ELEKTOR ELECTRONICS

THE INTERNATIONAL ELECTRONICS MAGAZINE
OCTOBER 1991
UK £1·90

RS

Digital function generator

Four-terminal networks

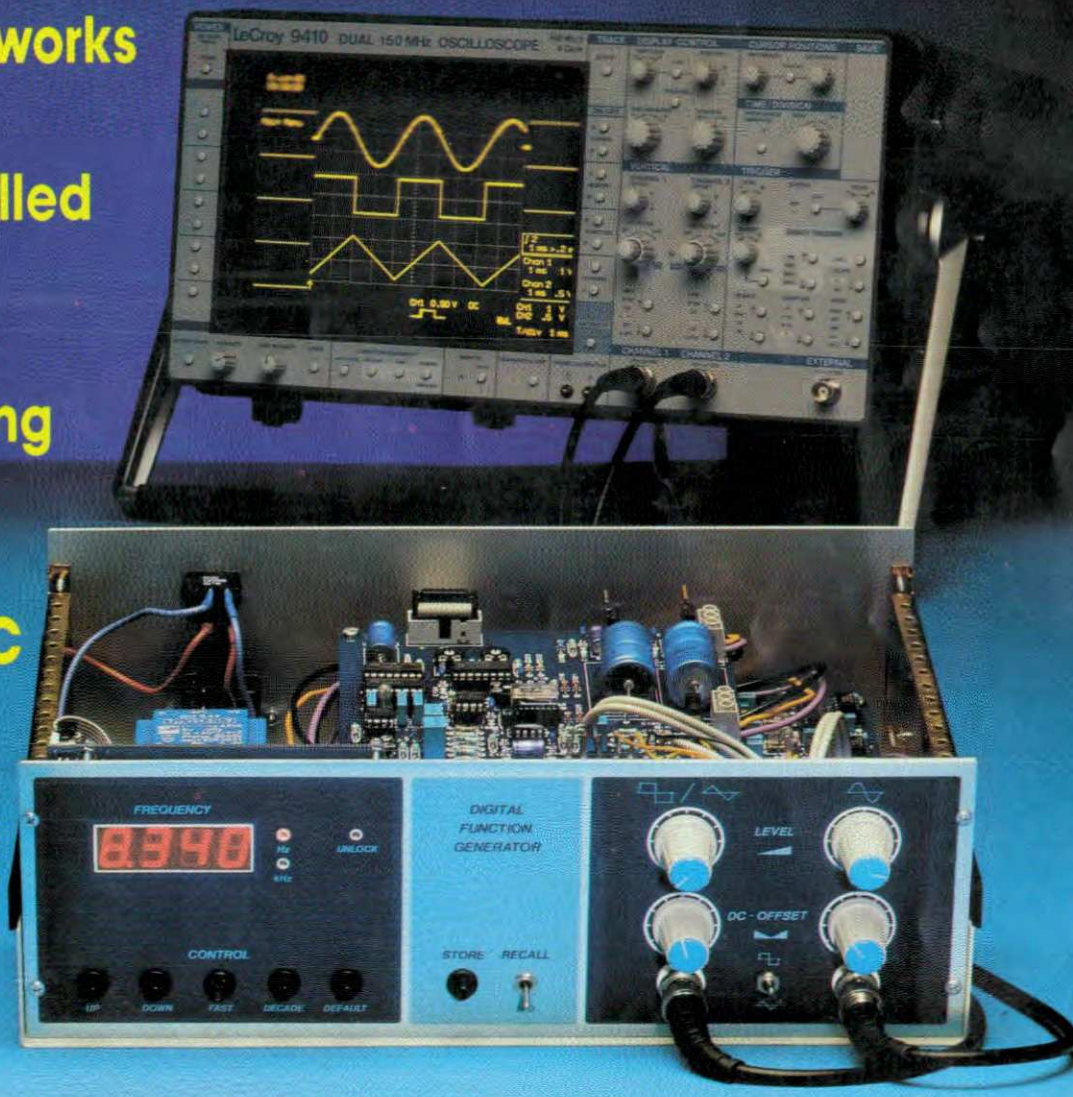
Computer-controlled weather station

A review of coding theory

50 MHz 8-bit DAC

Audio spectrum shift techniques

AM Receiver



10
9 770268 451036

ELEKTOR ELECTRONICS

CONTENTS

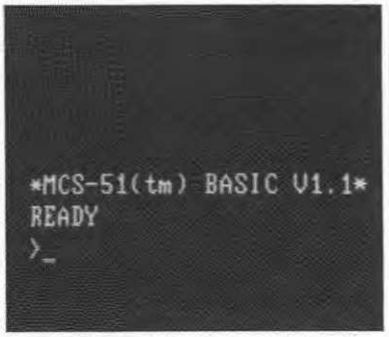
October 1991
Volume 17
Number 193

Front cover
Function generators are intended to produce several different waveforms and are, therefore, very flexible instruments. Most of them generate a sine, rectangular and triangular waveform over a frequency range of up to 1–50 MHz. Some also have the capability of producing pulse waveforms. Generally, the performance of a function generator is inferior to that of dedicated instruments, but the digital design of which the first of three parts is presented in this issue has the capability of accurate frequency setting coupled with very low distortion of the sine wave output.

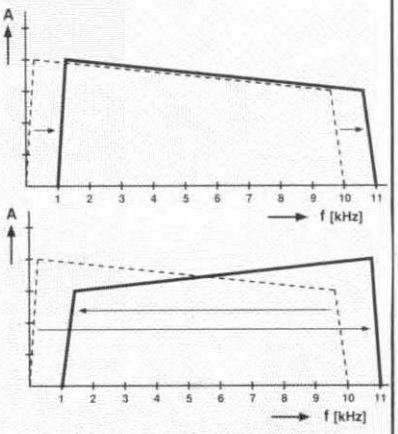
In next month's issue

- Class A amplifier
- Dissipation limiter
- Function generator [2]
- Experimental quadri-form ferrite antenna
- 24-bit full-colour video digitizer
- Four-terminal networks – Part 2
- Computer-aided electronics design

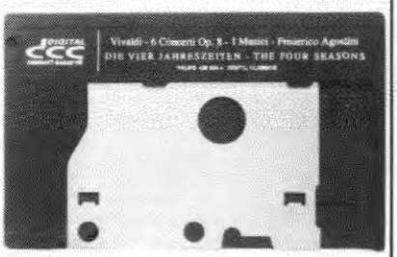
ABC
MEMBER OF THE AUDIT BUREAU OF CIRCULATIONS



