \frac{25}{2} = 2.7$ MHz. The output of detail smaller than 1/405 of picture height would be less than maximum. Detail generating 3.0 MHz, for example, would be about -3 dB. The total video channel rf bandwidth becomes 3.4 MHz and the lowest modulating frequency is zero or dc. The total radio spectrum space occupied by this first *system A* (as it became known) is shown in *Figure 1.6*.

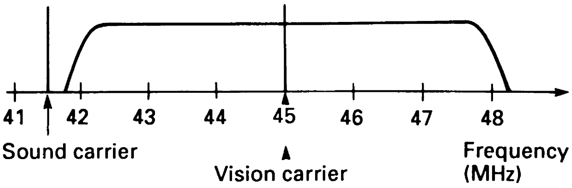


Figure 1.6 The full bandwidth of a dsb system A channel

1.6 Synchronism between scanning systems

The waveform in *Figure 1.4*(b) could represent the transmitter modulation during a one-line scan. At the receiver, the modulating signal is recovered and, after suitable processing, is used to modulate the beam current of the display tube, thus re-creating the detail seen by the scanner during the one-line scan.

Two further important items of information are necessary to ensure that the scanning spot at the transmitter, and the beam position on the face of the display tube, occupy identical positions in their 4:3 frame. One determines the position of the scanning spot in the vertical plane, and the other ensures the correct position in the horizontal plane.

It would appear from *Figure 1.4*(b) that the video waveform is very complete, and there is no way in which any additional information can be provided, but a development of system A showed how it could be done, in a manner that forms part of any television standard today. *Figure 1.7*(a) shows how the video signal modulation can be established between the two limits: the maximum transmitter output, and the carrier level corresponding to black. This leaves a region between zero carrier output and the black level into which we can put extra information.

However, the entire transmitter/receiver system can deal with only one bit of information at any instant, and we must remove the video information while we provide any extra information. A *blanking pulse* blanks out all video information down to the black level, at the start of the line scan. A narrow *line synchronizing pulse* is then inserted from the foot of the blanking pulse down to zero. The leading edge of this pulse is used to start the line scan on its traverse across the field from left to right. At the receiver, the leading edge of the demodulated narrow pulse is used to start the sweep of the display tube beam across the face of the tube. At the end of the line scan, the next blanking/synchronizing pulse triggers the return of the spot to the left-hand side, and the scanning cycle starts again, this time slightly displaced vertically, to trace a new line path.

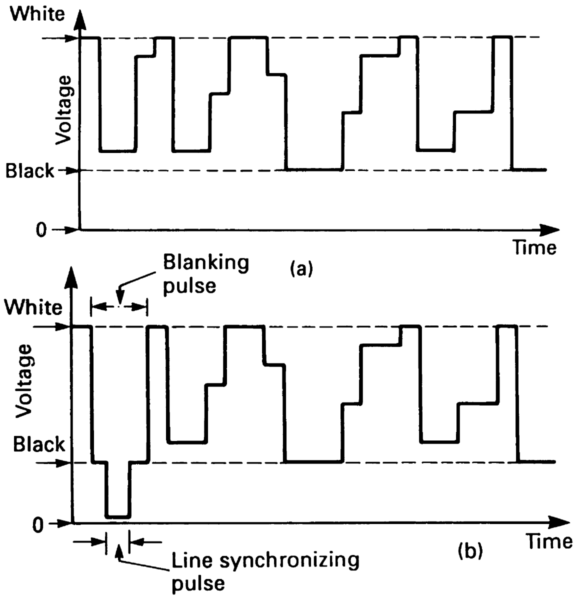


Figure 1.7 A typical camera output (a) before and (b) after the insertion of blanking and line sync pulses

The waveform of a line of the UK system I is shown in *Figure 1.8*. The heavy black line is the waveform of a black and white system, and the shaded areas are concerned with the colour information. The timing of all the pulses and other features is with reference to the leading edge of the narrow sync pulse. All television standards use the same type of waveform, but their timing and pulse widths may differ.

To effect vertical displacement of the scanning spot in synchronism, it is necessary to use a different type of pulse. The line sync pulses are narrow and relatively infrequent, so we can suppress video information for several lines, and transmit either a single, long pulse or a train of relatively broad pulses to effect vertical synchronism. *Figure 1.9* shows the four-field sequence of field sync pulses used in the UK system I. The 'burst' sequence is necessary for the PAL colour system.

At the receiver, the two types of pulse can be separated from the composite waveform by relatively simple forms of amplitude discrimination, and they can then be separated from each other by the passive circuitry shown in *Figures 1.10* and *1.11*.

Figure 1.10 shows one form of a *differentiation circuit* in which the time constant *CR* is short compared with *t*. It will be recalled that synchronizing information is required from the leading edge of the line sync pulse if all the timing is to be correct, and this type of circuit provides this discrimination between leading and lagging edges.

Figure 1.11 shows an *integrating circuit*. The time constant *CR* is long compared with the duration of the train of pulses provided at the input. Successive pulses build up the voltage on C to the value V_p and, at the end of the pulse train, C

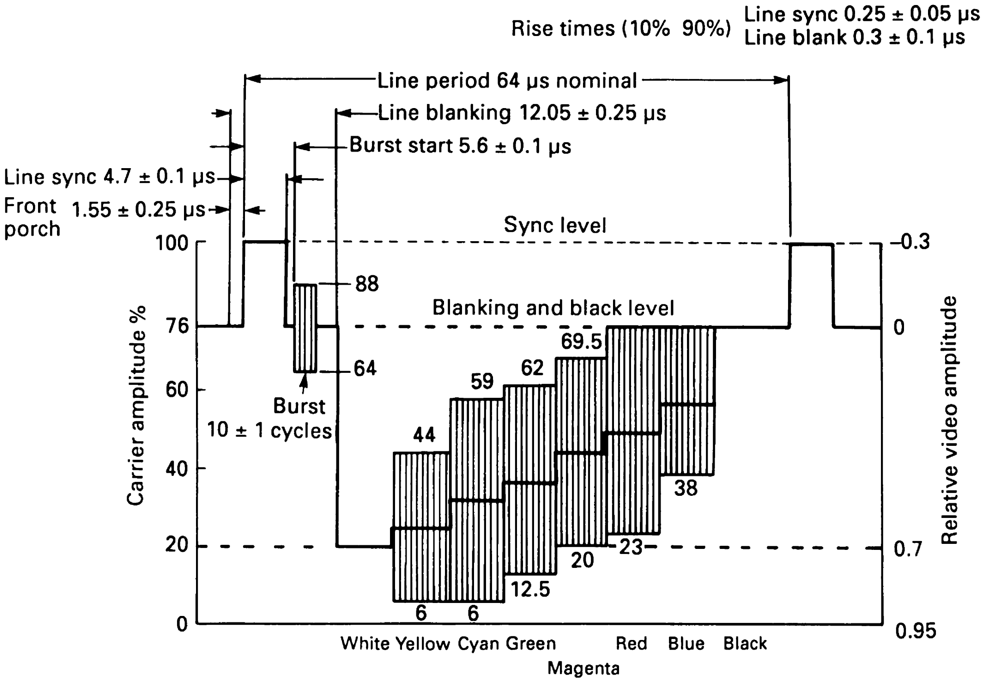


Figure 1.8 The thick line show a possible signal output resulting from a single line scan of a black and white system, the image being a grey 'staircase' from white to black. The shaded areas are concerned with the addition of colour (BBC)

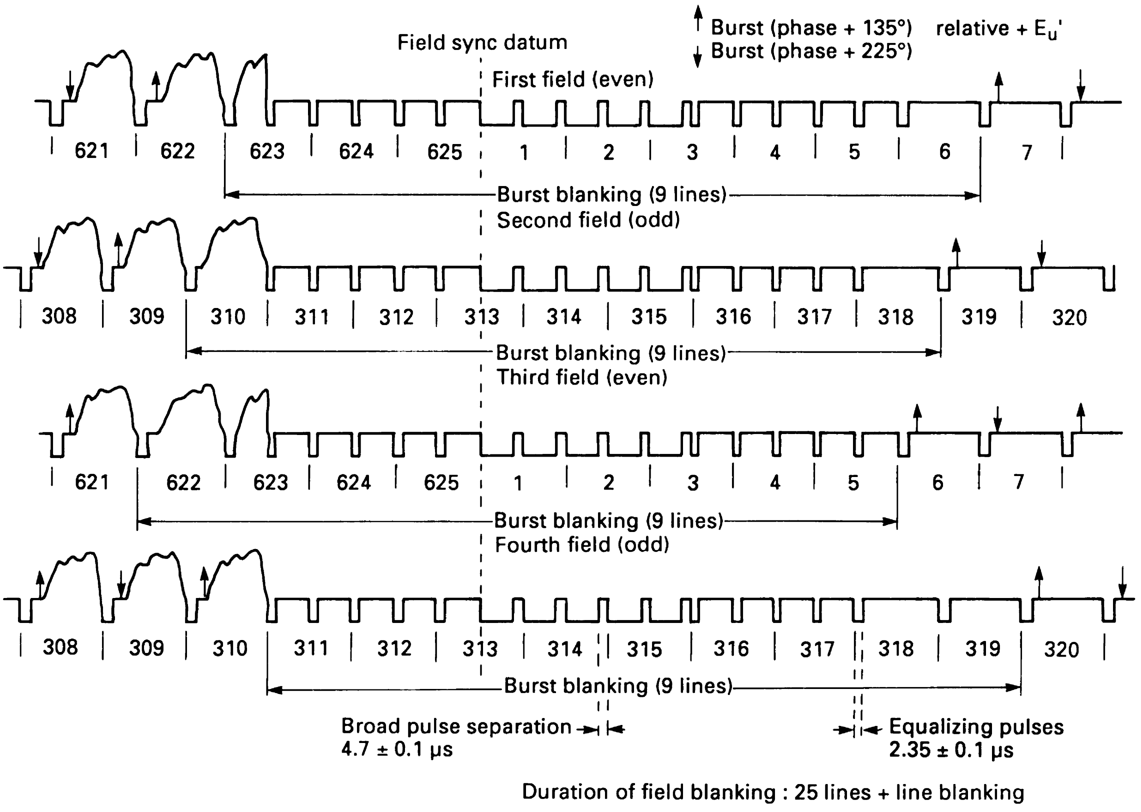


Figure 1.9 The broad pulses used for synchronism of the vertical scan during blanking of 25 lines

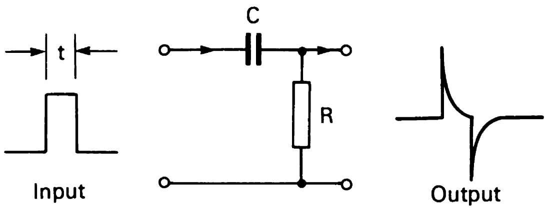


Figure 1.10 A differentiation circuit that provides an output when the input voltage changes in value

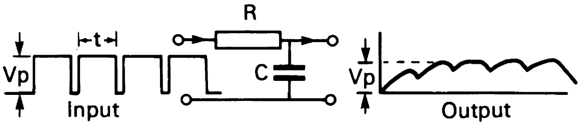


Figure 1.11 An integration system that enables a train of broad pulses to build up to a peak value

discharges. The output waveform thus constitutes a single broad pulse, which operates at a relatively slow speed to trigger the vertical sweep system.

The differentiation circuit does not distinguish between line or field pulses; it will generate 'spike' pulse output from the edges of any pulse fed into it. The integration circuit is the one that *does* discriminate between the two types of pulse. The narrow, infrequent line sync pulses will provide no output from the integrator.

1.7 Porches

Figure 1.8 shows two other features that are concerned with the line sync pulse region. Note that the pulse is not in the centre of the blanking interval.

In any system that contains inductance, capacitance and resistance, voltage or current changes cannot take place instantaneously. A video signal change from black to peak white, or white to black, takes time for completion, requiring possibly a capacitor to charge or discharge. A black/white edge results in a voltage change as shown in *Figure 1.12* (not to be confused with the aperture distortion shown in *Figure 1.1*). Ahead of the pulse in *Figure 1.8* is a narrow plateau, the *front porch*. Its purpose is to allow the video signal of the previous line (which might have been at peak white, for example) to reach black level before the sync pulse starts.

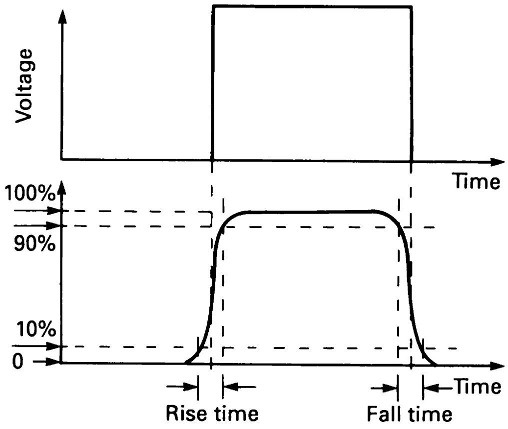


Figure 1.12 Voltage and current values cannot change instantly. Rise or fall times are determined by the values at 10 per cent and 90 per cent of peak values

Behind the sync pulse is another plateau, the *back porch*. The original purpose of the back porch, when first used in system A, was to make sure that there was plenty of time for the receiver line scan circuitry to effect complete retrace of the line scan, and for the beam to be in its correct position for commencement of the next line scan. By comparison with current technology, receiver scan circuits in 1936 were crude, ponderous and extravagant of power, and they required a lot of time for retrace. Today, the back porch provides plenty of time for retrace and it also provides the space necessary for colour information to be transmitted and extracted at the receiver.

1.8 DSB, ssb, asb and vsb

The UK system A operated from 1936 to 1939, and then had to close down due to the outbreak of World War II. The USA standardized its television system and immediately found a serious problem. The double sideband (dsb) system was very extravagant of channel space and, as many channels were required, a new system of modulation was devised. It was known as a *vestigial sideband* (vsb) system, sometimes termed an *asymmetric sideband* (asb) system, and is now used throughout the world because it saves a considerable amount of channel space.

To appreciate the operation of vsb, it is necessary to digress into some important differences between sound and video amplitude modulation. Single sideband (ssb) am systems have been used for radio transmission of speech and music for at least 60 years. Why cannot ssb be used for video modulation and thus save half the dsb bandwidth?

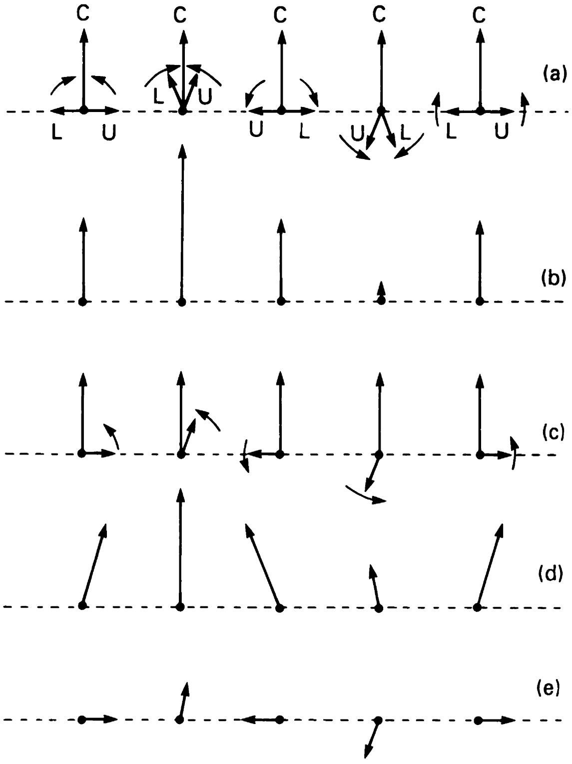


Figure 1.13 The vector relationships of dsb (a) and (b) ssb (c) and (d), and a single side frequency (e)

There are two sources of distortion in a dsb speech or music system. An amplitude constraint exists whereby any attempt at over-modulation results in severe waveform distortion, with unacceptable audible quality. The second effect occurs in single-channel sound reproduction where the quality of the reproduced sound is unaffected by changes of phase response. Many do not accept that our normal binaural experience distorts. Nevertheless, simply turning one's head, for example, results in a severe phase change of, say, a 5 kHz tone, where about 4 cm movement represents about 180°. We live with this effect and do not notice it, even when we use it for directional information.