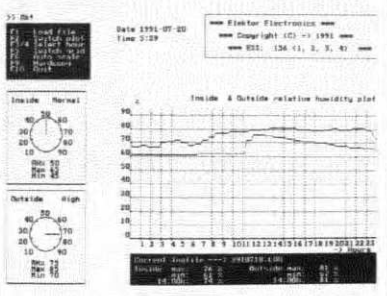
Upgrade for MCS-BASIC-52
p.14



Audio spectrum shift techniques – p. 16

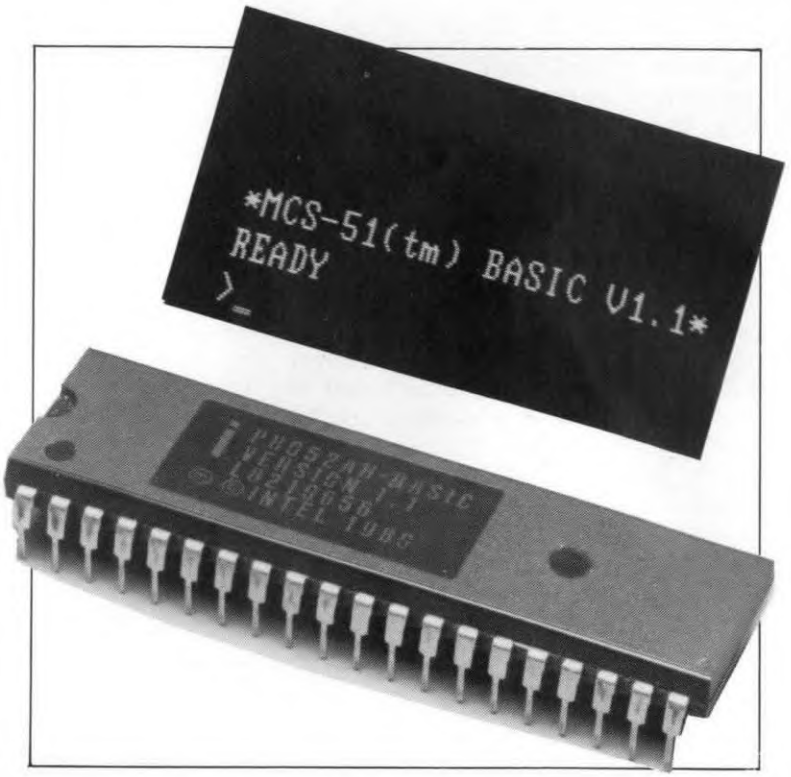


The digital compact cassette
p. 56



Electronic hygrometer – p. 28

UPGRADE FOR MCS® BASIC-52 V1.1 (Part 1)



The 8052AH-BASIC from Intel is a versatile microcontroller with a powerful BASIC interpreter lurking in its on-board mask-programmed ROM. The authors, having worked with this IC for some time, discovered certain flaws in the BASIC interpreter, and set out to produce a better, faster version that can be run from EPROM.

by Dusan Mudric and Zoran Stojsavljevic

TO be able to make changes to the MCS-52 BASIC interpreter in the 8052AH-BASIC microcontroller, it is necessary to first unload it from the IC. This is done basically as described in an earlier article on the MCS BASIC-52 interpreter, Ref. 1. The result of reading the 8-KByte ROM is a file in Intel-Hex format that contains the machine code of the MCS BASIC-52 interpreter (version 1.1).

MCS BASIC-52 (Ref. 3), extracted from the 8052AH-BASIC V1.1 microcontroller, was disassembled and texts, tables and constants were extracted in order to produce an assembler version of the interpreter. The size of this assembler file was approximately 4,000 lines. Studying the program, we found that the operation of the interpreter could be improved by rewriting certain lines of assembler code. Subsequently, a number of algorithms were developed and substituted for the ones originally implemented by Intel. Furthermore, errors found in a number of routines were corrected.

```
ADDR  CODE      INSTRUCTION

19F2H 752A00    MOV   2AH,#00H  ; +000T
19F5H 71C8      ACALL 1BC8H
19F7H 7F04      MOV   R7,#04H   ; +004T
19F9H 792E      MOV   R1,#2EH   ; +046T
19FBH 749E      MOV   A,#9EH    ; +158T
19FDH C3        CLR   C
19FEH 9C        SUBB  A,R4
19FFH D4        DA    A
1A00H CC        XCH   A,R4
1A01H 7001      JNZ   1A04H     ; $ + 03H
1A03H FC        MOV   R4,A
1A04H B45000    CJNE  A,#50H,1A07H ; +080T ; $ + 03H
1A07H 302318    JNB   23H,1A22H ; $ + 1BH
1A0AH B3        CPL   C
1A0BH 5119      ACALL 1A19H
1A0DH 5008      JNC   1A17H     ; $ + 0AH
1A0FH 052A      INC   2AH

                                        910128 - 12
```

Fig. 1. Original floating-point nucleus in the MCS BASIC-52 interpreter.

Floating point nucleus

One of the routines in the BASIC interpreter found to contain programming errors is the floating-point arithmetic nucleus. The errors can be demonstrated by running two small programs:

```
10   a=.10000001E30
20   b=.99999993E29
30   ?a-b
```

The result, 2.74E22, is erroneous, and should be 1.7E22. Similarly,

```
10   a=.10000001E30
20   b=.99999997E29
30   ?a-b
```

produces 1.34E22 instead of 1.3E22.

The disassembly listing of the original floating point nucleus developed by Intel is given in Fig. 1, and the version developed by the authors in Fig. 2. When implemented in the BASIC interpreter, the nucleus shown in Fig. 2 produces the correct answers to the above subtractions.

Other corrections

Further improvements were made to the hex-to-BCD conversion routine, both in regard of efficient programming and speed. For example, two approaches are possible for extracting BCD digits a, b, c, d, and e in

xyzwH = aD*10000D + bD*1000D + cD*100D + dD*10D + eD*1D

These possibilities are:

1. successive extraction of BCD digits starting with the most significant digit, a;
2. successive extraction of BCD digits starting with the least significant digit, e.

If the original version of the hex-to-BCD converter is studied, it is seen that the first procedure is employed. The DPTR is used as a 'weighted register', and the procedure is based on finding a suitably weighted subtraction number from a variable value.

```
ADDR CODE                       INSTRUCTION

19F2 752A00     5699            MOV     FP_MSB-1,#00        ; preparation of equal exponents
19F5 71C8       5700            ACALL   SHIFT_RIGHT
19F7 7F04       5701            MOV     R7,#LEN_BYTE
19F9 792E       5702            MOV     R1,#FP_LSB
                5703
                5704    ;FPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPE
                5705
                5706    ;       FLOATING POINT ERRORS, FOUND BY D.MUDRIC AND I.STOJSAVLJEVIC
                5707
                5708    ;       MOV     A,#9EH  ;ERROR NUMBER #1
                5709
                5710            ; VALUE IN R4 MUST BE COMPLEMENTED WITH 100D (#9AH), IT MUST BE THE
                5711            ; FIRST COMPLEMENT
                5712
                5713    ;       CLR     C
                5714    ;       SUBB    A,R4
                5715    ;       DA      A
                5716    ;       XCH     A,R4
                5717    ;       JNZ     $+3
                5718    ;       MOV     R4,A
                5719
                5720    ;       ERROR NUMBER #2
                5721
                5722    ;       WITH SUBSTRACTION, AFTER REDUCING BOTH THE  MINUEND  AND  THE
                5723    ;       SUBTRAHEND TO THE SAME EXPONENTS, WHEN R4 <> 0, IT IS OBVIOUS
                5724    ;       THAT ONE ALWAYS HAS TO MAKE A BORROWING FROM THE FIRST HIGHER
                5725    ;       POSITION OF THE MINUEND, NOT AS IT IS STATED BY THE  ORIGINAL
                5726    ;       WHERE IT IS MADE ONLY WHEN R4 => 50H
                5727
                5728    ;       CJNE    A,#50H,$+3          ; deal with rounding
                5729    ;       JNB     FRESER,FP_SUBB      ; test for subtraction (atention to carry)
                5730
                5731    ;FPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPEFPE
                5732
                5733            ; PROPER ROUNDING, DEVELOPED BY D.MUDRIC
                5734
19FB 749A       5735            MOV     A,#9AH
19FD C3         5736            CLR     C
19FE 9C         5737            SUBB    A,R4
19FF D4         5738            DA      A
1A00 CC         5739            XCH     A,R4
1A01 30231E     5740            JNB     FRESER,FP_SUBB
1A04 B45003     5741            CJNE    A,#50H,$+6
1A07 00         5742            DB      0,0,0
1A08 00
1A09 00
                5743    ;       continue normal code
                5744
1A0A B3         5745            CPL     C
1A0B 5119       5746            ACALL   ADD_MANTISSA
1A0C 5008       5747            JNC     JMP_COPY_TO_TOS
1A0E 052A       5748            INC     FP_MSB-1

                                                        910128 - 13
```

Fig. 2. Assembly listing of the improved floating-point nucleus.

OLD:

1080	0E A5 02 06	9F 12 05 73	7B 00 79 07	A3 11 1A 12
1090	05 6D 12 19	A3 12 0E A5	A3 B9 0D 00	40 22 A3 E0
19F0	7F 0A 75 2A	00 71 C8 7F	04 79 2E 74	9E C3 9C D4
1A00	CC 70 01 FC	B4 50 00 30	23 18 B3 51	19 50 08 05
1F00	E4 3B FB 22	E4 90 27 10	F1 21 90 03	E8 F1 21 90
1F10	00 64 F1 21	90 00 0A F1	21 90 00 01	F1 21 60 20
1F20	22 7E FF 0E	CA B5 83 00	CA 40 12 C8	95 82 C8 CA
1F30	95 83 CA 50	EE C8 25 82	C8 CA 35 83	CA 4E 60 E0
1F40	74 30 2E 8B	A0 F3 09 B9	00 01 0B 22	D1 A7 A2 36

NEW:

1080	0E A5 02 06	9F 12 05 73	7B 00 79 07	A3 12 05 6D
1090	12 19 A3 00	00 00 00 00	00 B9 0D 00	40 22 A3 E0
19F0	7F 0A 75 2A	00 71 C8 7F	04 79 2E 74	9A C3 9C D4
1A00	CC 30 23 1E	B4 50 03 00	00 00 B3 51	19 50 08 05
1F00	E4 3B FB 22	7D 00 EA 75	F0 0A 84 FA	E8 54 F0 45
1F10	F0 C4 75 F0	0A 84 C4 FE	E8 54 0F C4	45 F0 C4 75
1F20	F0 0A 84 4E	F8 E5 F0 24	30 0D C0 E0	EA 48 70 D6
1F30	D0 E0 8B A0	F3 09 B9 00	01 0B DD F4	22 00 00 00
1F40	00 00 00 00	00 00 00 00	00 00 00 00	D1 A7 A2 36

910128 - 11

Fig. 3. Overview of address locations to be modified in the 8-KByte interpreter. Remember, you can not overwrite data just like that in an EPROM — a 0 will remain a 0 unless you erase the EPROM using ultra-violet light. The modified version of the interpreter must, therefore, be loaded in a blank EPROM.

The second procedure is based on successive division of the variable value by 10, where, with each division, one BCD digit appears as the remainder.

By simplifying the conversion to successive division-by-10, the original code of 72 bytes is reduced to 57 bytes. The improved hex-to-BCD converter no longer requires the DPTR contents to be stored, returned and incremented after the conversion is finished. As a result, it is faster than the original converter implemented by Intel. Table 1 shows a few 'bench-mark' examples.

Table 1. Hex-to-decimal conversion time

P (hex)	P (dec)	t_{old} (ms)	t_{new} (ms)
0000	0000	141	56
0009	0009	276	56
4000	16384	507	268
EA5F	59999	820	268

Action!

If you are interested in the improvements made to the MCS BASIC-52 interpreter as described here, unload the interpreter from a 8052AH-BASIC, modify it, put it back into an EPROM and run it either with an 80C32, or as a turnkey-EPROM with an 8052AH-BASIC. All this is pretty straightforward and described at length in Refs. 1 and 2. First, copy the interpreter into an EPROM. Next, use the OLD function of the EPROM programmer on the BASIC computer (Refs. 4, 5) to make sure that an exact copy is available. Then load the contents of the EPROM into an EPROM programmer, and use the edit function to make the changes indicated in Fig. 3. Finally, program a new EPROM with the modified BASIC interpreter. The EPROM used may be a 27C64, a 27C128 or a 27C256. Alternatively, you may want to use an EEPROM Type 2864A. ❑

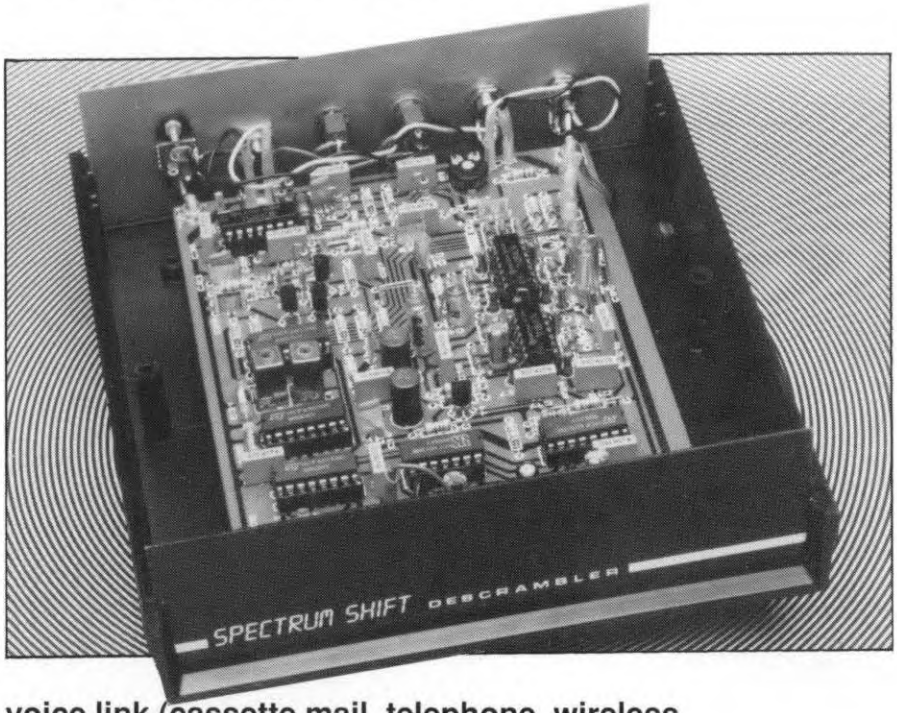
References:

1. "CMOS replacement for 8052AH-BASIC", Elektor Electronics January 1990.
2. "ROM-copy for 8052 BASIC computer",
3. "MCS BASIC-52 Users Manual", Intel Corp., 1986.
4. "BASIC computer", Elektor Electronics November 1987.
5. "8032/8052 Single-board computer", Elektor Electronics May 1991.

DESIGN IDEAS

The contents of this article are based solely on information obtained from the author and do not imply practical experience by *Elektor Electronics*.