Figure 1.13(a) shows the vectors relating to the carrier side frequencies during one cycle of a modulating tone of dsb am at maximum modulation; (b) shows the resultant vector addition of the vectors in (a), causing the rise and fall of carrier amplitude, and the phase of the resultant vector, which remains that of the carrier at all times during the modulating cycle.

Figure 1.13(c) is as (a), but with the lower side frequency suppressed, and (d) is the new resultant. Two features can be seen:

- The modulation, considered in terms of carrier amplitude, is halved.
- The angle of the resultant vector is now swinging with respect to the carrier vector.

We have produced phase modulation.

To determine the requirements of a video signal used for amplitude modulation, consider the simple video modulating signal shown in *Figure 1.14*(a). Such a pulse can be analysed into a number of discrete harmonic components, which extend out to infinity with a descending order of amplitude. The square waveform of (a) could be synthesized by addition of the frequency components of (b) at the correct amplitudes and phases. In a television system, we require the pulse signal to progress through the system in such a manner that the shape of the waveform that finally modulates the display tube beam current is a faithful copy of the signal derived by the scanning process at the transmitter.

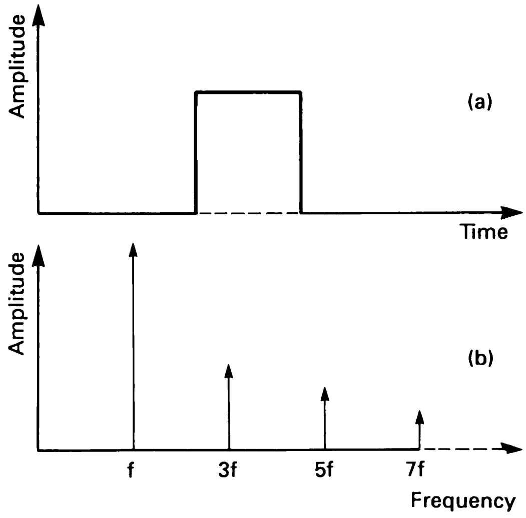


Figure 1.14 A theoretical rectangular pulse contains a fundamental frequency *f*, and all the odd harmonic frequencies. The practical pulse loses some of the higher harmonics due to the finite bandwith of the system, and has sloping sides with rounded corners

A pulse can experience degradation of its waveform in a number of ways as it goes through the system:

1 The bandwidth of the system is finite, and therefore some of the higher frequency harmonics will not be present.

2 An inadequate hf response at some point in the signal path can change the pulse shape by, for example, rounding the corners and sloping the sides of the pulse.

3 Inadequate lf response can produce a 'tilt' at the top of a pulse. The dc voltage may not hold up for the duration of the pulse.

4 The various discrete frequency components that constitute the signal waveform may experience differing transit times as they travel through the system. Some may go through faster or slower than others, distorting the wave shapes in varying degrees at the point where they are intended to arrive together, i.e. at the point where the display tube beam current is modulated.

Of the above, 1 is determined by a definition compromise as discussed in section *1.5*, and 2 and 3 are concerned with circuit behaviour, and any deficiencies can be resolved. The timing effect of 4 is much more important. The phase response of the system, clearly, determines the preservation of pulse shape. Each of the signal frequency components must travel through the system in the same time, although the actual time of transit is, within reason, of no importance.

If a fundamental frequency *f* passes through the system in time *t*, this time can be interpreted as a phase change ϕ. The frequency $2f$ will go through twice this phase angle in the same time. $3f$ will shift three times the fundamental phase angle, and so on for each harmonic component. If one of the harmonic components goes through too great a phase shift in the time *t*, it means that it is travelling too fast and will arrive early at the tube beam current control. It is seen that the phase response of the system should be such that the angle ϕ must be proportional to frequency. Any one of the lines shown in *Figure 1.15* would indicate a satisfactory phase response. The linear relationship is essential, and the only significance of the differing slopes is that they relate to different transit times. (The horizontal line would indicate zero time which is, of course, not possible.)

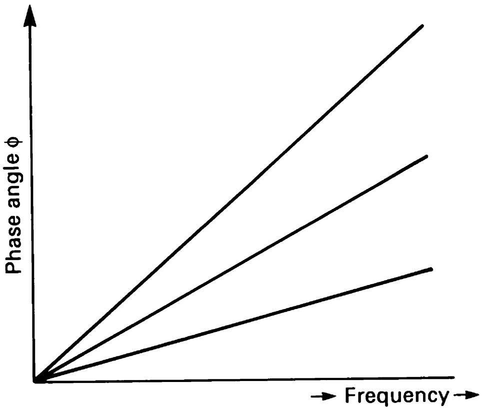


Figure 1.15 A video system phase response must be proportional to frequency

The main differences between an audio channel and video signal processing are now more clearly seen. Audio signal channels can tolerate a high degree of phase distortion, but amplitude distortion must be kept to very low limits. Video signal processing must ensure that phase distortion is minimized, but overload effects are far less serious. The peak value of *Figure 1.14* may represent peak white, for example, and any increase above this value would not be visually of much significance, although it may drive a transmitter into an overload condition. Alternatively, the peak value may correspond to a black signal, in which case 'blacker than black' is of no visual consequence.

The phase distortion produced by ssb (*Figure 1.13*(c) and (d)) may be acceptable for audio, but is quite unacceptable for video. However, it *is* possible to use partial suppression of one sideband (vsb) for a television system and save considerable channel space. Whereas phase distortion can be quite unacceptable where large picture areas are involved, phase errors become difficult to see on small detail in the picture. Thus, a practical system must have a very low phase distortion at low video frequencies, while the high frequencies generated by the fine detail in the picture can be severely distorted, but still acceptable.

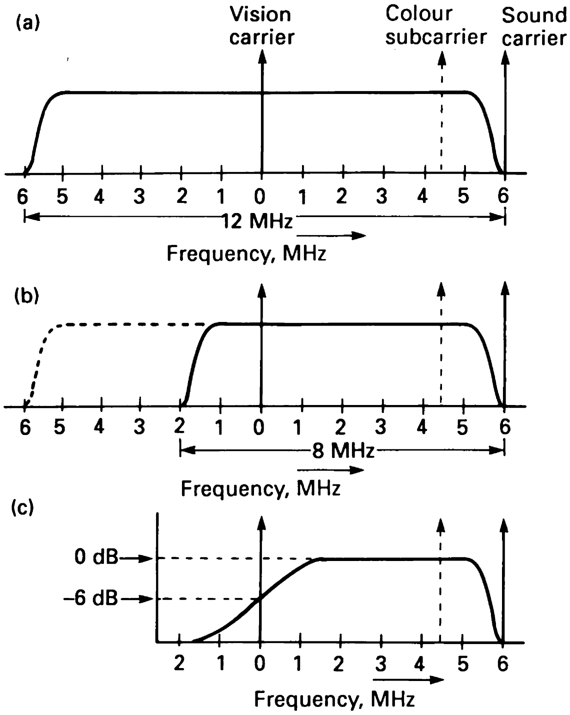


Figure 1.16 A full channel bandwith is shown with, at (b), the radiated signal of vsb modulation, and at (c), the required receiver if response

Figure 1.16 shows, at (a), the full dsb channel bandwidth that would result from video amplitude modulation, together with the sound channel, for the UK system I. The transmitted signals are shown in (b) where the lower sideband has had about 4 MHz filtered off, leaving a 'vestigial' 1.25 MHz. The 12 MHz channel has been reduced to 8 MHz.

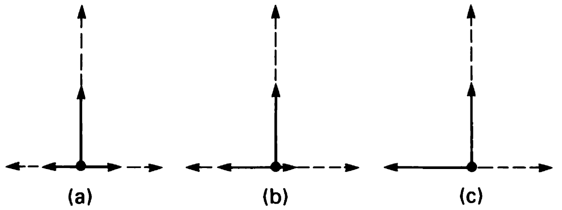


Figure 1.17 Demodulated output from the vsb input to a receiver. See description in text

For vsb to function correctly, it is necessary for the received signals to be processed before being presented to the demodulator. The pass-band of the receiver if channel must be shaped as shown in *Figure 1.16*(c). Consider, for example, a 50 Hz modulating frequency, and its position in *Figure 1.16.* Such a frequency would result from the scan of a large area, indeed a complete field, and the resulting side frequencies would be very close to the carrier. The signals presented to the demodulator by the receiver if amplifier are shown as vectors in *Figure 1.17*(a). The carrier and side frequencies are cut by 6 dB, but no phase distortion would result because the signal is a full double sideband am signal at 100 per cent modulation.

Consider next a modulating frequency resulting from the scan of a smaller detail, producing a frequency around, say, 500 kHz. *Figure 1.17*(b) shows the vector relationship between side frequencies and carrier. The overall length of the upper plus lower side vectors add to the same value as in (a), showing that the modulation factor has not changed. However, one vector is longer than the other, and some phase distortion is beginning to appear in the resultant addition of the three vectors. At frequencies of about 1.5 MHz and above, the vector situation is as shown in (c). The single side frequency resulting from the scan of a small detail has an amplitude equal to the attenuated carrier and, again, the level of modulation is unchanged. Phase distortion is present, but the detail is small and the errors become more difficult to see as the modulating frequency increases.

The US problem of minimizing channel space was eased considerably by the use of vsb, and all standard systems now use vsb. All UK system A stations that followed Channel I used vsb and, when the first station moved from its original site to Crystal Palace, the opportunity was taken to bring it into line and change it from dsb to vsb. The vestigial band was about 1 MHz, and the overall channel width was 5 MHz.

1.9 National standards

Many of the world's standards use 625 lines, the exceptions being the 525-line system M, used by North and South America, Japan and a few other countries, and the French system E on 619 lines. System E was developed before World War II in an attempt to provide a better picture quality than the 405-line standard then prevailing. This virtual doubling of the line structure does, indeed, provide an excellent picture quality, although at the expense of channel space. Consideration is currently being given to the adoption of new line standards that will approximately double the number of lines.

The almost universal adoption of a 625-line standard resulted from the first international conference on television standards after World War II. All countries planned to start a television service, and it was hoped that the 625-line standard would permit international links between European and other countries that used 625 lines. Unfortunately, widespread adoption of 625 lines did not result in a universal standard. There are many 625-line systems, but few are compatible. Some have positive modulation, others negative. Some have am sound, and others fm sound. There are further differences in the colour.

The US system M was the first practical broadcast colour system. A later European conference attempted to standardize a European colour system. It was considered that the NTSC system M could be improved, and several proposals were made for systems based on the NTSC principle that used a subcarrier for colour information. Germany proposed a modified NTSC system known as PAL (*Phase Alternation Line*). The French proposal was for the subcarrier to be frequency-modulated, the system being termed SECAM (*Sequentiel á Mémoire*). Other variants of the NTSC system were considered for a European standard, but the final outcome was that PAL was preferred

Table 1.1 Standard television systems

	A	*B*	*C*	*D(K)*	*E*	*G*	*H*	*I*	*L*	*M*	*N*
Lines per picture	405	625	625	625	819	625	625	625	625	525	625
Field frequency, Hz	50	50	50	50	50	50	50	50	50	60	50
Line frequency, kHz	10.125	15.625	15.625	15.625	20.475	15.625	15.625	15.625	15.625	15.734	15.625
Video bandwidth, MHz	3	5	5	6	10	5	5	5.5	6	4.2	4.2
Channel bandwidth, MHz	5	7	7	8	14	8	8	8	8	6	6
Sound/vision carrier spacing, MHz	3.5	5.5	5.5	6.5	11.15	5.5	5.5	6	6.5	4.5	4.5
Vestigial sideband width, MHz	0.75	0.75	0.75	0.75	2	0.75	1.25	1.25	1.25	0.75	0.75
Vision modulation polarity	+ ve	− ve	+ ve	− ve	+ ve	− ve	− ve	− ve	+ ve	− ve	− ve
Sound modulation	am	fm	am	fm	am	fm	fm	fm	am	fm	fm
Deviation, k Hz		50		50		50	50	50		25	25
Pre-emphasis, μs		50		50		50	50	50		75	75

and adopted by most countries, while France and Russia decided to use SECAM. (They actually have different versions of SECAM.)

A common feature of all colour systems is that they use a subcarrier for colour information, and transmit it in the form of a 'burst' of a few cycles for use by the receiver as a reference. The position for the burst is on the back porch, as shown in *Figure 1.8*.

Table 1.1 gives details of most of the standards in use throughout the world. (System A has been included, although it is no longer in use in the UK.) These are not the only differences between systems, but the radio frequencies on which they operate also differ between countries. The frequency allocations on which radio services operate are decided by the International Telecommunications Union (ITU). The world is divided by pole-to-pole boundaries into three regions:

Region 1 includes Europe, Africa and Russia. The eastern boundary includes the whole of Russia and Mongolia.
Region 2 includes North and South America, Greenland and Alaska.
Region 3 includes Australia, New Zealand, India and Pakistan, China and Japan.

This partition is intended to ensure that the various radio services should operate with minimum interference to each other and other services. However, this does not ensure that all television services will operate on the same frequency bands. Considerable differences exist, not only between regions, but within regions.

For example, *Table 1.2* shows the UK and European allocations for television services, but Africa and Scandinavia, also in Region 1, differ. South Africa has two bands, one on vhf 175–255 MHz and the other a uhf band on 471–632 MHz. The Danish bands are 55–68 MHz, 175–216 MHz and 615–856 MHz. Both countries use PAL. South Africa uses system I, as in the UK, and Denmark uses PAL on systems B and G which differ only in channel bandwidth.

Many similar examples can be quoted. A notable one is that Japan uses NTSC system M, but a receiver manufactured by the Japanese for home use could not be used in the USA because the two countries are in different regions and their radio frequencies are different.

Further examination of *Table 1.1* shows several very similar standards. Many have derived from basic systems, and differences are small. Systems D and K, for example, are the same for the listed parameters, but the D version uses PAL and the K version uses SECAM. The K system, in turn, has spawned the K′ system, which is similar to the K system but with a wider vestigial sideband.

1.10 Bands and channels

The radio frequency spectrum is divided up into bands as follows:

vlf	<30 kHz
lf	30–300 kHz
mf	300–3000 kHz
hf	3–30 MHz
vhf	30–300 MHz
uhf	300–3000 MHz
shf	3–30 GHz
ehf	30–300 GHz

The band classifications are international, and broadcasting takes place in some of them. Medium-wave broadcasting, for example, occurs in the mf band. The '3–30' range of each band is not so arbitrary as it may appear (see section *12.2.1*). Each band is characterized by its own propagation and antenna features.

Table 1.2 European bands and designations

Band	European band	Frequency range
vhf	I	41-68 MHz
vhf	II	88-108 MHz (fm sound)
vhf	III	174-216 MHz
uhf	IV	470-582 MHz
uhf	V	614-854 MHz
shf		11.7-12.5 GHz

Notes:
1. The UK used bands I and III for a television service, on standard A. This has now been discontinued, but other parts of Europe still use these bands for television.
2. Band II is not used for television, but for fm sound only.
3. The shf band has been allocated for satellite television services.

The sections of the spectrum that are allocated to television are in the vhf and uhf bands. There are three world zones with different frequency allocations for broadcasting, and *Table 1.2* shows the European bands and their designations.

The UK television service functions in Bands IV and V only. Channels 21–34 are in Band IV, and channels 39–68 are in Band V. The various standards in use have different channel bandwidths so, for example, a UK channel 40 does not necessarily occupy the same frequency space as a European channel 40.

Table 1.3 UK shf frequencies

SHF channel	Frequency GHz	
4	11.78502	all left-hand circular polarization
8	11.86174	
12	11.93846	
16	12.01518	
20	12.09190	

The actual frequencies in the shf band allocated to television broadcasting in the UK are shown in *Table 1.3*. There are 40 channels, each 20 MHz wide. The precise form of the transmission has yet to be decided.

1.11 Adding colour to a monochrome system

Colour television was demonstrated in 1928 by Baird, using red, green and blue filters before each field in a sequential manner, with appropriate synchronized filters before a modulated white light source at the receiver. Various unsatisfactory features appeared, such as *colour break-up* (a fast-moving object in the field was trailed by a colour), and flicker, requiring an increase in field rate.

A more sophisticated version of a mechanical sequential colour field system was proposed by CBS when the US sought a standard for a colour system, and this proposal was adopted in 1951, only to be subsequently rejected when the US National Television Standards Committee (NTSC) was formed to derive a wholly electronic standard. The NTSC system was adopted by the US in 1953, and had its first broadcast in 1955.