AUDIO SPECTRUM SHIFT TECHNIQUES



Although spectrum shift and spectrum inversion have been known for quite some time as a means of protecting voice links mainly over the telephone, little has been written on the design and operation of the electronics involved. This article aims at offering you a design base and two practical circuits to start experimenting with an interesting form of analogue audio scrambling and descrambling which may be used to, say, personalize a voice link (cassette mail, telephone, wireless babysitter, etc.). Remarkably, you will not be able to tell the difference between the encoder and the decoder until you compare the two units and spot a couple of components with different values.

by C. White Halfoat

ONE may wonder why analogue audio scrambling is still used when digital encoding systems are relatively simple to implement whilst offering a good degree of security. The reason is simple: digitized audio data requires a bandwidth which is greater than that of the analogue source signal. Thus, where a limited bandwidth is available to convey an audio signal — say, 3 kHz for voice signals, and 15 kHz for music signals — digital encoding is simply out of the question, and different means must be sought to protect the information as it is conveyed from the transmitter to the licensed receiver.

Encoding principle

An all-analogue encoding system is described that renders an audio signal totally unintelligible by shifting the entire spectrum between 50 Hz and about 10 kHz by 1 to 2 kHz within the existing channel. When the shift is 1 kHz, a source signal of 50 Hz is shifted to 1,050 Hz, while a source signal of 10 kHz is shifted to 11 kHz. In this system, the frequency range below 1,050 Hz is, in principle, empty after encoding. The shift operation is shown diagrammatically in Fig. 1a. Figure 1b shows an alternative version, where the audio spectrum is additionally mirrored around a certain frequency. Both encoding systems yield a new audio signal that is virtually unintelligible, with speech pauses and stress on utterances (i.e., volume variations) as practically the only recognizable features (in fact, we have heard such scrambled signals described as the sound of "*a lot of squabbling chipmunks*").

Each of the two encoding systems described by Figs. 1a and 1b has its advantages and disadvantages. The second system (Fig. 1b) is simple to realize both at the transmitter and the receiver. However, when applied to an FM (frequency modulation)

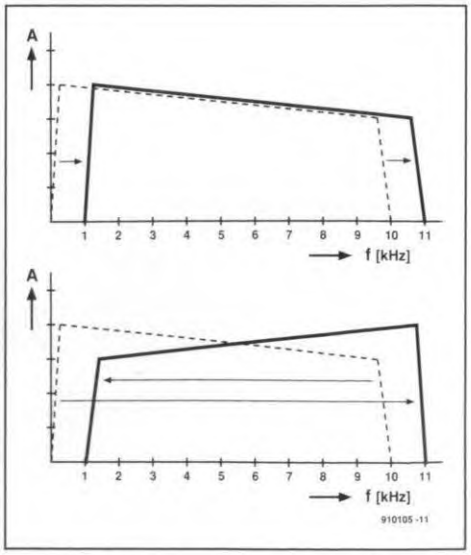


Fig. 1. Principle of spectrum shift (upper drawing) and spectrum inversion (lower drawing). Both techniques are used to scramble audio signals conveyed over air, via cables or magnetic media.

communication link, the signal-to-noise (*S/N*) ratio at the receiver side suffers because the shifted audio signal contains more high-frequency components than the original (source-) signal. By contrast, when the first encoding system (Fig. 1a) is applied, the bulk of the intelligence is carried by components in the lower part of the spectrum, so that *S/N* degrading will not occur (in fact, there is psychoacoustic evidence that most information in speech and music is contained in the frequency range from 300 Hz to 1,000 Hz).

The system whose spectrum is shown in Fig. 1a yields almost unintelligible output signals, and is not easy to decode. The decoding process required to restore the signal to its original spectrum can be subdivided into a number of operations as illustrated in Fig. 2. As shown in Fig. 2a, the first function of the decoder is to limit the bandwidth of the encoded signal, i.e., to fit this into a frequency range from 500 Hz to 10 kHz. The filter slope at the high end of the spectrum prevents intermodulation products being generated in the decoding process, while the high-pass filter prevents whistles between 1 kHz and 2 kHz as a result of low-frequency components that are possibly added to the channel.

The block diagram of the decoder (Fig. 3) shows that the input signal is taken through a low-pass filter before it is amplified and applied to a high-pass filter. A switch connected ahead of the output buffer allows the user to select between encoded and non-encoded signals. When the switch is set to 'non-encoded', the input signal is fed direct to the output buffer, i.e., it is not amplified. This is done to ensure that the signal level of non-encoded signals is not changed by the decoder. Thus, the amplifier merely serves to compensate losses introduced by the phase filters used for encoding or decoding the input signal.

Frequency transforms: Hilbert & Fourier

Shifting an audio spectrum can be achieved in three ways, which are familiar from radio communication techniques used for generating and demodulating SSB (single-sideband) signals. The first two systems are based on narrow-band filters. These are widely applied, and do not present problems with audio signals whose frequency range, for radio communication purposes, extends from 300 Hz to 3,000 Hz only.

Transmitting music via a communication link based on spectrum shifting poses more problems than speech because a much larger frequency spectrum must be conveyed without distortion. It is for this, and other, reasons, that the present encoding system is based on the so-called 'third method', a term familiar to most licensed radio amateurs and radio engineers (Ref. 3).

So how does the encoder work? First, the audio signal is split into two components. Both have the same amplitude, and contain the source signal. The only difference between these components is the phase shift between the frequencies they consist of. This phase shift is 90°, and one of the components is called the Hilbert transform of the other (the two signals are called a Hilbert pair). A Hilbert transform is a frequency shifting operation that unfortunately exists in theory only. In practice, however, a number of systems have been developed that yield close approximations of such an ideal transform. One of these systems is applied in the present encoder.

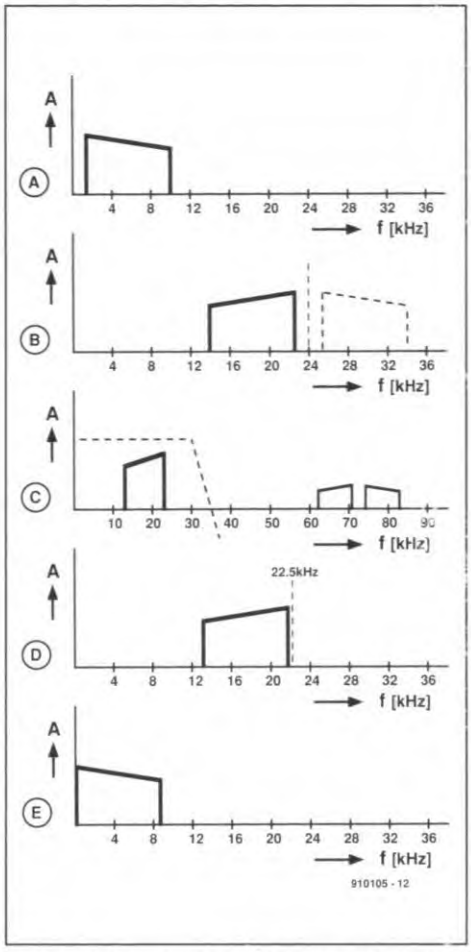


Fig. 2. Multiplication with two phase-locked carriers, filtering and sideband suppression are the main functions into which the decoding process can be subdivided.

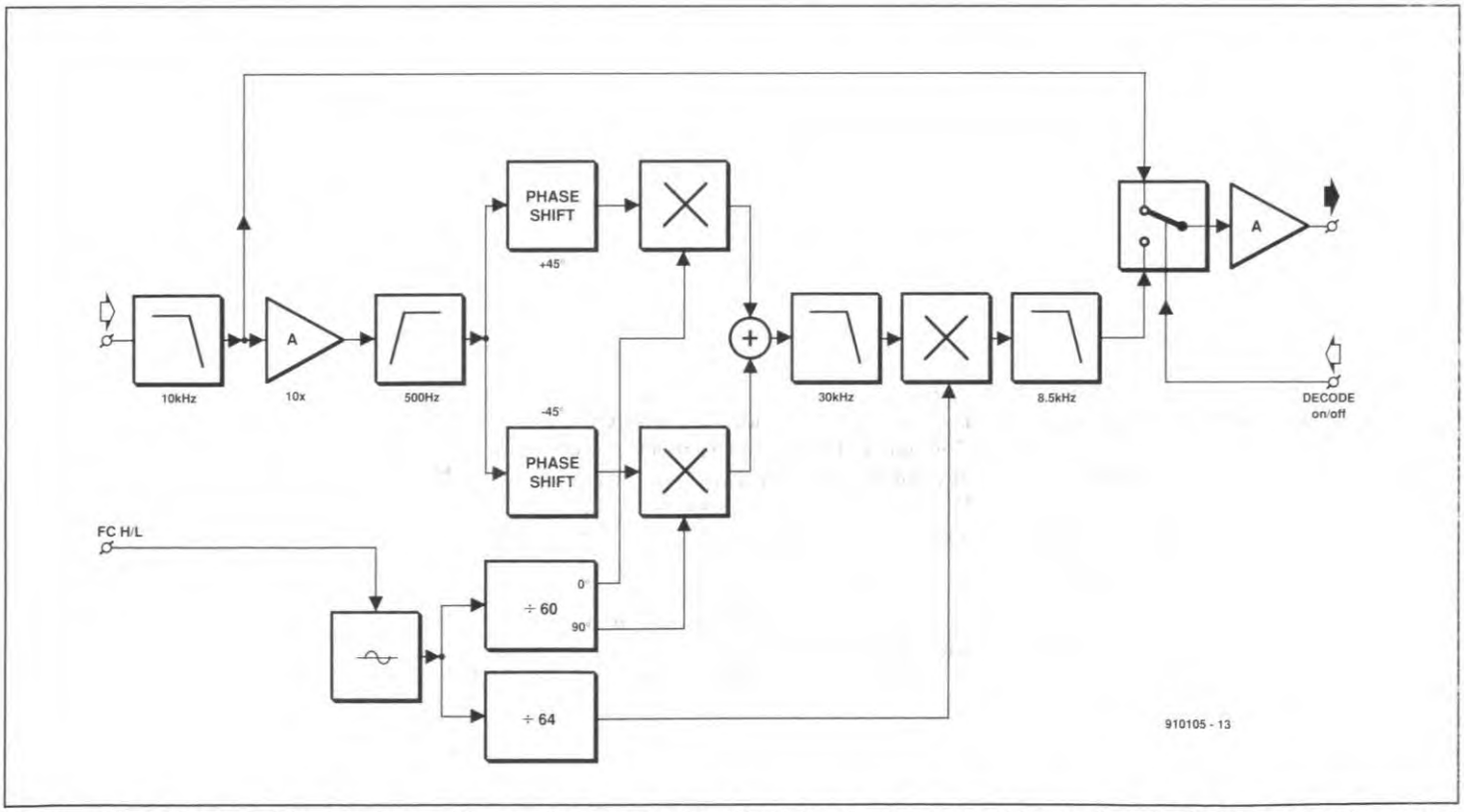


Fig. 3. Block diagram of the frequency shift encoder/decoder. The FC H/L input selects one of two system clocks from which the amount of frequency shift is derived. Note that the encoder/decoder has a bypass function, which is controlled via the S. 0/1 input.

The Hilbert pair is required to suppress one of the sidebands that is generated when the filtered input signal is shifted ('transformed') to a higher frequency. The multiply operation carried out after shifting the phase of the input signal over +45° and –45° gives rise to two double-sideband (DSB) signals. In the decoder, these signals appear around a 24-kHz component — the actual frequency depends on the shift used during the encoding, and will be discussed further on.

As shown in Fig. 2b, one of the sidebands disappears when the two DSB signals are added. The mathematics underlying this operation are shown inset in Fig. 4, a drawing reproduced from an earlier article in this magazine on frequency transformation (Ref. 1). The operations shown in Fig. 4 describe largely what happens in the decoder. The equations go to show that there is no way to avoid a Hilbert transform function when one of the sidebands is to be selected without resorting to (very complex) filters. As shown in Figs. 2b and 2c, the upper sideband is removed in the present design.

Multipliers or double-balanced mixers can take a number of practical shapes. Although their design is mostly plain sailing, they do have to be balanced with the aid of adjustments, and may have insufficient linearity for some applications. Good results at a very small outlay, and without the need of adjustments, can be achieved with electronic analogue bilateral switches, e.g., those contained in the 4053 CMOS multiplexer IC. These switches are particularly well suited to use in the decoder where they enable high suppression of the 24 kHz component to be achieved. When this suppression is insufficient, a 1.5-kHz signal (whistle) would readily appear in the decoded spectrum, corresponding to the amount of frequency shift applied. The main disadvantage of the analogue switches is that they generate a frequency spectrum with spurious components centred around odd-numbered harmonics of the 24-kHz switching signal. It will be recalled that switching is the same as multiplying with a rectangular signal. Thus, in terms of the spectrum, switching can be described as a multiplication with the Fourier series

$$\sin(\omega t) - \tfrac{1}{3}\sin(3\omega t) + \tfrac{1}{5}\sin(5\omega t) - \tfrac{1}{7}\sin\ldots$$

The sidebands of the third harmonic, which are visible in Fig. 2c, are removed with the aid of a low-pass filter designed for a steep roll-off at 30 kHz. This is done to prevent the sidebands causing interference in the mixing operation that restores the original frequency band.

The actual decoding process is relatively simple. As shown in Fig. 2d, the encoded signal is subjected to a second multiply operation, this time with a frequency of 22.5 kHz. The equations in Fig. 4 once more prove that the decoding operation works and yields the desired output signal.

Finally, the decoded signal is taken through a low-pass filter to remove spurious components generated by, among others, the switching mixers.

Which frequencies?

The frequency shift of 1.5 kHz used in the decoder is achieved by first shifting the spectrum to 24 kHz. Next, the spectrum is shifted down again with the aid of a signal of 22.5 kHz, or the difference between 24 kHz and 1.5 kHz. This restores the spectrum to the frequency range it has at the input of the encoder (which may be thousands of miles away, or just round the corner). In other words, we have decoded the scrambled signal. Note that it is also possible to shift the spectrum 'direct' by 1.5 kHz, but not when analogue switches are used as mixers.