Modern colour television has resulted from the addition of suitable techniques to existing monochrome systems. The transmission, reception and processing of black and white signals require three separate items of information: the video information and the synchronizing information for two-dimensional scanning.

It is shown in Part *3* that colour requires three further types of information. It is necessary to determine *luminance*, *hue* and *saturation* of any colour at any instant. Fortunately, existing monochrome systems are luminance-only systems, and they can supply the luminance information. This leaves two extra items of information to be transmitted and received which, when suitably processed, can determine the hue and saturation encountered during scanning.

Apart from the obvious difficulties, some severe constraints exist on the addition of colour information:

- It must not require extra channel space.
- There must be no interference between the existing luminance information and the added information.
- It must be totally compatible, i.e. a monochrome receiver must display a colour transmission in black and white, and a colour receiver must be capable of displaying a monochrome transmission in black and white.

1.11.1 Colour bandwidth

There are some aspects of human colour perception that can be used to engineering advantage. One of these concerns the bandwidth required for colour display. Our ability to perceive fine detail in colour is considerably inferior to our awareness of fine detail as a brightness or luminance variation, thus colour information does not require the same wide bandwidth that is needed by a luminance system.

The eye can resolve small variations of luminance detail over the very small angle of 0.5–1.0′ (the precise figure varies with individuals). Visual acuity for colour is far less sensitive than for luminance, and depends to some degree on the range and saturation of the colours involved. Over an orange/cyan range of colours the resolution angle becomes 1.5–3.0′, and over a green/magenta range, detail resolution requires about 4–8′. The bandwidth required for colour information can thus be reduced to between $1/3$ and $1/8$ of the width required for luminance. Whatever methods we are going to use to process colour, it is only necessary to use about 1 MHz for the relatively low-detail colour information.

1.11.2 Spectrum utilization

Let us examine more closely the distribution of energy over the band of frequencies occupied by a carrier which is modulated by a video waveform. Without any video modulation, the carrier is already modulated by a number of frequencies. The line scanning rate is at a constant frequency (15 625 lines/second in system I), and it establishes side frequencies on each side of the carrier, spaced at line frequency and from each other, as shown in *Figure 1.18*. Interlaced scanning modulates the carrier at 50 Hz, and the side frequency of 50 Hz and the harmonics are spaced on both sides of the carrier and the line scanning harmonics. The picture frequency is another modulating signal at 25 Hz, and its energy joins the clusters around each line frequency harmonic.

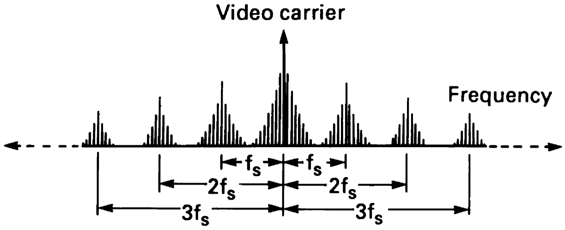


Figure 1.18 The signal components resulting from amplitude modulation. f_s is the line scanning frequency

The line frequency harmonics with their side frequencies extend out to the bandwidth limits of the video channel, and they are of a descending order of energy level as they become more remote from the carrier. Over the channel bandwidth, the energy generated across the band, without video modulation, is contained in clusters, spaced at line-scanning frequency from the carrier and from each other. Large regions of the channel space contain little or no energy.

In 1934, long before any broadcast television system existed, two mathematicians, Merz and Gray, studied the situation when video signals were included in the modulation process and concluded that the video energy simply joined the existing clusters round each line-scan harmonic, leaving the overall pattern substantially unchanged. The only differences between one picture and another, or whether one is moving or stationary, are relatively small variations in the magnitude of the energy in the clusters. The low or zero energy gaps between the line frequency harmonics remain with all types of video modulation.

The original work of Merz and Gray was directed at reducing interference between two adjacent transmitters, operating on nominally the same frequency. If their carrier frequencies were 'offset' by half the line frequency, the line harmonics of each would drop into the low-energy gaps of the other, thus minimizing mutual interference. However, in the years 1951–1953, the gaps were seen by the NTSC as the place for colour information.

A subcarrier can be inserted in the video frequency band, at a frequency about 1 MHz lower than the top of the video band, where the line-scan harmonics are low in amplitude. This subcarrier can then be modulated with colour information in such a manner that, when extracted and processed in the receiver, the colour display tube is driven with appropriate display information.

All existing standard colour systems are derived from the concept of the use of a subcarrier in the video band to supply colour information.

Acknowledgement

This section derives from a series of lectures given by the author during courses on Television Engineering, organized for the Royal Television Society. The RTS course lectures were arranged into a book *Television Engineering* and published by Pentech Press. We acknowledge permission given by Pentech Press for the use of some diagrams and text from the RTS book.

L.W.Turner FIEE,
Consultant Engineer

2

Quantities and Units

2.1 International unit system

The International System of Units (SI) is the modern form of the metric system agreed at an international conference in 1960. It has been adopted by the International Standards Organisation (ISO) and the International Electrotechnical Commission (IEC) and its use is recommended wherever the metric system is applied. It is now being adopted throughout most of the world and is likely to remain the primary world system of units of measurement for a very long time. The indications are that SI units will supersede the units of existing metric systems and all systems based on Imperial units.

SI units and the rules for their application are contained in *ISO Resolution* R1000 (1969, updated 1973) and an informatory document *SI-Le Système International d'Unités,* published by the Bureau International des Poids et Mesures (BIPM). An abridged version of the former is given in British Standards Institution (BSI) publication PD 5686 *The use of SI Units* (1969, updated 1973) and BS 3763 *International System (SI) Units;* BSI (1964) incorporates information from the BIPM document.

The adoption of SI presents less of a problem to the electronics engineer and the electrical engineer than to those concerned with other engineering disciplines as all the practical electrical units were long ago incorporated in the metre-kilogram-second (MKS) unit system and these remain unaffected in SI.

The SI was developed from the metric system as a fully coherent set of units for science, technology and engineering. A coherent system has the property that corresponding equations between quantities and between numerical values have exactly the same form, because the relations between units do not involve numerical conversion factors. In constructing a coherent unit system, the starting point is the selection and definition of a minimum set of independent 'base' units. From these, 'derived' units are obtained by forming products or quotients in various combinations, again without numerical factors. Thus the base units of length (metre), time (second) and mass (kilogram) yield the SI units of velocity (metre/second), force (kilogram-metre/second-squared) and so on. As a result there is, for any given physical quantity, only one SI unit with no alternatives and with no numerical conversion factors. A single SI unit (joule = kilogram metre-squared/second-squared) serves for energy of any kind, whether it be kinetic, potential, thermal, electrical, chemical..., thus unifying the usage in all branches of science and technology.

The SI has seven base units, and two supplementary units of angle. Certain important derived units have special names and can themselves be employed in combination to form alternative names for further derivations.

Each physical quantity has a quantity-symbol (e.g., *m* for mass) that represents it in equations, and a unit-symbol (e.g., kg for kilogram) to indicate its SI unit of measure.

2.1.1 Base units

Definitions of the seven base units have been laid down in the following terms. The quantity-symbol is given in italics, the unit-symbol (and its abbreviation) in roman type.

Length: *l;* metre (m). The length equal to 1 650 763.73 wavelengths in vacuum of the radiation corresponding to the transition between the levels $2p_{10}$ and $5d_5$ of the krypton-86 atom.

Mass: *m;* kilogram (kg). The mass of the international prototype kilogram (a block of platinum preserved at the International Bureau of Weights and Measures at Sèvres).

Time: *t;* second (s). The duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium-133 atom.

Electric current: *i;* ampere (A). The current which, maintained in two straight parallel conductors of infinite length, of negligible circular cross-section and 1 m apart in vacuum, produces a force equal to 2×10^{-7} newton per metre of length.

Thermodynamic temperature: *T;* kelvin (K). The fraction 1/273.16 of the thermodynamic (absolute) temperature of the triple point of water.

Luminous intensity: *I;* candela (cd). The luminous intensity in the perpendicular direction of a surface of 1/600 000 m^2 of a black body at the temperature of freezing platinum under a pressure of 101 325 newtons per square metre.

Amount of substance: *Q;* mole (mol). The amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kg of carbon-12. The elementary entity must be specified and may be an atom, a molecule, an ion, an electron, etc., or a specified group of such entities.

2.1.2 Supplementary units

Plane angle: α, β...; radian (rad). The plane angle between two radii of a circle which cut off on the circumference an arc of length equal to the radius.

Solid angle: Ω; steradian (sr). The solid angle which, having its vertex at the centre of a sphere, cuts off an area of the surface of the sphere equal to a square having sides equal to the radius.

Force: The base SI unit of electric current is in terms of force in newtons (N). A force of 1 N is that which endows unit mass (1 kg) with unit acceleration (1 $\mathrm{m/s^2}$). The newton is thus not only a coherent unit; it is also devoid of any association with gravitational effects.

2.1.3 Temperature

The base SI unit of thermodynamic temperature is referred to a point of 'absolute zero' at which bodies possess zero thermal energy. For practical convenience two points on the Kelvin temperature scale, namely 273.15 K and 373.15 K, are used to define the Celsius (or Centigrade) scale (0°C and 100°C). Thus in terms of temperature *intervals*, 1 K = 1°C; but in terms of temperature *levels*, a Celsius temperature *0* corresponds to a Kelvin temperature (0+273.15)K.

2.1.4 Derived units

Nine of the more important SI derived units with their definitions are given below.

Quantity	*Unit name*	*Unit symbol*
Force	newton	N
Energy	joule	J
Power	watt	W
Electric charge	coulomb	C
Electrical potential difference and EMF	volt	V
Electric resistance	ohm	Ω
Electric capacitance	farad	F
Electric inductance	henry	H
Magnetic flux	weber	Wb

Newton That force which gives to a mass of 1 kilogram an acceleration of 1 metre per second squared.

Joule The work done when the point of application of 1 newton is displaced a distance of 1 metre in the direction of the force.

Watt The power which gives rise to the production of energy at the rate of 1 joule per second.

Coulomb The quantity of electricity transported in 1 second by a current of 1 ampere.

Volt The difference of electric potential between two points of a conducting wire carrying a constant current of 1 ampere, when the power dissipated between these points is equal to 1 watt.

Ohm The electric resistance between two points of a conductor when a constant difference of potential of 1 volt, applied between these two points, produces in this conductor a current of 1 ampere, this conductor not being the source of any electromotive force.

Farad The capacitance of a capacitor between the plates of which there appears a difference of potential of 1 volt when it is charged by a quantity of electricity equal to 1 coulomb.

Henry The inductance of a closed circuit in which an electromotive force of 1 volt is produced when the electric current in the circuit varies uniformly at a rate of 1 ampere per second.

Weber The magnet flux which, linking a circuit of one turn, produces in it an electromotive force of 1 volt as it is reduced to zero at a uniform rate in 1 second.

Some of the simpler derived units are expressed in terms of the seven basic and two supplementary units directly. Examples are listed in Table 2.1.

Table 2.1 Directly derived units

Quantity	*Unit name*	*Unit symbol*
Area	square metre	$\mathrm{m^2}$
Volume	cubic metre	$\mathrm{m^3}$
Mass density	kilogram per cubic metre	$\mathrm{kg/m^3}$
Linear velocity	metre per second	m/s
Linear acceleration	metre per second squared	$\mathrm{m/s^2}$
Angular velocity	radian per second	rad/s
Angular acceleration	radian per second squared	$\mathrm{rad/s^2}$
Force	kilogram metre per second squared	$\mathrm{kg\ m/s^2}$
Magnetic field strength	ampere per metre	A/m
Concentration	mole per cubic metre	$\mathrm{mol/m^3}$
Luminance	candela per square metre	$\mathrm{cd/m^2}$

Units in common use, particularly those for which a statement in base units would be lengthy or complicated, have been given special shortened names (see Table 2.2). Those that are named from scientists and engineers are abbreviated to an initial capital letter: all others are in lower-case letters.

Table 2.2 Named derived units

Quantity	*Unit name*	*Unit symbol*	*Derivation*
Force	newton	N	$\mathrm{kg\ m/s^2}$
Pressure	pascal	Pa	$\mathrm{N/m^2}$
Power	watt	W	J/s
Energy	joule	J	N m, W s
Electric charge	coulomb	C	A s
Electric flux	coulomb	C	A s
Magnetic flux	weber	Wb	V s
Magnetic flux density	tesla	T	$\mathrm{Wb/m^2}$
Electric potential	volt	V	J/C, W/A
Resistance	ohm	Ω	V/A
Conductance	siemens	S	A/V
Capacitance	farad	F	A s/V, C/V
Inductance	henry	H	V s/A, Wb/A
Luminous flux	lumen	lm	cd sr
Illuminance	lux	lx	$\mathrm{lm/m^2}$
Frequency	hertz	Hz	l/s

The named derived units are used to form further derivations. Examples are given in Table 2.3.

Names of SI units and the corresponding EMU and ESU CGS units are given in Table 2.4.

2.1.5 Gravitational and absolute systems

There may be some difficulty in understanding the difference between SI and the Metric Technical System of units which has been used principally in Europe. The main difference is that while mass is expressed in kg in both systems, weight (representing a force) is expressed as kgf, a gravitational unit, in the MKSA system and as N in SI. An absolute unit of force differs from a gravitational unit of force because it induces unit acceleration in a unit mass whereas a gravitational unit imparts gravitational acceleration to a unit mass.

A comparison of the more commonly known systems and SI is shown in Table 2.5.

2.1.6 Expressing magnitudes of SI units

To express magnitudes of a unit, decimal multiples and submultiples are formed using the prefixes shown in Table 2.6. This method of expressing magnitudes ensures complete adherence to a decimal system.

Table 2.3 Further derived units

Quantity	*Unit name*	*Unit symbol*
Torque	newton metre	N m
Dynamic viscosity	pascal second	Pa s
Surface tension	newton per metre	N/m
Power density	watt per square metre	W/m^2
Energy density	joule per cubic metre	J/m^3
Heat capacity	joule per kelvin	J/K
Specific heat capacity	joule per kilogram kelvin	J/(kg K)
Thermal conductivity	watt per metre kelvin	W/(m K)
Electric field strength	volt per metre	V/m
magnetic field strength	ampere per metre	A/m
Electric flux density	coulomb per square metre	C/m^2
Current density	ampere per square metre	A/m^2
Resistivity	ohm metre	Ω m
Permittivity	farad per metre	F/m
Permeability	henry per metre	H/m

Table 2.4 Unit names

Quantity	*Symbol*	*SI*	*EMU & ESU*
Length	l	metre (m)	centimetre (cm)
Time	t	second (s)	second
Mass	m	kilogram (kg)	gram (g)
Force	F	newton (N)	dyne (dyn)
Frequency	f, v	hertz (Hz)	hertz
Energy	E, W	joule (J)	erg (erg)
Power	P	watt (W)	erg/second (erg/s)
Pressure	p	newton/metre^2 (N/m^2)	dyne/centimetre^2 (dyne/cm^2)
Electric charge	Q	coulomb (C)	coulomb
Electric potential	V	volt (V)	volt
Electric current	I	ampere (A)	ampere
Magnetic flux	ϕ	weber (Wb)	maxwell (Mx)
Magnetic induction	B	tesla (T)	gauss (G)
Magnetic field strength	H	ampere turn/ metre (At/m)	oersted (Oe)
Magnetomotive force	Fm	ampere turn (At)	gilbert (Gb)
Resistance	R	ohm (Ω)	ohm
Inductance	L	henry (H)	henry
Conductance	G	mho (Ω^{-1}) (siemens)	mho
Capacitance	C	farad (F)	farad

Table 2.5 Commonly used units of measurement

		SI (absolute)	*FPS (gravitational)*	*FPS (absolute)*	*cgs (absolute)*	*Metric technical units (gravitational)*
Length		metre (m)	ft	ft	cm	metre
Force		newton (N)	lbf	poundal (pdl)	dyne	kgf
Mass		kg	lb or slug	lb	gram	kg
Time		s	s	s	s	s
Temperature		°C K	°F	°F °R	°C K	°C K
Energy	mech.	joule*	ft lbf	ft pdl	dyne cm = erg	kgf m
	heat		Btu	Btu	calorie	kcal
Power	mech.	watt	hp	hp	erg/s	metric hp
	elec.		watt	watt		watt
Electric current		amp	amp	amp	amp	amp
Pressure		N/m^2	lbf/ft^2	pdl/ft^2	dyne/cm^2	kgf/cm^2

* 1 joule = 1 newton metre or 1 watt second.