Even though a special kind of frequency synthesis had to be implemented to ensure the required stability of the frequency transform, the 'third method' is used in the decoder for reasons already mentioned. For example, a stability of ±5 Hz is equal to 0.33% at 1,500 Hz. However, the 1,500 Hz signal is not generated directly, but indirectly as the difference between 24 kHz and 22.5 kHz. At 24 kHz, an error of ±5 Hz is equal to 0.021%, while at 22.5 kHz it equals 0.022%. When these two switching frequencies are generated independently, the accuracy of each of them must be 0.01% to ensure that a total error of ±5 Hz is not exceeded. It is readily seen that 0.01% is a pretty tough requirement as regards stability, particularly when the relevant oscillators are to be designed as free-running types to enable a free choice of the shift frequency. However, it can be shown that if these two frequencies have fixed phase relation, the common (central) oscillator from which they are derived is good enough if it has a relative deviation equal to the maximum permissible error at about 1,500 Hz, which is 0.33%. In the present decoder, this is realized with the aid of a central oscillator in conjunction with two digital dividers. Basically, the first divider is set for a divisor of 15, the second for a divisor of 16. The frequency of the central clock is

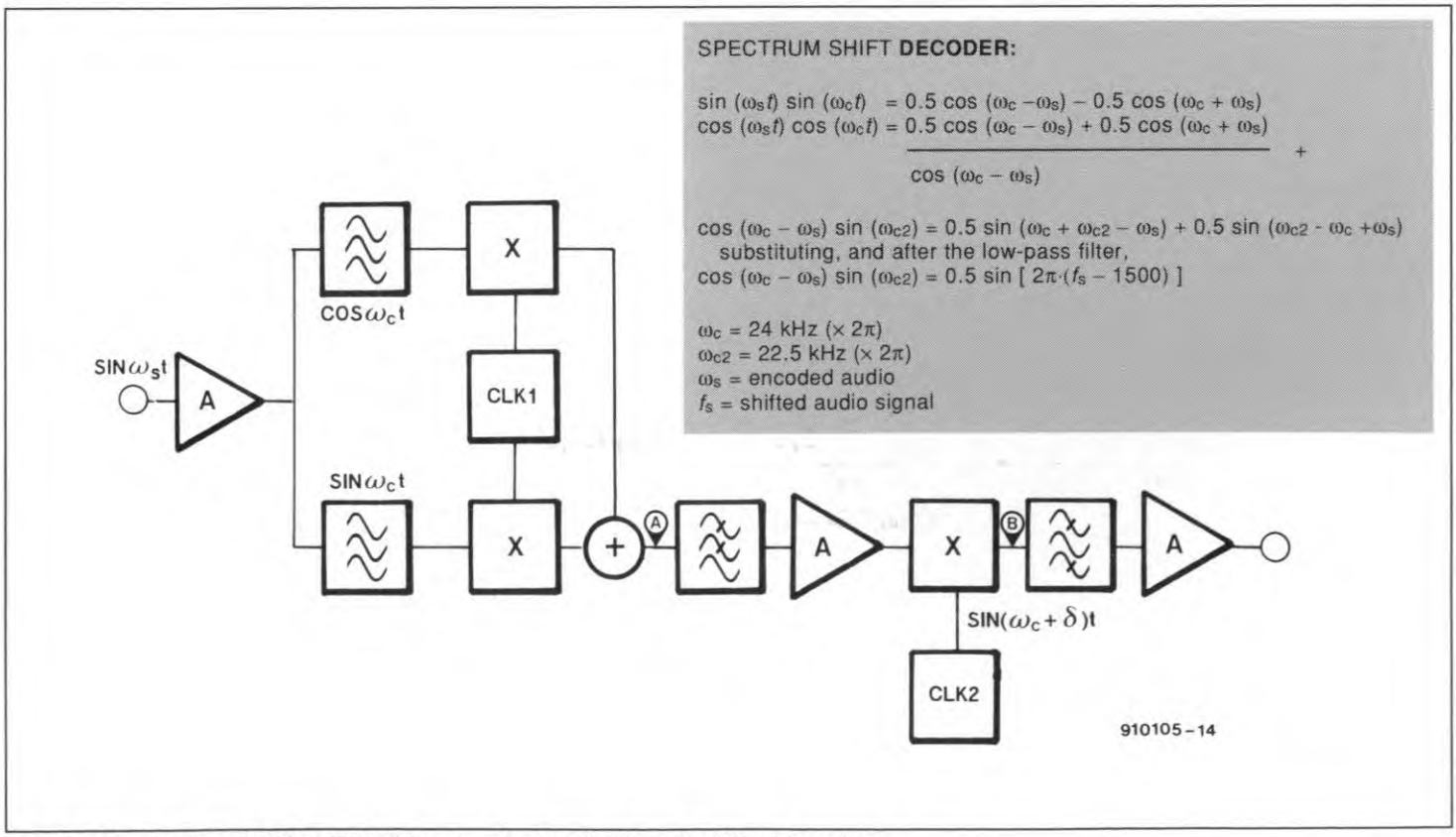
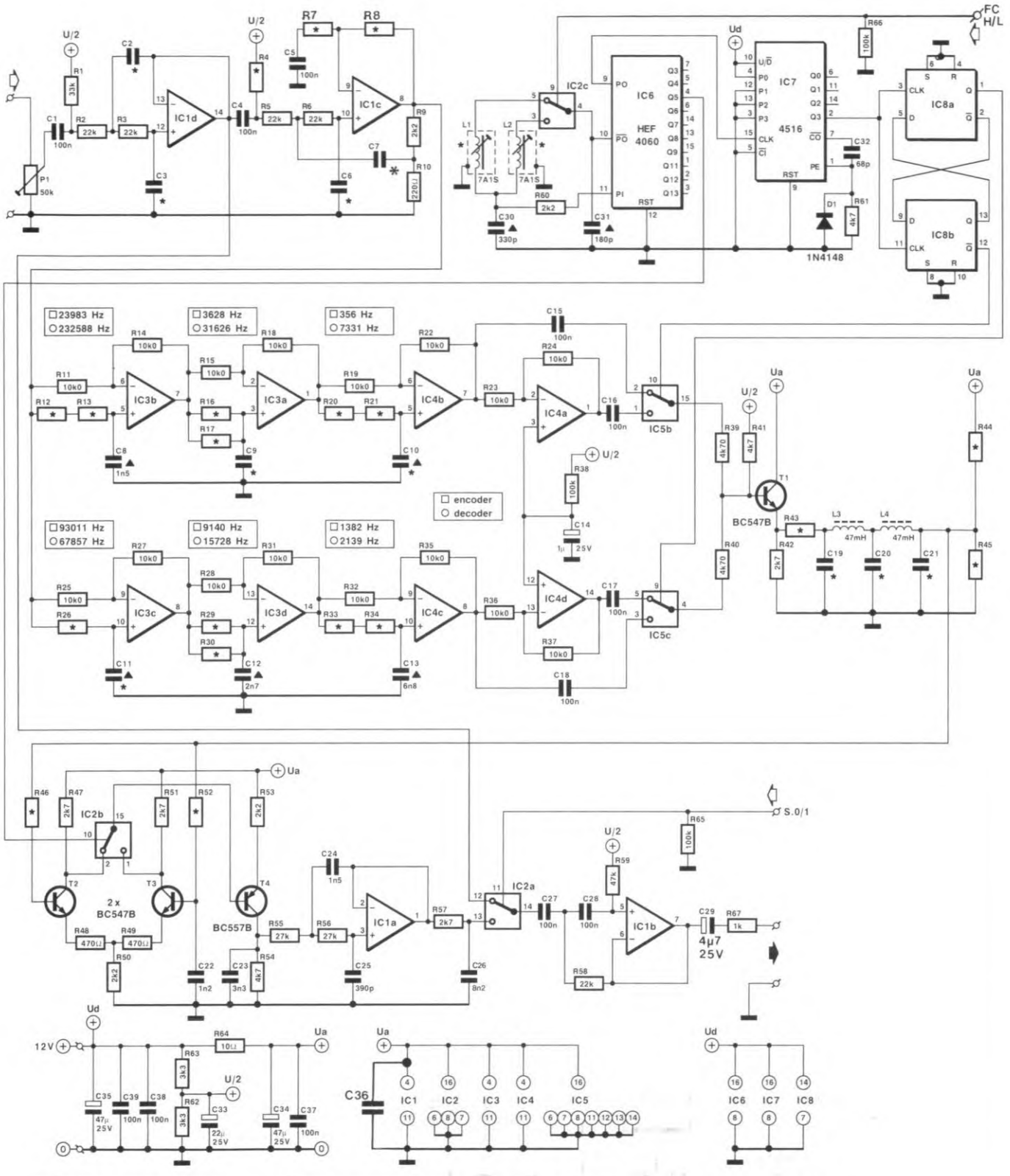


Fig. 4. The 'Feedback killer' (Ref. 1) is a practical realization of a Hilbert transform.