Table 2.6 The internationally agreed multiples and submultiples

Factor by which the unit is multiplied		*Prefix*	*Symbol*	*Common everyday examples*
One million million (billion)	10^{12}	tera	T	
One thousand million	10^9	giga	G	gigahertz (GHz)
One million	10^6	mega	M	megawatt (MW)
One thousand	10^3	kilo	k	kilometre (km)
One hundred	10^2	hecto*	h	
Ten	10^1	deca*	da	decagram (dag)
UNITY	1			
One tenth	10^{-1}	deci*	d	decimetre (dm)
One hundredth	10^{-2}	centi*	c	centimetre (cm)
One thousandth	10^{-3}	milli	m	milligram (mg)
One millionth	10^{-6}	micro	μ	microsecond (μs)
One thousand millionth	10^{-9}	nano	n	nanosecond (ns)
One million millionth	10^{-12}	pico	p	picofarad (pF)
One thousand million millionth	10^{-15}	femto	f	
One million million millionth	10^{-18}	atto	a	

* To be avoided wherever possible.

2.1.7 Auxiliary units

Certain auxiliary units may be adopted where they have application in special fields. Some are acceptable on a temporary basis, pending a more widespread adoption of the SI system. Table 2.7 lists some of these.

Table 2.7 Auxiliary units

Quantity	*Unit symbol*	*SI equivalent*
Day	d	86 400 s
Hour	h	3600 s
Minute (time)	min	60 s
Degree (angle)	°	π/180 rad
Minute (angle)	′	π/10 800 rad
Second (angle)	″	π/648 000 rad
Acre	a	1 dam^2 = 10^2 m^2
Hectare	ha	1 hm^2 = 10^4 m^2
Barn	b	100 fm^2 = 10^{-28} m^2
Standard atmosphere	atm	101 325 Pa
Bar	bar	0.1 MPa = 10^5 Pa
Litre	l	1 dm^3 = 10^{-3}m^3
Tonne	t	10^3 kg = 1 Mg
Atomic mass unit	u	1.660 53 × 10^{-27} kg
Angström	A	0.1 nm = 10^{-10} m
Electron-volt	eV	1.602 19 × 10^{-19} J
Curie	Ci	3.7 × 10^{10} s^{-1}
Röntgen	R	2.58 × 10^{-4} C/kg

2.2 Universal constants in SI units

Table 2.8 Universal constants

The digits in parentheses following each quoted value represent the standard deviation error in the final digits of the quoted value as computed on the criterion of internal consistency. The unified scale of atomic weights is used throughout ($^{12}C=12$). C=coulomb; G=gauss; Hz=hertz; J=joule; N=newton; T=tesla; u=unified nuclidic mass unit; W=watt; Wb=weber. For result multiply the numerical value by the SI unit.

Constant	*Symbol*	*Numerical value*	*SI unit*
Speed of light in vacuum	c	2.997 925(1)	10^{8} m/s
Gravitational constant	G	6.670(5)*	10^{-11} N m^{2} kg^{2}
Elementary charge	e	1.602 10(2)	10^{-19} C
Avogadro constant	N_A	6.022 52(9)	10^{26} kmol^{-1}
Mass unit	u	1.660 43(2)	10^{-27} kg
Electron rest mass	m_e	9.109 08(13)	10^{-31} kg
		5.485 97(3)	10^{-4} u
Proton rest mass	m_p	1.672 52(3)	10^{-27} kg
		1.007 276 63(8)	u
Neutron rest mass	m_n	1.674 82(3)	10^{-27} kg
		1.008 665 4(4)	u
Faraday constant	F	9.684 70(5)	10^{4} C/mol
Planck constant	h	6.625 59(16)	10^{-34} J s
	$h/2\pi$	1.054 494(25)	10^{-34} J s
Fine-structure constant	$\propto$	7.297 20(3)	10^{-3}
	$1/\alpha$	137.038 8(6)	
Charge-to-mass ratio for electron	e/m_e	1.758 796(6)	10^{11} C/kg
Quantum of magnetic flux	hc/e	4.135 56(4)	$10\$^{11}$ Wb
Rydberg constant	R_∞	1.097 373 1(1)	10^{7} m^{-1}
Bohr radius	a_0	5.291 67(2)	10^{-11} m
Compton wavelength of electron	h/m_ec	2.426 21(2)	10^{-12} m
	$\lambda C/2\pi$	3.861 44(3)	10^{-13} m
Electron radius	$e^2/m_ec^2=r_e$	2.817 77(4)	10^{-15} m
Thomson cross-section	$8\pi r_e^2/3$	6.651 6(2)	10^{-29} m^{2}
Compton wavelength of proton	λc,p	1.321 398(13)	10^{-15} m
	λc,p$/2\pi$	2.103 07(2)	10^{-16} m
Gyromagnetic ratio of proton	γ	2.675 192(7)	10^{8} rad/(s T)
	$\gamma/2\pi$	4.257 70(1)	10^{7} Hz/T
(uncorrected for diamagnetism of H_2O)	γ'	2.675 123(7)	10^{8} rad/(s T)
	$\gamma'/2\pi$	4.257 59(1)	10^{7} Hz/T
Bohr magneton	μB	9.273 2(2)	10^{-24} J/T
Nuclear magneton	μN	5.050 50(13)	10^{-27} J/T
Proton magnetic moment	μp	1.410 49(4)	10^{-26} J/T
	$\mu p/\mu N$	2.792 76(2)	
(uncorrected for diamagnetism in H_2O sample)	$\mu' p/\mu N$	2.792 68(2)	
Gas constant	R_0	8.314 34(35)	J/K mol
Boltzmann constant	k	1.380 54(6)	10^{-23} J/K
First radiation constant ($2\pi hc^2$)	c_1	3.741 50(9)	10^{-16} W/m^{2}
Second radiation constant (hc/k)	c_2	1.438 79(6)	10^{-2} m K
Stefan-Boltzmann constant	σ	5.669 7(10)	10^{-8} W/m^{2} K^{4}

* The universal gravitational constant is not, and cannot in our present state of knowledge, be expressed in terms of other fundamental constants. The value given here is a direct determination by P.R. Heyland and P. Chrzanowski, *J. Res. Natl. Bur. Std.* (U.S.) 29, 1 (1942).

The above values are extracts from *Review of Modern Physics* Vol. 37 No. 4 October 1965 published by the American Institute of Physics.

2.3 Metric to Imperial conversion factors

Table 2.9 Conversion factors

SI units	*British units*
SPACE AND TIME	
Length:	
1 μm (micron)	= 39.37 × 10^{-6} in
1 mm	= 0.039 370 1 in
1 cm	= 0.393 701 in
1 m	= 3.280 84 ft
1 m	= 1.093 61 yd
1 km	= 0.621 371 mile
Area:	
1 mm^{2}	= 1.550 × 10^{-3} in^{2}
1 cm^{2}	= 0.155 0 in^{2}
1 m^{2}	= 10.763 9 ft^{2}
1 m^{2}	= 1.195 99 yd^{2}
1 ha	= 2.471 05 acre
Volume:	
1 mm^{3}	= 61.023 7 × 10^{-6} in^{3}
1 cm^{3}	= 61.023 7 × 10^{-3} in^{3}
1 m^{3}	= 35.314 7 ft^{3}
1 m^{3}	= 1.307 95 yd^{3}
Capacity:	
10^{6} m^{3}	= 219.969 × 10^{6} gal
1 m^{3}	= 219.969 gal
1 litre (l)	= 0.219 969 gal
	= 1.759 80 pint

Table 2.9 *continued*

SI units	*British units*
Capacity flow:	
10^3 m^3/s	= 791.9 × 10^6 gal/h
1 m^3/s	= 13.20 × 10^3 gal/min
1 litre/s	= 13.20 gal/min
1 m^3/kW h	= 219.969 gal/kW h
1 m^3/s	= 35.314 7 ft^3/s (cusecs)
1 litre/s	= 0.588 58 × 10^{-3} ft^3/min (cfm)
Velocity:	
1 m/s	= 3.280 84 ft/s = 2.236 94 mile/h
1 km/h	= 0.621 371 mile/h
Acceleration:	
1 m/s^2	= 3.280 84 ft/s^2
MECHANICS	
Mass:	
1 g	= 0.035 274 oz
1 kg	= 2.204 62 lb
1 t	= 0.984 207 ton = 19.684 1 cwt
Mass flow:	
1 kg/s	= 2.204 62 lb/s = 7.936 64 klb/h
Mass density:	
1 kg/m^3	= 0.062 428 lb/ft^3
1 kg/litre	= 10.022 119 lb/gal
Mass per unit length:	
1 kg/m	= 0.671 969 lb/ft = 2.015 91 lb/yd
Mass per unit area:	
1 kg/m^2	= 0.204 816 lb/ft^2
Specific volume:	
1 m^3/kg	= 16.018 5 ft^3/lb
1 litre/tonne	= 0.223 495 gal/ton
Momentum:	
1 kg m/s	= 7.233 01 lb ft/s
Angular momentum:	
1 kg m^2/s	= 23.730 4 lb ft^2/s
Moment of inertia:	
1 kg m^2	= 23.730 4 lb ft^2
MECHANICS	
Force:	
1 N	= 0.224 809 lbf
Weight (force) per unit length:	
1 N/m	= 0.068 521 lbf/ft
	= 0.205 566 lbf/yd
Moment of force (or torque):	
1 N m	= 0.737 562 lbf ft
Weight (force) per unit area:	
1 N/m^2	= 0.020 885 lbf/ft^2
Pressure:	
1 N/m^2	= 1.450 38 × 10^{-4} lbf/in^2
1 bar	= 14.503 8 lbf/in^2
1 bar	= 0.986 923 atmosphere
1 mbar	= 0.401 463 in H_2O
	= 0.029 53 in Hg
Stress:	
1 N/mm^2	= 6.474 90 × 10^{-2} tonf/in^2
1 MN/m^2	= 6.474 90 × 10^{-2} tonf/in^2
1 hbar	= 0.647 490 tonf/in^2
Second moment of area:	
1 cm^4	= 0.024 025 in^4
Section modulus:	
1 m^3	= 61 023.7 in^3
1 cm^3	= 0.061 023 7 in^3
Kinematic viscosity:	
1 m^2/s	= 10.762 75 ft^2/s = 10^6 cSt
1 cSt	= 0.038 75 ft^2/h
Energy, work:	
1 J	= 0.737 562 ft lbf
1 MJ	= 0.372 5 hph
1 MJ	= 0.277 78 kW h
Power:	
1 W	= 0.737 562 ft lbf/s
1 kW	= 1.341 hp = 737.562 ft lbf/s
Fluid mass:	
(Ordinary) 1 kg/s	= 2.204 62 lb/s = 793 6.64 lb/h
(Velocity) 1 kg/m^{2}s	= 0.204 815 lb/ft^{2}s
HEAT	
Temperature:	
(Interval) 1 K	= $^9/_5$ deg R (Rankine)
1°C	= $^9/_5$ deg F
(Coefficient) 1°R^{-1}	= 1 deg F^{-1} = $^5/_9$ deg C
1°C^{-1}	= $^5/_9$ deg F^{-1}
Quantity of heat:	
1 J	= 9.478 17 × 10^{-4} Btu
1 J	= 0.238 846 cal
1 kJ	= 947.817 Btu
1 GJ	= 947.817 × 10^3 Btu
1 kJ	= 526.565 CHU
1 GJ	= 526.565 × 10^3 CHU
1 GJ	= 9.478 17 therm
Heat flow rate:	
1 W(J/s)	= 3.412 14 Btu/h
1 W/m^2	= 0.316 998 Btu/ft^2 h
Thermal conductivity:	
1 W/m °C	= 6.933 47 Btu in/ft^2 h °F
Coefficient and heat transfer:	
1 W/m^2 °C	= 0.176 110 Btu/ft^2 h °F
Heat capacity:	
1 J/°C	= 0.526 57 × 10^{-3} Btu/°R
Specific heat capacity:	
1 J/g °C	= 0.238 846 Btu/lb °F
1 kJ/kg °C	= 0.238 846 Btu/lb °F
Entropy:	
1 J/K	= 0.526 57 × 10^{-3} Btu/°R
Specific entropy:	
1 J/kg °C	= 0.238 846 × 10^{-3} Btu/lb °F
1 J/kg K	= 0.238 846 × 10^{-3} Btu/lb °R
Specific energy/specific latent heat:	
1 J/g	= 0.429 923 Btu/lb
1 J/kg	= 0.429 923 × 10^{-3} Btu/lb
Calorific value:	
1 kJ/kg	= 0.429 923 Btu/lb
1 kJ/kg	= 0.773 861 4 CHU/lb
1 J/m^3	= 0.026 839 2 × 10^{-3} Btu/ft^3
1 kJ/m^3	= 0.026 839 2 Btu/ft^3
1 kJ/litre	= 4.308 86 Btu/gal
1 kJ/kg	= 0.009 630 2 therm/ton
ELECTRICITY	
Permeability:	
1 H/m	= $10^7/4\pi\ \mu_0$
Magnetic flux density:	
1 tesla	= 10^4 gauss = 1 Wb/m^2
Conductivity:	
1 mho	= 1 reciprocal ohm
1 siemens	= 1 reciprocal ohm
Electric stress:	
1 kV/mm	= 25.4 kV/in
1 kV/m	= 0.025 4 kV/in

2.4 Symbols and abbreviations

Table 2.10 Quantities and units of periodic and related phenomena (based on ISO Recommendation R31)

Symbol	*Quantity*
T	periodic time
τ, (T)	time constant of an exponentially varying quantity
f, ν	frequency

Table 2.10 *continued*

Symbol	*Quantity*
η	rotational frequency
ω	angular frequency
λ	wavelength
σ $(\bar{\nu})$	wavenumber
K	circular wavenumber
$\log_e (A_1/A_2)$	natural logarithm of the ratio of two amplitudes
$10 \log_{10} (P_1/P_2)$	ten times the common logarithm of the ratio of two powers
δ	damping coefficient
Λ	logarithmic decrement
α	attenuation coefficient
β	phase coefficient
γ	propagation coefficient

Table 2.11 Symbols for quantities and units of electricity and magnetism (based on ISO Recommendation R31)

Symbol	*Quantity*
I	electric current
Q	electric charge, quantity of electricity
ρ	volume density of charge, charge density (Q/V)
σ	surface density of charge (Q/A)
$E, (K)$	electric field strength
$V, (\varphi)$	electric potential
$U, (V)$	potential difference, tension
E	electromotive force
D	displacement (rationalised displacement)
D'	non-rationalised displacement
ψ	electric flux, flux of displacement (flux of rationalised displacement)
ψ'	flux of non-rationalised displacement
C	capacitance
ε	permittivity
ε_0	permittivity of vacuum
ε'	non-rationalised permittivity
ε'_0	non-rationalised permittivity of vacuum
ε_r	relative permittivity
χ_e	electric susceptibility
χ'_e	non-rationalised electric susceptibility
P	electric polarisation
$p, (P_e)$	electric dipole moment
$J, (S)$	current density
$A, (\alpha)$	linear current density
H	magnetic field strength
H'	non-rationalised magnetic field strength
U_m	magnetic potential difference
$F, (F_\mathrm{m})$	magnetomotive force
B	magnetic flux density, magnetic induction
Φ	magnetic flux
A	magnetic vector potential
L	self-inductance
$M, (L)$	mutual inductance
$k, (x)$	coupling coefficient
σ	leakage coefficient
μ	permeability
μ_o	permeability of vacuum
μ'	non-rationalised permeability
μ'_o	non-rationalised permeability of vacuum
μ_r	relative permeability
$k, (\chi_\mathrm{m})$	magnetic susceptibility
$k', (\chi'_\mathrm{m})$	non-rationalised magnetic susceptibility
m	electromagnetic moment (magnetic moment)
$H_\mathrm{v}, (M)$	magnetisation
$J, (B_i)$	magnetic polarisation
J'	non-rationalised magnetic polarisation
w	electromagnetic energy density
S	Poynting vector
c	velocity of propagation of electromagnetic waves *in vacuo*
R	resistance (to direct current)
G	conductance (to direct current)
ρ	resistivity
y, σ	conductivity
R, R_m	reluctance
$\mathrm{A}, (P)$	permeance
N	number of turns in winding
m	number of phases
p	number of pairs of poles
φ	phase displacement
Z	impedance (complex impedance)
$[Z]$	modulus of impedance (impedance)
X	reactance
R	resistance
Q	quality factor
Y	admittance (complex admittance)
$[Y]$	modulus of admittance (admittance)
B	susceptance
G	conductance
P	active power
$S, (P_\mathrm{s})$	apparent power
$Q, (P_\mathrm{q})$	reactive power

Table 2.12 Symbols for quantities and units of acoustics (based on ISO Recommendation R31)

Symbol	*Quantity*
T	period, periodic time
f, ν	frequency, frequency interval
ω	angular frequency, circular frequency
λ	wavelength
k	circular wavenumber
ρ	density (mass density)
P_s	static pressure
p	(instantaneous) sound pressure
$\varepsilon, (x)$	(instantaneous) sound particle displacement
u, v	(instantaneous) sound particle velocity
a	(instantaneous) sound particle acceleration
q, U	(instantaneous) volume velocity
c	velocity of sound
E	sound energy density
$P, (N, W)$	sound energy flux, sound power
I, J	sound intensity
$Z_s, (W)$	specific acoustic impedance
$Z_a, (Z)$	acoustic impedance
$Z_m, (w)$	mechanical impedance
$L_p, (L_N, L_W)$	sound power level
$L_p, (L)$	sound pressure level
δ	damping coefficient
Λ	logarithmic decrement
α	attenuation coefficient
β	phase coefficient
γ	propagation coefficient
δ	dissipation coefficient
r, τ	reflection coefficient
γ	transmission coefficient
$\alpha, (\alpha_\mathrm{a})$	acoustic absorption coefficient
R	sound reduction index sound transmission loss
A	equivalent absorption area of a surface or object
T	reverberation time
$L_\mathrm{N}, (\Lambda)$	loudness level
N	loudness

Table 2.13 Some technical abbreviations and symbols

Quantity	*Abbreviation*	*Symbol*
Alternating current	ac	
Ampere	A or amp	
Amplification factor		μ
Amplitude modulation	am	
Angular velocity		ω
Audio frequency	af	
Automatic frequency control	afc	
Automatic gain control	agc	
Bandwidth		Δf
Beat frequency oscillator	bfo	
British thermal unit	Btu	
Cathode-ray oscilloscope	cro	
Cathode-ray tube	crt	
Celsius	C	
Centi-	c	
Centimetre	cm	
Square centimetre	cm^2 or sq cm	
Cubic centimetre	cm^3 or cu cm or cc	
Centimetre-gram-second	cgs	
Continuous wave	cw	
Coulomb	C	
Deci-	d	
Decibel	dB	
Direct current	dc	
Direction finding	df	
Double sideband	dsb	
Efficiency		η
Equivalent isotropic radiated power	eirp	
Electromagnetic unit	emu	
Electromotive force instantaneous value	emf	E or V, e or v
Electron-volt	eV	
Electrostatic unit	esu	
Fahrenheit	F	
Farad	F	
Frequency	freq.	f
Frequency modulation	fm	
Gauss	G	
Giga-	G	
Gram	g	
Henry	H	
Hertz	Hz	
High frequency	hf	
Independent sideband	isb	
Inductance-capacitance		L-C
Intermediate frequency	if	
Kelvin	K	
Kilo-	k	
Knot	kn	
Length		l
Local oscillator	lo	
Logarithm, common		log or log$_{10}$
Logarithm, natural		ln or log$_e$
Low frequency	lf	
Low tension	lt	
Magnetomotive force	mmf	F or M
Mass		m
Medium frequency	mf	
Mega-	M	
Metre	m	
Metre-kilogram-second	mks	
Micro-	μ	
Micromicro-	p	
Micron		μ
Milli-	m	
Modulated continuous wave	mcw	
Nano-	n	
Neper	N	
Noise factor		N
Ohm		Ω
Peak to peak	p-p	
Phase modulation	pm	
Pico-	p	
Plan-position indication	PPI	
Potential difference	pd	V
Power factor	pf	
Pulse repetition frequency	prf	
Radian	rad	
Radio frequency	rf	
Radio telephony	R/T	
Root mean square	rms	
Short-wave	sw	
Single sideband	ssb	
Signal frequency	sf	
Standing wave ratio	swr	
Super-high frequency	shf	
Susceptance		B
Travelling-wave tube	twt	
Ultra-high frequency	uhf	
Very high frequency	vhf	
Very low frequency	vlf	
Volt	V	
Voltage standing wave ratio	vswr	
Watt	W	
Weber	Wb	
Wireless telegraphy	W/T	

Table 2.14 Greek alphabet and symbols

Name	*Symbol*		*Quantities used for*
alpha	A	α	angles, coefficients, area
beta	B	β	angles, coefficients
gamma	Γ	γ	specific gravity
delta	Δ	δ	density, increment, finite difference operator
epsilon	E	ε	Napierian logarithm, linear strain, permittivity, error, small quantity
zeta	Z	ζ	coordinates, coefficients, impedance (capital)
eta	H	η	magnetic field strength, efficiency
theta	Θ	θ	angular displacement, time
iota	I	ι	inertia
kappa	K	κ	bulk modulus, magnetic susceptibility
lambda	Λ	λ	permeance, conductivity, wavelength
mu	M	μ	bending moment, coefficient of friction, permeability
nu	N	ν	kinematic viscosity, frequency, reluctivity
xi	Ξ	ζ	output coefficient
omicron	O	*o*	
pi	Π	π	circumference ÷ diameter
rho	P	ρ	specific resistance

Table 2.14 *continued*

Name	*Symbol*		*Quantities used for*
sigma	Σ	σ	summation (capital), radar cross-section, standard deviation
tau	Τ	τ	time constant, pulse length
upsilon	Υ	*u*	
phi	Φ	φ	flux, phase
chi	Χ	χ	reactance (capital)
psi	Ψ	ψ	angles
omega	Ω	ω	angular velocity, ohms

References

1 COHEN, E.R. and TAYLOR, B.N., *Journal of Physical and Chemical Reference Data*, vol. 2, 663 (1973)

2 'Recommended values of physical constants'. CODATA (1973)

3 McGLASHAN, M. L., *Physicochemical quantities and units*, London: The Royal Institute of Chemistry (1971)

P Sproxton,
Alpha Image Ltd

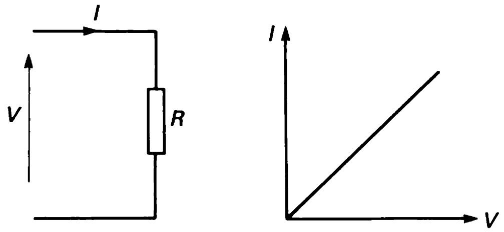
3

Analogue and Digital Circuit Theory

3.1 Analogue circuit theory

3.1.1 Resistors

Resistors are the simplest linear components encountered when applying circuit theory techniques. The symbol for a resistor is shown in *Figure 3.1* (a). If an emf V (measured in volts) is applied across the terminals of the resistor (resistance R, measured in ohms), it can be shown, as in (b), that the current flowing through the resistor is linearly proportional to the current I (measured in amperes).



(a) (b)

Figure 3.1 Resistor and voltage/current characteristic

This relationship is better known as Ohm's law, i.e.:

$$I = \frac{V}{R} \tag{3.1}$$

Another important relationship here is the power P (in watts) which is dissipated in the resistor. This is given by:

$$\begin{aligned} P &= VI \\ &= I^2R \end{aligned} \tag{3.2}$$

3.1.2 Series resistance circuits

A series configuration for resistors is illustrated in *Figure 3.2*. In this circuit, the current flowing through all three resistors is the same. Therefore according to equation (3.1) the potential difference across each resistance is given by:

$$V_1 = R_1,I \; ; V_2 = R_2I \; ; V_3 = R_3I$$

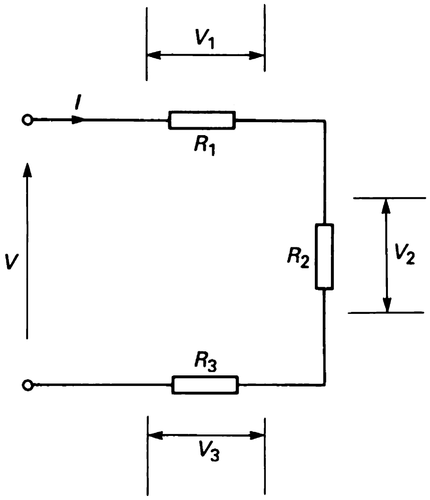


Figure 3.2 Series resistance circuit

The sum of these three potential differences is equal to the applied voltage V, i.e.:

$$V = V_1 + V_2 + V_3 \tag{3.3}$$

Equation (3.3) indicates that the algebraic sum of the potential differences around any complete circuit is equal to zero.

Using equation (3.1) we can re-arrange equation (3.3) to give:

$$\begin{aligned} V &= IR_1 + IR_2 + IR_3 \\ &= I(R_1 + R_2 + R_3) \end{aligned}$$

Therefore

$$I = \frac{V}{R_1 + R_2 + R_3} = \frac{V}{R_e} \tag{3.4}$$

The equivalent resistance of the circuit in *Figure 3.2* is therefore R_e, which is given by

$$R_e = R_1 + R_2 + R_3 \tag{3.5}$$

Equation (3.5) can be stated more generally: If any number of resistors are connected in series, then the equivalent resistance is the sum of the individual values.

3.1.3 Parallel resistance circuits

A second way of configuring resistors is shown in *Figure 3.3*. In this circuit, the voltage across each resistor is the same, but the current in each is given by:

$$I_1 = \frac{V}{R_1} \;;\; I_2 = \frac{V}{R_2} \;;\; I_3 = \frac{V}{R_3} \qquad (3.6)$$

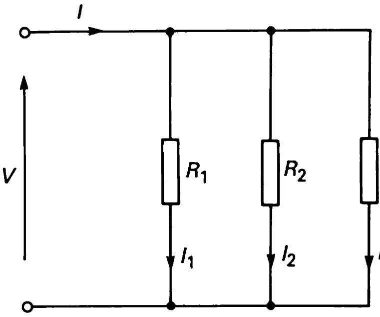


Figure 3.3 Parallel resistance circuit

Here, the total supplied current equals the sum of the currents through each resistor. i.e.:

$$I = I_1 + I_2 + I_3 \qquad (3.7)$$

Substituting for currents in equation (3.6):

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3} = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right) \qquad (3.8)$$

From Ohm's law (equation (3.1)) we have

$$I = \frac{V}{R_e} = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right) \qquad (3.9)$$

Therefore

$$\frac{1}{R_e} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \qquad (3.10)$$

This can be more generally stated by:

$$\frac{1}{R_e} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \;....\; + \frac{I}{R_n} \qquad (3.11)$$

i.e. the reciprocal of the equivalent resistance is equal to the sum of the reciprocals of the component resistances.

3.1.4 Voltage dividers

One very common implementation of resistor networks is the voltage divider (see *Figure 3.4*). For this circuit, the output is always smaller than the input, in a ratio which is determined by the value of R_1 and R_2. The current through R_1 and R_2 is given by:

$$I = \frac{V_i}{R_1 + R_2} \qquad (3.12)$$

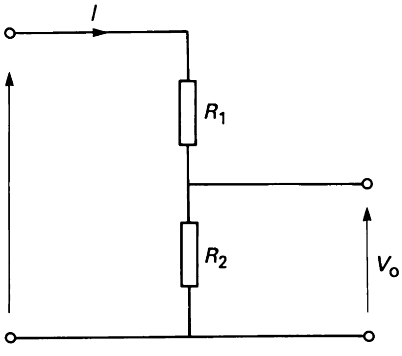


Figure 3.4 Voltage divider

and, as $V_o = IR_2$:

$$V_o = \frac{R_2}{R_1 + R_2} V_i \qquad (3.13)$$

3.1.5 Kirchoff's laws

The circuit in *Figure 3.5* (a) contains a parallel set of resistances with two *nodes* indicated. A node is a point where three or more conductors are joined. Kirchoff's first law states: The total current flowing towards a node is equal to the total current flowing away from that node, i.e. the algebraic sum of currents flowing towards a node is zero. Thus in *Figure 3.5* (b) at node B:

$$I_1 + I_2 + I_3 = I$$

or

$$I_1 + I_2 + I_3 - I = O$$

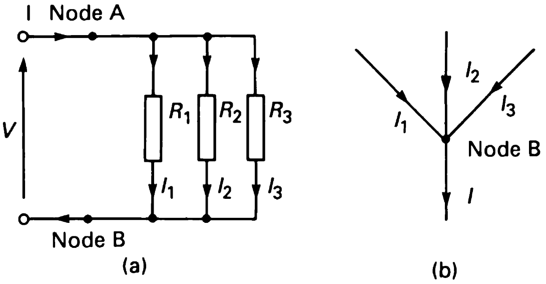


Figure 3.5 Kirchoff's current law

In general, for any node,

$$\Sigma I = O \qquad (3.14)$$

The circuit in *Figure 3.6* illustrates Kirchoff's second law, which states: The algebraic sum of the potential differences is zero, i.e.:

$$V = V_1 + V_2 + V_3$$

or

$$V_1 + V_2 + V_3 - V = O$$

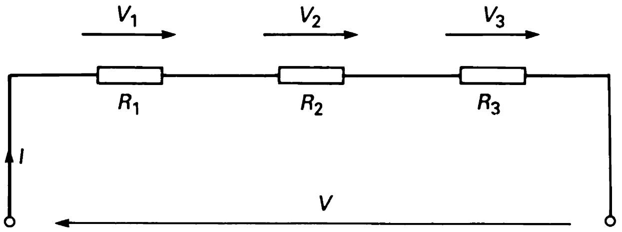


Figure 3.6 Kirchoff's voltage law

Generally, for any node,

$$\Sigma V = O \qquad (3.15)$$

3.1.6 Equivalent circuits

To simplify circuit analysis, it is often necessary to reduce portions of a circuit to a simpler equivalent form. This is in order to clarify areas of the circuit that are of particular interest. The following two sections describe two theorems that enable some networks to be simplified.

3.1.6.1 Thevenin's theorem

Consider the network and load resistance in *Figure 3.7* (a). Thevenin's theorem can be stated as follows: An active network, having two terminals, A and B, can be replaced by a constant voltage source having an emf V_o and an internal resistance r. The value of V_o is equal to the open circuit potential between A and B, and r is the resistance of the network measured between A and B, with the load disconnected and the sources of emf replaced by their internal resistances. On this basis, the network in *Figure* 7(a) can be redrawn as in (b).

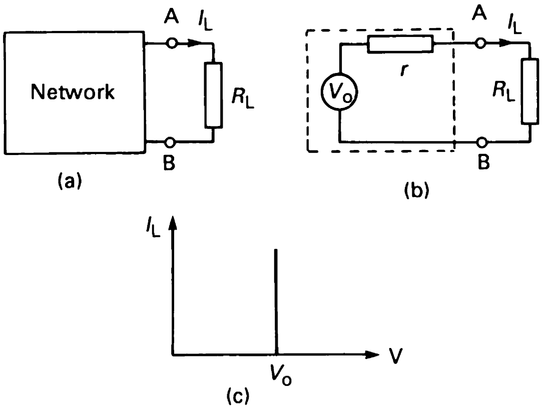


Figure 3.7 Thevenin's equivalent circuit

3.1.6.2 Norton's theorem

This is another theorem for an equivalent circuit. Norton's theorem states that: Any two-terminal network consisting of dc voltage sources and resistors can be replaced by a parallel combination of a current source I_e and a resistance R_e (see *Figure 3.8*). The current source is the short-circuit current at the output terminals and R_e is the same as for Thevenin's theorem. The Norton equivalent is illustrated in *Figure 3.8*, with the characteristic of the current generator in (c).