Pos.*	Encoder	Decoder
C2	1n2	820p
C3	1n	680p
C6	470p	1n
C7	2n7	1n8
C9	6n8 2.5%	2n7 2.5%
C10	24n 2.5%	6n8 2.5%
C11	1n5 2.5%	2n7 2.5%
C19	330p	560p
C20	1n	1n8
C21	330p	560p

Pos.*	Encoder	Decoder
R4	10k 5%	3k3 5%
R7	10k 5%	3k3 5%
R8	100k 5%	33k 5%
R12	25k5 1%	2k49 1%
R13	2k37 1%	390Ω 5%
R16	43k2 1%	12k1 1%
R17	680k 5%	390k 5%
R20	100k 1%	10k 1%
R21	16k2 1%	10k 1%
R26	7k15 1%	5k49 1%

Pos.*	Encoder	Decoder
R29	43k2 1%	24k3 1%
R30	680k 5%	820k 5%
R33	100k 1%	47k5 1%
R34	6k8 5%	21k5 1%
R43	6k8 5%	5k6 5%
R44	18k 5%	15k 5%
R45	18k 5%	15k 5%
R46	27k 5%	22k 5%
R52	27k 5%	22k 5%

IC1, IC3, IC4 = TL084
IC2, IC5 = 4053
IC8 = 4013

★ see table
▲ = 2.5% polystyrene

910105 - 15

Fig. 5. Circuit diagram of the spectrum shift encoder/decoder.

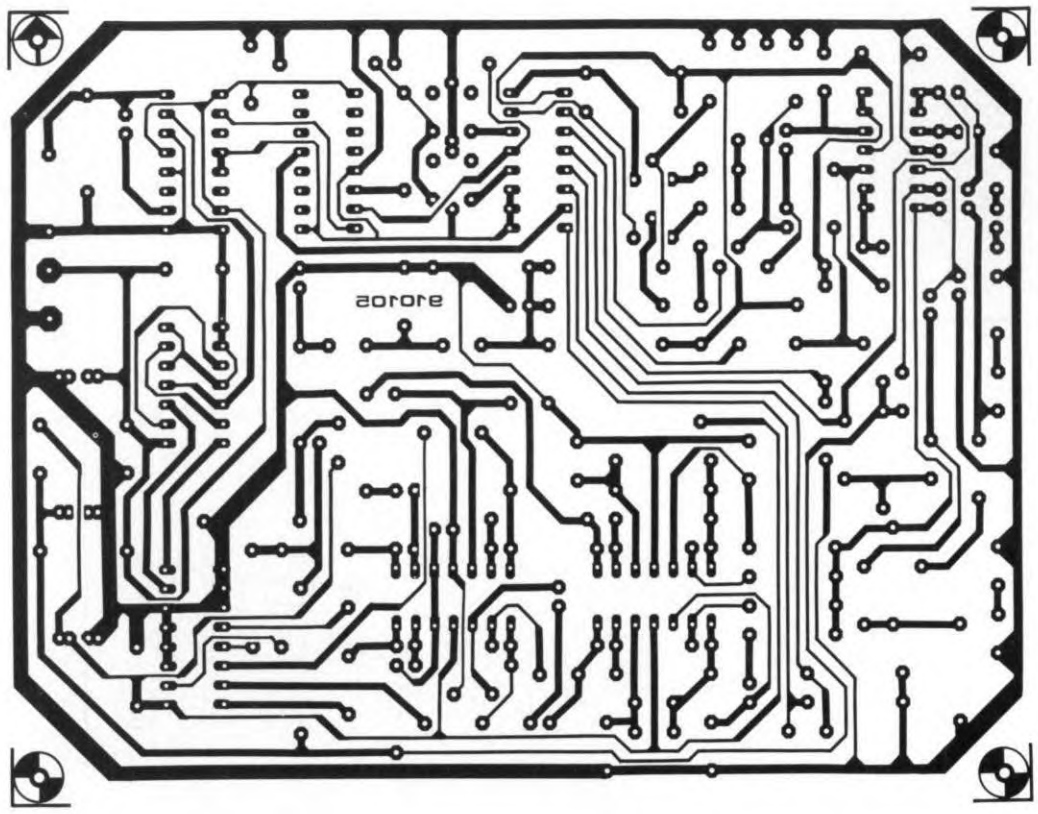


Fig. 6a. Track layout (mirror image) of the single-sided printed circuit board designed for the spectrum shift encoder/decoder.

calculated as follows. Assuming a spectrum shift of 1,500 Hz, then:

$$f_{osc} / 15 - f_{osc} / 16 = 1{,}500$$
$$\therefore 16 f_{osc} / 240 - 15 f_{osc} / 240 = 1{,}500$$
$$\therefore f_{osc} / 240 = 1{,}500$$
$$\therefore f_{osc} = 360 \text{ kHz}.$$

Since the 24-kHz frequency must be available as four signals with a phase shift of 90°, i.e., 0°, 90°, 180° and 270°, the actual oscillator frequency is four times higher than calculated, i.e., 1.44 MHz, while the divisors of 15 and 16 change to 60 and 64 respectively. This is illustrated in Fig. 3. The 24-kHz signal is fed in quadrature to two identical mixers (4053). After adding the results of the multiplication, the upper sideband is compensated. Via a 30-kHz low-pass filter, the signal is fed to the third multiplier where it is transformed to the original frequency band.

The encoder

So far, we have not discussed the encoder, and this has a very good reason: it is virtually identical to the decoder! In fact, the same circuit is used, which allows the encoder and the decoder to be built on identical printed-circuit boards. The only difference between the two units is the value of 29 passive components, of which a number determine the frequency response of the phase filter built around two opamps, IC3 and IC4 (see the circuit diagram in Fig. 5). These components define a pass-band of 1 kHz to 10 kHz in the decoder, and a pass-band of 120 Hz to 7 kHz in the encoder. The accuracy of the phase filter is determined by the accuracy of the passive components used. The performance of the filter depends on the number of sections it consists of, and the ratio between the highest and the lowest input frequency. This ratio is about 10 in the decoder, which, based on a phase filter with three sections, achieves an *S/N* ratio of about 60 dB. Also based on three filter sections, but designed to handle a max/min frequency ratio of about 58 (7,000 Hz/120 Hz), the encoder has a worse *S/N* performance of about 40 dB, which is not bad given the simplicity of the design. In any case, the results are perfectly acceptable for speech.

The shift of +1.5 kHz in the encoder (instead of –1.5 kHz in the decoder) is achieved by suppressing the lower sideband instead of the upper sideband (see also Fig. 2b). Since the spectrum in Fig. 2b then extends from 24 kHz to about 31 kHz, the roll-off frequency of the 30-kHz low-pass filter must be shifted up to about 36 kHz. In addition, the input and output filters must be adapted to the spectrum to be passed. All the changes necessary to effect this, i.e., to turn the decoder into an encoder (or vice versa), are summarized in the components table you find inset in Fig. 5.