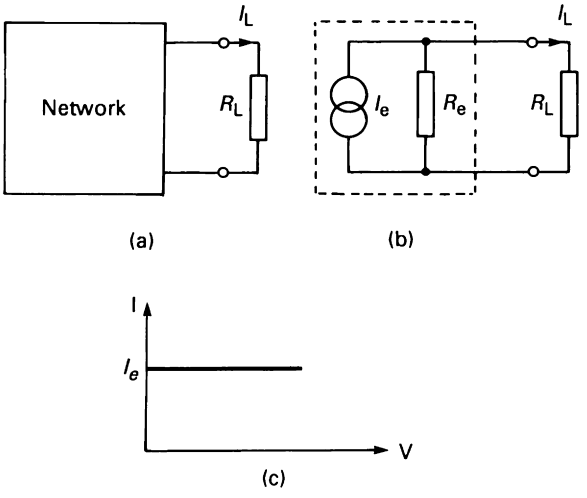


Figure 3.8 Norton's equivalent circuit

3.2 Alternating current circuits

In practice, the currents and voltages most often found in electronic circuits vary with time. The simplest time-varying waveform is one where the current or voltage changes direction periodically; this is termed an *alternating current* (ac).

The simplest ac waveform is the sine-wave, where current or voltage vary sinusoidally with time. A sinusoidal waveform (*Figure 3.9*) is generated by the variation of the vertical component of a vector rotating counter-clockwise with uniform angular velocity. A single revolution is termed a *cycle*, where the elapsed time interval for one revolution is the period *T*. The number of cycles per second is the *frequency*, *f*, of the sine-wave and the *period* is the reciprocal of the frequency.

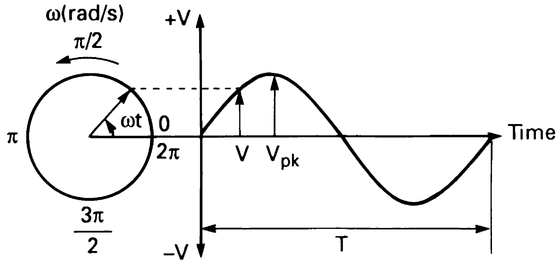


Figure 3.9 Rotating vector representation of a sinusoidal quantity

For a complete cycle, the vector will turn through 2π radians, therefore the angular frequency (in radians per second) is given by:

$$\omega = 2\pi f \qquad (3.16)$$

If the magnitude of the rotating vector is V_{pk}, the instantaneous value of v, at any time t, is given by:

$$v = V_{pk}\sin\omega t \qquad (3.17)$$

To produce a more generalized expression for the instantaneous voltage, the *phase* of the sine-wave must be considered. *Figure 3.10* illustrates two sinusoidal voltage waveforms with different phases.

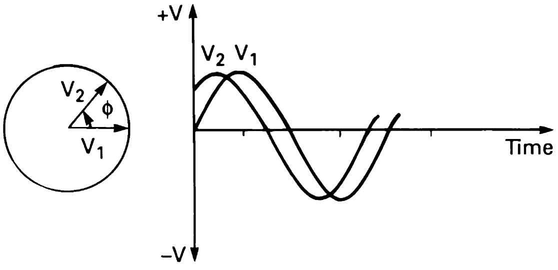


Figure 3.10 Vector representation of phase sinusoidal quantity

This shows that voltage V_2 is leading V_1, because V_2 has passed through zero in advance of V_1. In fact the voltage V_2 is *leading* voltage V_1 by the phase angle ϕ. The phase angle can only be specified between sine-waves of the same frequency. It is not possible to completely describe a sine-wave in terms of its amplitude and frequency unless it is being compared to a reference waveform of the same frequency. Hence, a more general expression for instantaneous voltage can be stated thus:

$$v = V_{pk}\sin(\omega t + \phi) \qquad (3.18)$$

In working with ac circuits, there are two circuit elements that become particularly significant. These are capacitors and inductors.

3.2.1 Capacitors

The circuit in *Figure 3.11*(a) contains a capacitor C consisting of two parallel plates separated by an insulator (e.g. air) with a dc voltage applied across the plates.

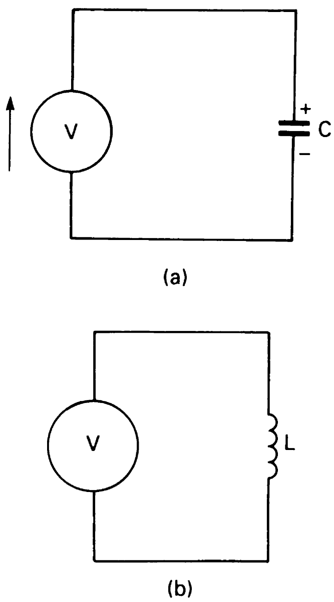


Figure 3.11 Capacitor and inductor circuits

A potential difference exists between the plates of the capacitor, and hence a positive charge will develop on the plate connected to the positive battery terminal and a negative charge on the plate connected to the negative terminal. The charge, *q*, on the plates is proportional to the voltage across them, i.e.:

$$q = CV \tag{3.19}$$

where C is a constant, called the capacitance, which depends upon the size, shape and separation of the plates, and the type of insulator between the plates.

In an ac circuit, the effect of the voltage changing with time will give rise to a time-varying charge on the capacitor plates. This is equivalent to a current through the circuit:

$$i = \frac{dq}{dt} \tag{3.20}$$

Substituting for *q* from equation (3.19),

$$i = C\frac{dv}{dt} \tag{3.21}$$

For a sinusoidal exciting voltage *v*, equations (3.19) and (3.20) give the current *i* in the circuit as:

$$\begin{aligned} i &= C\frac{d}{dt}(V_{pk}\sin\omega t) \\ &= \omega CV_{pk}\cos\omega t \qquad (3.22) \\ &= \omega CV_{pk}\sin(\omega t + \pi/2) \end{aligned}$$

Equation (3.22) shows the current to be sinusoidal and leading the voltage by a phase $\pi/2$.

3.2.2 Inductors

An inductor consists of a coil of wire around a magnetic circuit, a circuit which may, for example, consist of iron or air. The symbol for the inductor is shown connected to a voltage source V in *Figure 3.11*(b). The current in an inductor produces a magnetic flux around that inductor. The magnetic field will vary as the current changes, which will, in turn, induce an emf in the circuit. This emf is given by:

$$v = L\frac{di}{dt} \tag{3.23}$$

Equation (3.23) indicates that the voltage *v* across the inductance is proportional to the rate of change of current through it. The proportional constant, L, is determined by the size, shape and magnetic properties of the inductor. Substituting a sinusoidal exciting current for the current in equation (3.23).

$$\begin{aligned} V &= L\frac{d}{dt}(I_{pk}\sin\omega t) \\ &= \omega LI_{pk}\cos\omega t \\ &= \omega LI_{pk}\sin(\omega t + \pi/2) \qquad (3.24) \end{aligned}$$

Equation (3.24) shows that the current through the inductor *lags* the voltage by a phase angle of $\pi/2$.

3.2.2.1 Complex representation of sinusoidal quantities

We have seen in section *3.2* that sinusoidally varying voltages or currents can be represented by a rotating vector (*Figure 3.9*). By representing this vector on an Argand diagram, it can be described by a complex number.

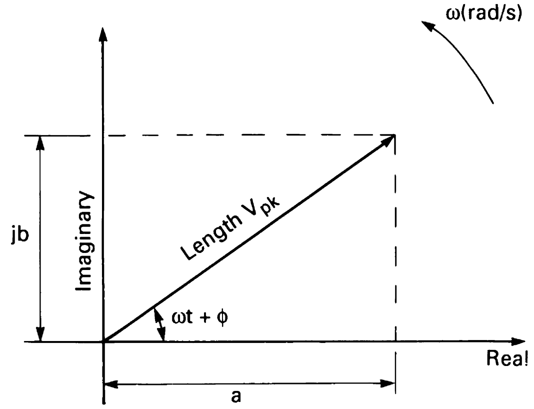


Figure 3.12 Complex representation of a sinusoidal quantity

For a complex number of the form $a + jb$, *Figure 3.12* shows that

$$a = V_{pk}\cos(\omega t + \phi)$$
$$b = V_{pk}\sin(\omega t + \phi)$$

Therefore

$$v = V_{pk}\cos(\omega t + \phi) + jV_{pk}\sin(\omega t + \phi) \tag{3.25}$$

This result can also be represented exponentially:

$$v = V_{pk}e^{j(wt + \phi)} \tag{3.26}$$

By this means, the inductor and capacitor can be represented in terms of their complex impedance. Consider first the inductor;

$$\begin{aligned} V_{pk}e^{j(wt+\phi)} &= L\frac{d}{dt}(I_{pk}e^{jwt}) \\ &= j\omega LI_{pk}e^{jwt} \\ \therefore V &= j\omega LI \qquad (3.27) \end{aligned}$$

By comparing equation (3.27) with Ohm's law (equation (3.1)), it can be reduced to

$$V = ZI \tag{3.28}$$

This is the ac form of Ohm's law, where *Z* is the complex impedance. Similarly, for the capacitor;

$$I_{pk}e^{j(w+\phi)} = C\frac{d}{dt}(V_{pk}e^{jwt})$$

$$= j\omega CV_{pk}e^{j\omega t}$$

$$\therefore I = j\omega CV \qquad (3.29)$$

Here the complex impedance for the capacitor is $Z = 1/j\omega C$. From equation (3.29) it can be seen that the impedance for the capacitor is $1/j\omega C$, and from equation (3.27) the impedance for the inductor is $j\omega L$. This indicates that they are both imaginary quantities. This means that the voltage and current are always 90° out of phase, i.e. these circuits are purely *reactive*. In practice circuits always have some resistance, which results in a real part to the impedance.

Consider the circuit in *Figure 3.13*.

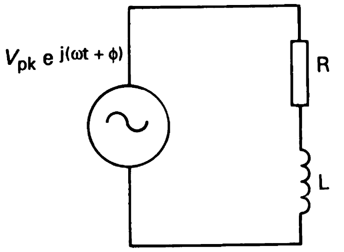


Figure 3.13 RL series circuit

The differential equation for this circuit is:

$$v = R_i + \frac{di}{dt} \qquad (3.30)$$

Allowing for a phase angle between the current and voltage, equation (3.30) can be expressed in its complex form, i.e.:

$$\begin{aligned} V_{pk}e^{j(wt+\varnothing)} &= RI_{pk}e^{jwt} + j\omega LI_{pk}e^{jwt} \\ &= (R+j\omega L)I_{pk}e^{jwt} \\ \therefore V &= (R + j\omega L)I \end{aligned} \qquad (3.31)$$

Equation (3.31) shows that the impedance for the RL series circuit is given by

$$Z = R + j\omega L \qquad (3.32)$$

which has a real part R and an imaginery part $j\omega L$. The impedance phase angle can be found for equation (3.32), as generally, for a complex number $a + jb$, the angle is arctan a/b, or for equation (3.32):

$$\varnothing = \arctan \frac{\omega L}{R}$$

or more generally:

$$\varnothing = \arctan \frac{X}{R} \qquad (3.33)$$

where X is the pure reactance, and R is the pure resistance.

To summarize, sinusoidally excited circuits, consisting of inductance, capacitance and resistance can be analysed using the component complex impedance:

pure resistive impedance	$Z_r = R$
pure capacitive impedance	$Z_c = 1/j\omega c$
pure inductive impedance	$Z_l = j\omega L$

For complex impedance, Ohm's law can be rewritten:

$$V = IZ$$

where Z is the equivalent impedance of the circuit. The complex impedance obeys the same rules as for parallel and series resistance circuits, i.e.:

series:	$Z = Z_1 + Z_2 + Z_3$
parallel:	$Z = 1(1/Z_1 + 1/Z_2 + 1/Z_3)$

3.2.2.2 RLC series circuits

Consider the circuit in *Figure 3.14*. The complex impedance method of determining the total equivalent impedance in the circuit gives:

$$Z = R + j\omega L + \frac{1}{j\omega C} \qquad (3.34)$$

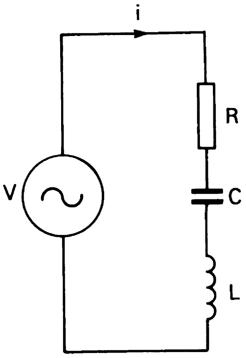


Figure 3.14 RLC series circuit

Separating into real and imaginary parts:

$$Z = R + j(\omega L - \frac{1}{\omega C}) \qquad (3.35)$$

Therefore the current in the circuit is:

$$I = \frac{V}{Z} = \frac{V}{\{R + j[\omega L - (1/\omega c)]\}} \qquad (3.36)$$

and the impedance angle is given by:

$$\varnothing = \arctan \frac{[\omega L - (1/\omega c)]}{R} \qquad (3.37)$$

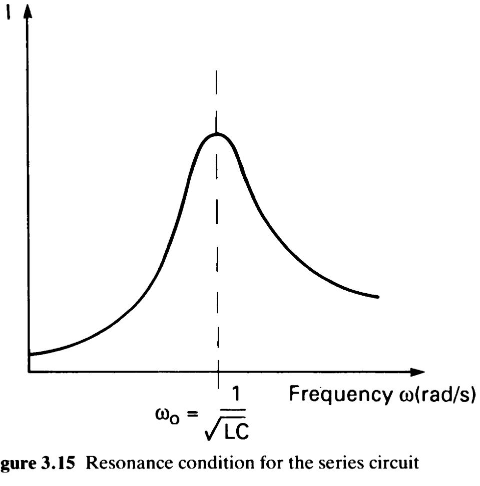


Figure 3.15 Resonance condition for the series circuit

Observation of equation (3.36) indicates that, as $\omega \rightarrow 0$, the current will be very small due to a high capacitive reactance. Also, as $\omega \rightarrow$ infinity, the inductive reactance becomes very high and hence the current very small. However, at a value of ω where the capacitive reactance is equal to the inductive reactance, the current is only dependent on the pure resistance

R, i.e.:

$$\text{for } \omega L = \frac{1}{\omega C}, \quad I = \frac{V}{R}$$

The value of ω for this condition is the *resonant frequency*, ω_o; i.e.:

$$\omega_o = \frac{1}{\sqrt{LC}} \tag{3.38}$$

In *Figure 3.15* the RLC series current is plotted against frequency, showing the current peaking at resonant frequency. Two points to notice about the series resonance condition are that the circuit appears as a pure resistance, and its current is in phase with the applied voltage. Also, in calculating the voltage drops across C and L, Kirchoff's laws would appear to be breached. This is explained, however, by indicating that the voltages are in antiphase, and hence cancel each other.

3.2.2.3 Parallel RLC circuits

In the circuit shown in *Figure 3.16* (a), the impedance for the parallel LC combination is given by:

$$Z = j\frac{\omega L}{1 - \omega^2 LC} \tag{3.39}$$

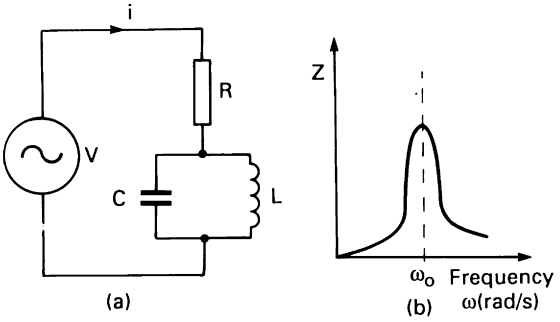


Figure 3.16 RLC parallel circuit

Equation (3.39) shows that the impedance is infinite when

$$\omega_o^2 LC = 1 \tag{3.40}$$

Here, ω_o is the resonant frequency. Furthermore, equation (3.40) can be rewritten:

$$\omega_o = \frac{1}{\sqrt{LC}} \tag{3.41}$$

Equation (3.41) is identical to equation (3.38) for the series LC combination. The difference between series and parallel resonance, however, is that the impedance is a *minimum* for series circuits at resonance, whereas it is a *maximum* for parallel circuits at resonance. The impedance characteristic for the parallel circuit is shown in *Figure 3.16* (b).

3.2.2.4 Q-factor

About the resonant frequency for a series circuit, there is a voltage magnification across the inductor (or capacitor), which is given by (voltage across inductor)/(supply voltage).

Thus

$$Q = \frac{\omega_o LI}{RI} = \frac{\omega_o L}{R} \tag{3.42}$$

Equation (3.42) indicates that the voltage magnification factor (the Q-factor) is equal to the ratio of the inductive reactance to the resistance. This means that near resonance, the resistance inherent in all inductors (due to multiple turns of wire) becomes significant in determining the sharpness or selectivity of the resonant circuit (*Figure 3.17*).

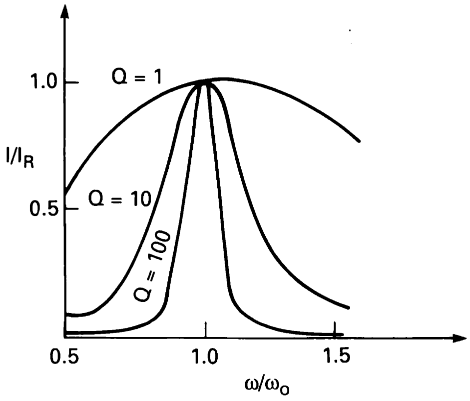


Figure 3.17 Q-factor

For circuits that require very high frequency selectivity, crystals are usually used instead of inductors and capacitors. This is because the Q-factor for LC circuits is typically in the range 10–100, whereas the Q for crystals can be as high as several thousand.

3.2.3 RC circuits as filters

RC circuits can be arranged to discriminate selected frequency bands by using low-pass and high-pass configurations.

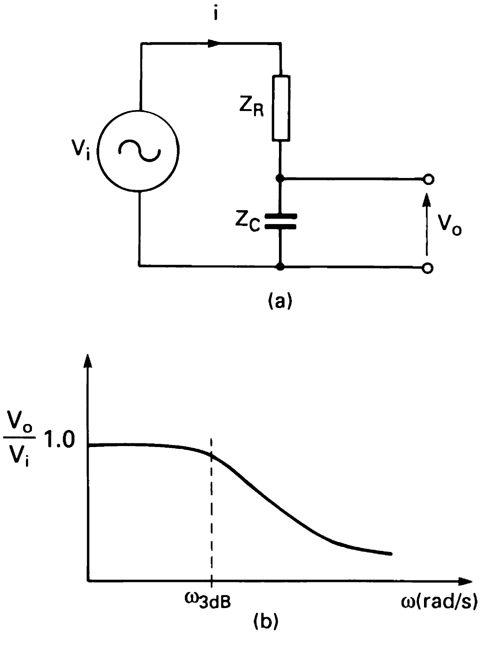


Figure 3.18 Low-pass RC filter circuit and characteristic

3.2.3.1 Low-pass RC filters

Consider the circuit in *Figure 3.18* (a). The output, V_o, of this circuit can be defined as a proportion of the input V_i:

$$V_o = \frac{Z_C}{Z_R + Z_C} V_i$$
$$= \frac{1/j\omega C}{R + 1/j\omega C} V_i$$
$$= \frac{1}{1 + j\omega CR} V_i \quad (3.43)$$

The amplitude response for this circuit is found by taking the modulus of equation (3.43), i.e.:

$$\frac{V_o}{V_i} = \frac{1}{\sqrt{1 + \omega^2C^2R^2}} \qquad (3.44)$$

This response is plotted in *Figure 3.18*(b). Equation (3.44) shows that as $\omega \rightarrow 0$, $V_i = V_o$, and that as $\omega \rightarrow \infty, V_o = 0$. Thus, the larger the frequency the lower the magnitude of the output V_o. For this reason, this configuration is called the *low-pass filter*. *Figure 3.18*(b) indicates the *half power* frequency, where:

$$V_o = \frac{1}{\sqrt{2}} V_i \qquad (3.45)$$

which, from equation (3.44), occurs when $\omega CR = 1$, or

$$\omega_{3dB} = \frac{1}{RC} \qquad (3.46)$$

3.2.3.2 High-pass RC filters

The configuration for the high-pass RC filter is shown in *Figure 3.19*(a). The current in this circuit is given by:

$$I = \frac{V_i}{Z} = \frac{V_i}{R + \frac{1}{j\omega C}} \qquad (3.47)$$

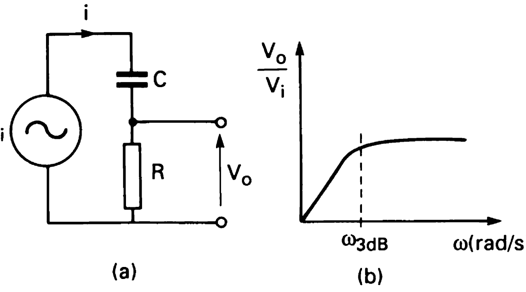


Figure 3.19 High-pass RC filter circuit and characteristic

Therefore, multiplying numerator and denominator by the complex conjugate of the denominator

$$I = V_i \left\{ \frac{j\omega C}{1 + j\omega CR} \cdot \frac{1 - j\omega CR}{1 - j\omega CR} \right\}$$
$$= V_i \frac{[R + (j/\omega C)]}{R^2 + (1/\omega^2C^2)} \qquad (3.48)$$

The voltage across R in *Figure 3.19*(a) is given by:

$$V_o = IR = V_i \frac{[R + (j/\omega C)]R}{R^2 + (1/\omega^2C^2)} \qquad (3.49)$$

Therefore, the magnitude response is given by

$$\frac{V_o}{V_i} = \frac{R}{\sqrt{R^2 + (1/\omega^2C^2)}} \qquad (3.50)$$

Equation (3.50) results in the response in *Figure 3.19*(b), which shows a zero amplitude response at dc ($\omega = 0$). For high frequencies, there is no attenuation at the output, V_o. This figure also indicates the half power frequency, which is again given by:

$$\omega_{3db} = \frac{I}{RC}$$

3.3 Digital circuit theory

Electronic circuits used for digital systems are designed to generate only two recognized output voltage levels, and probably the most common definition for these voltage levels is 5V for the high level and 0V for the lower level. In practice, a certain range is allowed for each of the two levels, again the most common being 0–0.8V for 0V, and 2.7–5V for 5V. Any voltage levels that exist outside these two ranges are invalid, and if present in a digital system will give rise to error conditions.

These voltage ranges are as defined for the ttl (transistor-transistor logic) family of digital circuits; other digital circuit families can use different voltage ranges, e.g. ecl (emitter coupled logic) uses levels -2.1 – -1.7V and -1.3 – -0.9V.

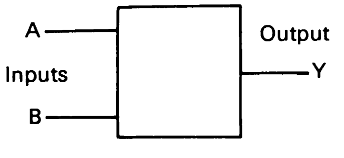
Ensuring digital circuits operate on two distinct voltage ranges, these two ranges can be equated to the two binary conditions 1 and 0 which are used in logic circuits. For ttl type circuits, binary 1 can be equated to the range 2.7V – 5V and the binary 0 condition can be equated to the range 0 – 0.8V. This is known as the *positive logic convention*, where the negative logic definition would equate 1 to the range 0 – 0.8V and 0 to the range 2.7 – 5V. Most digital systems use positive or mixed logic conventions.

3.3.1 Logic gates

The construction of the logic gate from transistor or fet type devices is beyond the scope of this section, as the aim is to describe the use of logic gates, not their construction. Any logic gate manufacturer provides this information in his data sheets.

All digital systems consist, at the most fundamental level, of individual logic gates. Although there are a number of different gate characteristics, there are three types from which all other logic functions can be synthesized: OR, AND and NOT gates.

Logic gates are circuits with one output and one or more inputs. The gate in *Figure 3.20* illustrates a two input logic gate. The truth table in (b) lists the gate output for all possible combinations of binary inputs.



(a)

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(b)

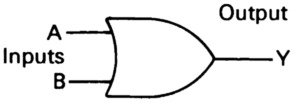
Figure 3.20 Logic gate and truth table

From this truth table we can see that the output of the gate will be binary 0 unless both input A and input B are binary 1. Manufacturers will normally quote their truth tables in terms of positive logic.

The truth table is used in the following descriptions of the basic logic gates.

3.3.1.1 OR gates

The symbol for the OR gate along with its truth table is shown in *Figure 3.21*. Considering the truth table for the two input OR gate, it can be seen that the output Y will be a binary 1 if input A is 1, *or* input B is 1, *or* both are 1, hence its name.



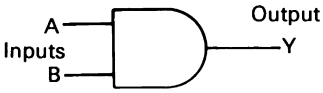
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

(a) (b)

Figure 3.21 OR gate and truth table

3.3.1.2 AND gates

The symbol for the AND gate along with its truth table is shown in *Figure 3.22*. Observing the truth table for this device, we see that the output Y will only be binary 1 when input A is 1 *and* input B is 1.



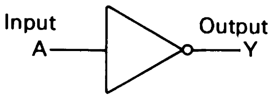
A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(a) (b)

Figure 3.22 AND gate and truth table

3.3.1.3 The inverter

The symbol for the NOT gate along with its truth table is shown in *Figure 3.23*. Observation of the truth table indicates the output Y to be the logical inverse of whichever state is present on the input A.



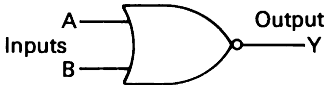
A	Y
0	1
1	0

(a) (b)

Figure 3.23 INVERTER and truth table

3.3.1.4 NOR gates

The symbol for the NOR gate along with its truth table is shown in *Figure 3.24*. The NOR gate is simply an OR gate followed by an inverter; this can be seen from the truth table in *Figure 3.24*(b) where the output Y is the inverse of the output Y for the OR gate in *Figure 3.21*(b).



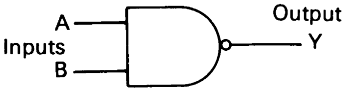
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

(a) (b)

Figure 3.24 NOR gate and truth table

3.3.1.5 NAND gates

The symbol for the NAND gate, and its truth table, is shown in *Figure 3.25*. As with the NOR function, the NAND gate is an AND gate followed by an inverter as can be seen from the two truth tables which show the output Y of the NAND gate to be the inverse of the output of the AND gate.



A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

(a) (b)

Figure 3.25 NAND gate and truth table

3.3.2 Implementing AND / OR functions from NAND / NOR gates

Both the NAND and the NOR gates can be used to implement inverters (see *Figure 3.26*). In *Figure 3.27* a NAND gate has been followed by a second NAND configured as an inverter. The truth table demonstrates that the function implemented is that of an AND gate. This means that NAND elements can be used to implement AND functions, and that NOR functions can be used to implement OR functions, hence the NAND and NOR gates are functionally complete. Thise becomes particularly important when considering that most manufacturers provide multiple gates within a single integrated circuit.

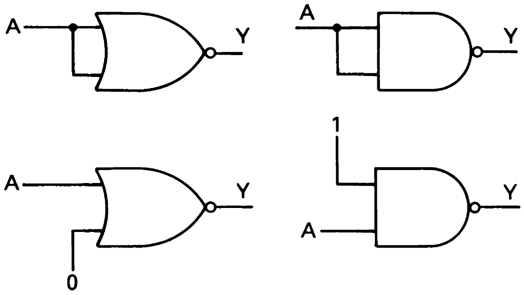
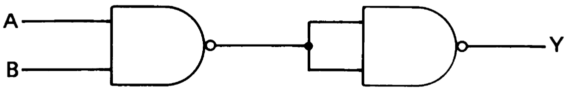


Figure 3.26 Implementing inverters with NAND and NOR gates

3.4 Boolean algebra

3.4.1 Combinational and sequential circuits

All logic circuits can be subdivided into two types: combinational logic and sequential logic. A *combinational* circuit can be described by stating that its output will be true for only certain combinations of input variables; all other input combinations will cause the output to be false. The output(s) for a *sequential* circuit depend upon current input variables, time and past input variables. Sequential circuits use combinational circuits as building blocks, so it is essential to understand these elements. Examples of the implementation of sequential logic circuits are covered in some later chapters.



A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

Figure 3.27 Synthesizing an AND gate with NAND gates

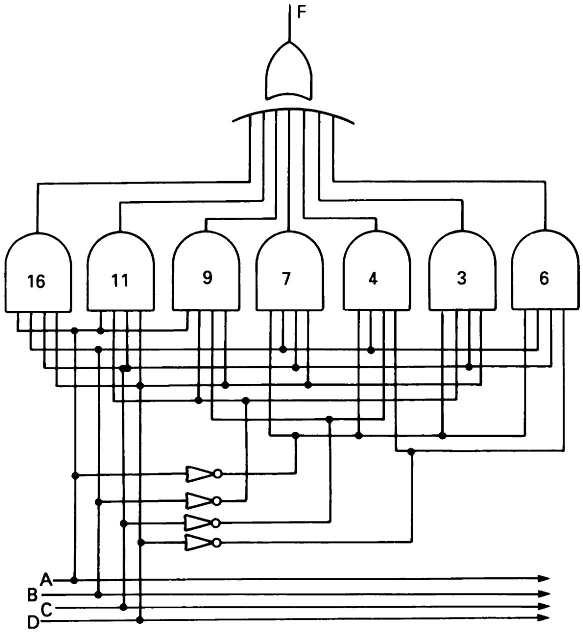


Figure 3.28 Logic system example

For effective design of digital systems, the designer must be able to specify clearly the function of the system, ensure the design will be reliable, and ensure the minimum number of logic elements are used. This section covers a method of defining a logic system in terms of an algebraic equation, and this method will be extended in the discussion of Karnaugh maps for minimizing logic circuit resources.

Consider the logic circuit in *Figure 3.28*. In this example, we have a four bit parallel data transmission link, where there is a requirement to recognize the presence of particular binary codes. The detection circuit is designed such that its output will be logic 1 for the presence of any of the decimal codes 16,11,9,7,4,3 or 6 and a logic 0 for any other combination of codes on the link.

Each of the logic system input variables are designated an alphabetic character (in this case A,B,C,D). Each of these variables implies its complement, i.e. $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$. For a positive logic convention, $A = 1$ (true) and $\bar{A} = 0$ (false). Thus, for the binary code 0100 on the transmission link in *Figure 3.28*, it is required to recognize $\bar{A}B\bar{C}\bar{D}$.

To recognize the code 0100, in this example, we can use a four input AND gate to detect the presence of $\bar{A}B\bar{C}\bar{D}$. The other required codes are similarly detected.

Instead of using unwieldy grammar to describe this example, Boolean algebra can be used. The basic rules for Boolean algebra are covered in the following sections.

3.4.2 Boolean OR/AND identities

The truth table for the logical OR relation is shown in *Figure 3.21*(b). It can be represented in terms of a Boolean expression:

$$A + B = X \qquad (3.51)$$

where the + symbol indicates the Boolean OR operation, A and B are the input variables, and X is the output. There are a number of important Boolean identities associated with the OR function, which can all be verified using the OR truth table:

$$A + 0 = A \qquad (3.52)$$

$$A + 1 = 1 \qquad (3.53)$$

$$A + A = A \qquad (3.54)$$

$$A + B + C = (A+B)+C = A + (B+C) \qquad (3.55)$$

$$A + B = B + A \qquad (3.56)$$

All these relations can be directly realized using OR, as illustrated in *Figure 3.29*.

The truth table for the AND gate is shown in *Figure 3.22*(b). It can be written in the Boolean expression

$$A.B = X \qquad (3.57)$$

As with the OR gate, there are a number of important AND identities, these are:

$$AO = O \qquad (3.58)$$

$$A1 = A \qquad (3.59)$$

$$AA = A \qquad (3.60)$$

$$ABC = (AB)C = A(BC) \qquad (3.61)$$

$$AB = BA \qquad (3.62)$$

All these relations can be directly realized using AND, as illustrated in *Figure 3.30*.

There are a number of other very important Boolean identities:

$$\bar{\bar{A}} = A \qquad (3.63)$$

$$A + \bar{A} = 1 \text{ (OR complement)} \qquad (3.64)$$

$$A\bar{A} = O \text{ (AND complement)} \qquad (3.65)$$

$$A(B+C) = AB+AC \qquad (3.66)$$

3.4.3 De Morgan's theorem

De Morgan's theorem indicates a useful relationship between AND and OR functions. It can be stated in the form of two laws, i.e.:

$$\bar{A} + \bar{B} + \bar{C} + \ldots + \bar{N} = \overline{ABC \ldots N} \qquad (3.67)$$

$$\bar{A}\,\bar{B}\,\bar{C} \ldots \bar{N} = \overline{A + B + C + \ldots + N} \qquad (3.68)$$

Figure 3.31 illustrates the physical realization of De Morgan's laws. In terms of gates, it can be seen that a NAND gate is equivalent to an OR gate with inverted inputs, and that a NOR gate is equivalent to an AND gate with inverted inputs.

Application of these laws can enable a designer to implement OR/NOR functions when there are only AND/NAND functions available, and vice versa.

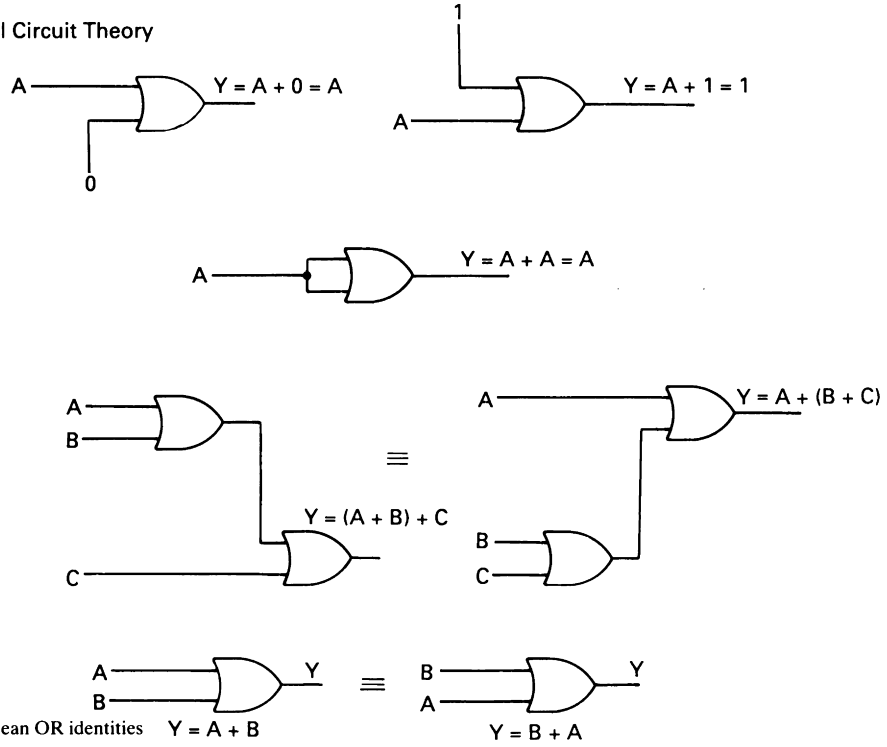


Figure 3.29 Realization of Boolean OR identities

3.5 Karnaugh maps

There are a number of different methods of minimizing Boolean expressions, but for functions of up to six variables, the Karnaugh map provides a method most suitable for manipulating expressions by hand. Functions above six variables are best processed using computer algorithms. Such algorithms are often supplied by the manufacturers of programmable logic devices.

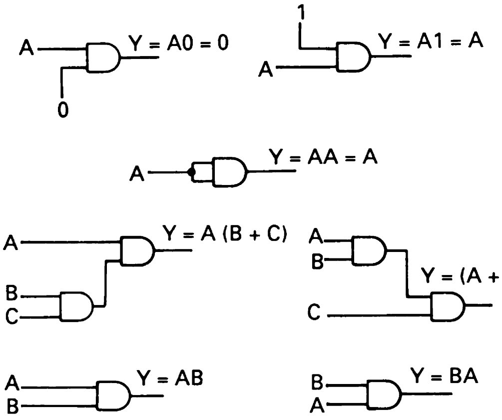


Figure 3.30 Realization of Boolean AND identities

3.5.1 Preparing a Boolean expression for plotting on a Karnaugh map

Boolean expressions will often exist in two distinct forms: the standard sum of products (*SOP*) form and the product of sums (*POS*). Expression (3.69) shows an example of the SOP form, whereas expression (3.70) gives an example of the POS form.

Before a Boolean expression can be plotted on the Karnaugh map, it must be converted into the standard sum of products form (*SSOP*).

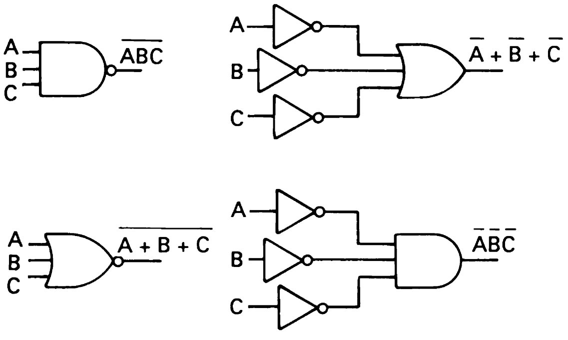


Figure 3.31 Realization of De Morgan's laws

$$AB + BC + \bar{B}D \qquad (3.69)$$

$$(A+B+C)(\bar{B} + \bar{C}) \qquad (3.70)$$

Using the distributive law (equation (3.66)), the POS expression (3.70) can be converted into the SOP form:

$$A\bar{B} + B\bar{B} + C\bar{B} + A\bar{C} + B\bar{C} + C\bar{C}$$

$$= A\bar{B} + C\bar{B} + A\bar{C} + B\bar{C} \qquad (3.71)$$

This form, now in SOP form, needs one further conversion into SSOP form. The original expression contains three variables A, B and C; for SSOP form, each product term must include all of these variables or their complements. This can be done by taking each product term in equation (3.71) with a missing variable, and ANDing that term with the sum of the missing variable and its complement, i.e.:

$$A\bar{B}(C+\bar{C}) + (A+\bar{A})\bar{B}C + A(B+\bar{B})\bar{C} + (A+\bar{A})B\bar{C}$$

$$= A\bar{B}C + A\bar{B}\bar{C} + A\bar{B}C + \bar{A}\bar{B}C + AB\bar{C} + A\bar{B}\bar{C} + AB\bar{C} + \bar{A}B\bar{C} \qquad (3.72)$$

This operation does not affect the expression, as the sum of a variable and its own complement is 1 (see equation (3.64)). Duplicate terms can now be removed from equation (3.72), giving:

$$A\overline{B}C + A\overline{B}\,\overline{C} + \overline{A}\,\overline{B}C + AB\overline{C} + \overline{A}B\overline{C} \qquad (3.73)$$

The SSOP terms in an expression are called *minterms*. Minterms are numbered according to the decimal code they represent, e.g.:

$$F = \overline{A}\,\overline{B}C + \overline{A}BC + A\overline{B}\,\overline{C} \qquad (3.74)$$

contains the minterms m1, m3 and m4. This shows that F = 1 when minterms m1, m2 or m4 are present and that F = 0 for all other possible minterms. (For equation (3.74), F will be 1 when, for example, the third term, minterm 4 is 1 i.e. A=1, B=0 and C=0.)

3.5.2 Entering an expression on the Karnaugh map

A Karnaugh map is plotted by entering each term of the SSOP expression in one of the map locations. The map for a two variable function is shown in *Figure 3.32*. It contains four locations, as there are four possible combinations of the two input variables, A and B. The minterm numbers are also shown in *Figure 3.32*, although they are not normally drawn in.

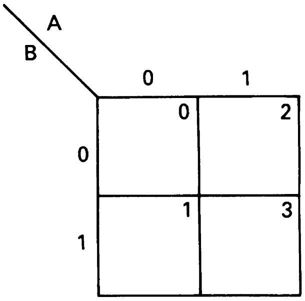


Figure 3.32 Two variable Karnaugh map

Consider the Boolean function

$$F = AB + \overline{A}B \qquad (3.75)$$

which indicates that F will be 1 when A is 1 and B is 1 and when A is 1 and B is 0. F will be 0 for the other two combinations of variables A and B. This can be plotted on the Karnaugh map as in *Figure 3.33*. Note the minterms present and their number and position on the Karnaugh map.

3.5.3 Reducing an expression using a Karnaugh map

Once an expression has been plotted, it can be reduced by forming the entries into logically adjacent groups. Consider *Figure 3.33*, where the function $F=AB+ A\overline{B}$ has been plotted. In this example, the bottom left and right entries are logically adjacent, forming a *couple*. This couple indicates that the function will be a 1 regardless of the state of A, because A is present in its complemented and uncomplemented form. Therefore

$$F_R = B \qquad (3.76)$$

Diagonal coupling is not valid, as it would mean more than one variable changing state at one time, e.g., minterm 2 is not

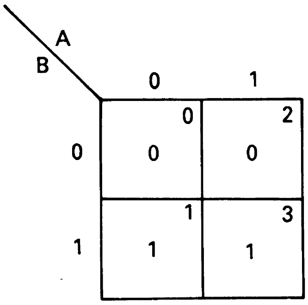


Figure 3.33 Karnaugh map of the function $F = AB + A\overline{B}$

logically adjacent to minterm 1. The result in equation (3.76) can be verified using the Boolean identities, i.e.:

$$F = BA + B\overline{A} = B(A+\overline{A}) = B \qquad (3.77)$$

where $(A+\overline{A})$ is the OR complement.

A three variable Karnaugh map is shown in *Figure 3.34*. It contains eight locations to represent each of the three variable combinations. The locations on the map are arranged so that all entries are logically adjacent. This means that only one variable will change state between any two adjacent map locations. Logical adjacency also exists between the extreme opposite squares for a row or column; for the example in *Figure 3.34*, location $A\overline{BC}$ (minterm 4) is adjacent to the location $\overline{ABC}$ (minterm 0). Note that due to the logical adjacency, the minterm numbers do not follow in order.

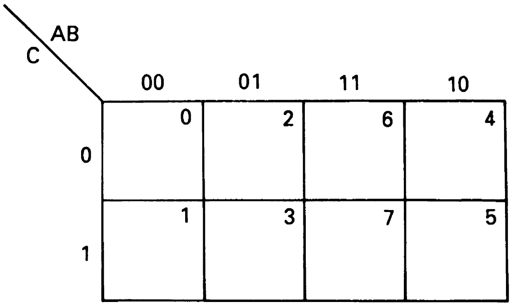


Figure 3.34 Three variable Karnaugh map

An example of a three variable expression reduction is shown in *Figure 3.35*, which maps the following expression:

$$F = \overline{A}\,\overline{B}C + A\overline{B}\,\overline{C} + A\overline{B}C + AB\overline{C} + ABC \qquad (3.78)$$

Grouping logically adjacent locations to form the largest possible groups, the map reveals a *quad* and a *couple* of adjacent terms. The quad indicates the constituent terms to be independent of the variables B and C because the complemented and non-complemented form exists for this quad. The couple indicates its constituent terms to be independent of A, because its complemented and non-complemented forms appear. Thus the function F will depend only upon A for the quad, and $\overline{B}C$ for the couple. Thus the expression in equation (3.78) can be reduced to:

$$F_R = A + \overline{B}C \qquad (3.79)$$

Using Boolean algebraic reduction, the result for the quad can be verified, i.e.:

$$\begin{aligned} &A\overline{B}\,\overline{C} + A\overline{B}C + AB\overline{C} + ABC \\ &\quad = A(\overline{B}\,\overline{C}+\overline{B}C+B\overline{C}+BC) \\ &\quad = A[\overline{B}(\overline{C}+C) + B(\overline{C}+C)] \\ &\quad = A(\overline{B}+B) \\ &\quad = A \end{aligned} \qquad (3.80)$$

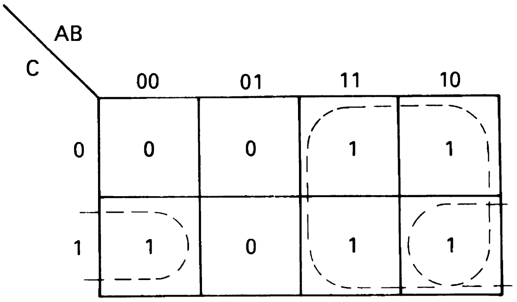


Figure 3.35 Karnaugh map of the function $F = \overline{A}\,\overline{B}C + A\overline{B}\,\overline{C} + A\overline{B}C + AB\overline{C} + ABC$

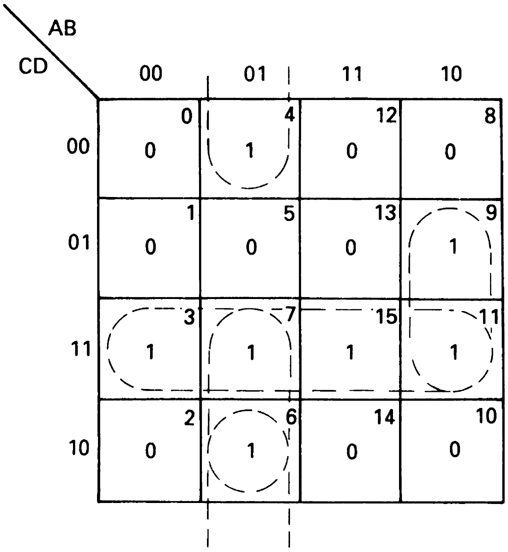


Figure 3.36 Karnaugh map of the function $F = ABCD + A\overline{B}CD + A\overline{B}\,\overline{C}D + \overline{A}BCD + \overline{A}B\overline{C}\,\overline{D} + \overline{A}\,\overline{B}CD + \overline{A}BC\overline{D}$

The example in *Figure 3.28* describes a logic system whose function can be described by the Boolean expression:

$$F = ABCD + A\overline{B}CD + A\overline{B}\,\overline{C}D + \overline{A}BCD + \overline{A}B\overline{C}\,\overline{D} + \overline{A}\,\overline{B}CD + \overline{A}BC\overline{D} \qquad (3.81)$$

This can be plotted on the four variable Karnaugh map as in *Figure 3.36*. In this example there are 16 possible input combinations for the four variables, so the map contains 16 locations. From this map it can be seen that the largest group of terms that can be formed is the quad; the other entries on the map can be grouped into two quads, as shown. The quad indicates that both A and B exist in their complemented and non-complemented forms; therefore it reduces to CD. Reducing the two couples, the resultant reduced expression becomes:

$$F_R = CD + \overline{A}B\overline{C} + A\overline{B}D \qquad (3.82)$$

The power of this technique can be illustrated by considering the implementation of the unreduced expression in equation (3.81) and comparing it with the reduced expression in equation (3.82). The unreduced implementation is shown in *Figure 3.28* and the reduced implementation in *Figure 3.37*; the reduction in gates can be clearly seen.

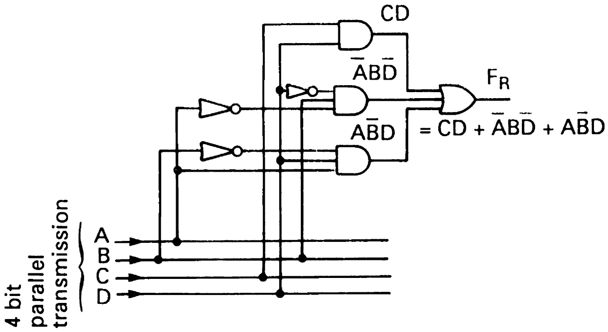


Figure 3.37 Implementation of function $F_R = CD + \overline{A}B\overline{C} + A\overline{3}D$

The operation of using a Karnaugh map for reducing an expression can be summarized as follows:

- Form standard sum of products of the unreduced expression.
- Plot the minterms on the map.
- Form the largest and least number of groups of logically adjacent entries (these groups will always contain a number which is a power of 2, e.g. 2,4,8).

Generally, a couple allows two original terms to be reduced to one smaller term, a quad allows four original terms to be reduced to one smaller term, etc. A map entry surrounded by 0s cannot be reduced, and will appear in the final expression without any reduction.

3.5.4 Prime implicants

Given the reduction method described in section *3.53*, it is still possible to produce a non-minimal result. This can be illustrated by considering the *prime implicants*.

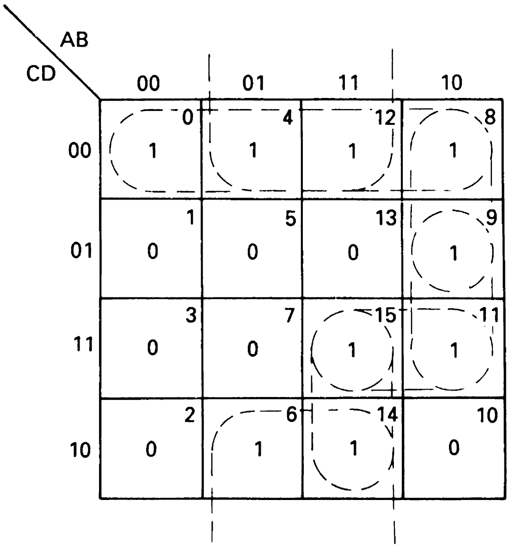


Figure 3.38 Function with non-essential prime implicants

Once the Karnaugh map has been plotted, logically adjacent entries are grouped together; these groups are the prime implicants. The four variable example in *Figure 3.38* shows the following prime implicants are formed:

$\overline{C}\,\overline{D}$	$A\overline{B}D$
$B\overline{D}$	ACD
$A\overline{B}\,\overline{C}$	ABC