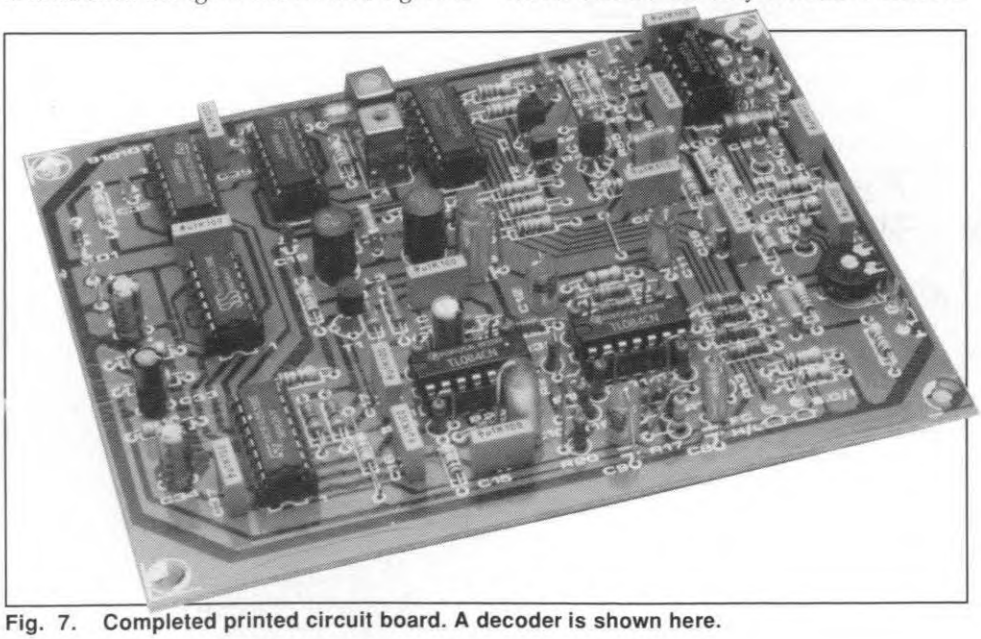


Fig. 7. Completed printed circuit board. A decoder is shown here.

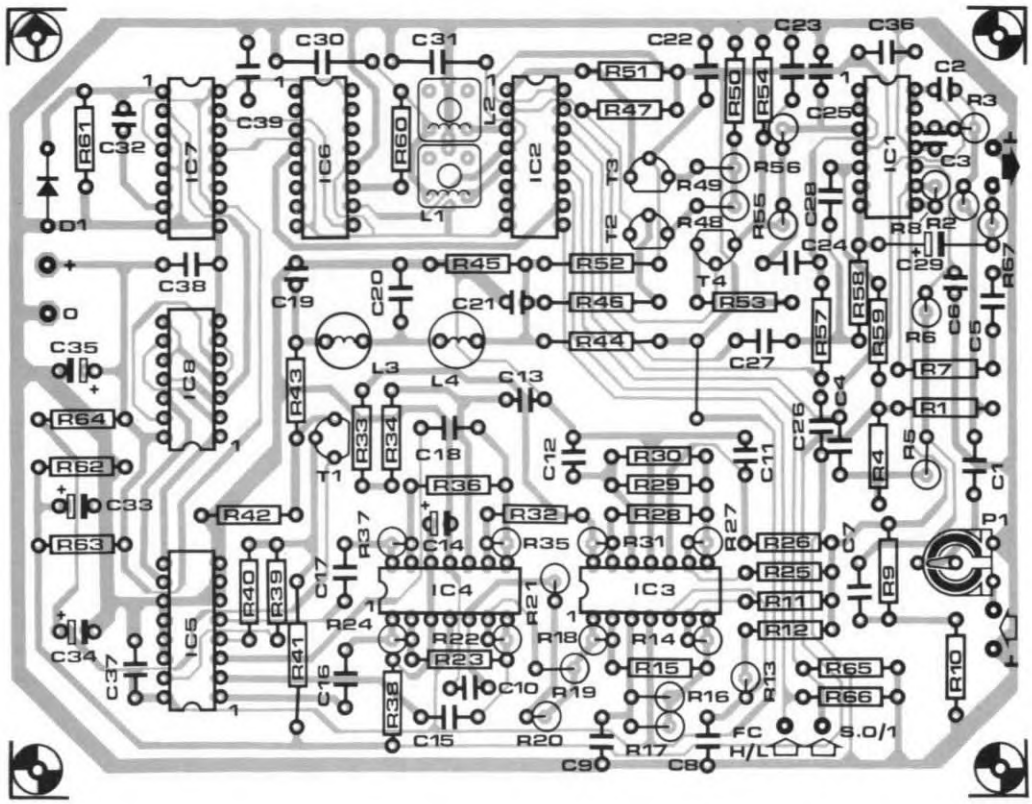


Fig. 6b. Component mounting plan.

COMPONENTS LIST

Resistors:

1	33kΩ	R1
5	22kΩ	R2;R3;R5;R6;R58
4	2kΩ2	R9;R50;R53;R60
1	220Ω	R10
16	10k0 1%	R11;R24;R14;R15; R18;R19;R22;R23; R27;R28;R31;R32; R35;R36;R37;R25
3	100kΩ	R38;R65;R66
2	4kΩ70 1%	R39;R40
3	4kΩ7	R41;R54;R61
4	2kΩ7	R42;R47;R51;R57
2	470Ω	R48;R49
2	27kΩ	R55;R56
2	3kΩ3	R62;R63
1	10Ω	R64
1	47kΩ	R59
1	1kΩ	R67
1	50kΩ preset H	P1

Capacitors:

13	100nF	C1;C4;C5;C15; C16;C17;C18; C27;C28;C36-C39
1	1nF5 2.5% polystyrene	C8
1	2nF7 2.5% polystyrene	C12
1	6nF8 2.5% polystyrene	C13
1	1µF 25V radial	C14
1	1nF2	C22
1	3nF3	C23
1	390pF ceramic	C25
1	1nF5	C24
1	8nF2	C26
1	4µF7 25V axial	C29
1	330pF 2.5% polystyrene	C30
1	180pF 2.5% polystyrene	C31
1	68pF ceramic	C32
1	22µF 25V radial	C33
2	47µF 25V radial	C34;C35

Inductors:

2	7A1S assembly (Neosid); for winding details see text	L1;L2
2	47mH radial choke; Toko 181LY473	L3;L4

Semiconductors:

1	1N4148	D1
3	TL084	IC1;IC3;IC4
2	4053	IC2;IC5
1	HEF4060*	IC6
1	4516	IC7
1	4013	IC8
3	BC547B	T1;T2;T3
1	BC557B	T4

* do not substitute by CD4060, HCF4060 or MC14060

Miscellaneous:

1	printed-circuit board	910105

ENCODER/DECODER COMPONENT VALUES

Pos.	Encoder	Decoder
C2	1nF2	820pF
C3	1nF	680pF
C6	470pF	1nF
C7	2nF7	1nF8
C9	6nF8 2.5%	2nF7 2.5%
C10	24nF 2.5%	6nF8 2.5%
C11	1nF5 2.5%	2nF7 2.5%
C19	330pF	560pF
C20	1nF	1nF8
C21	330pF	560pF
R4	10kΩ 5%	3kΩ3 5%
R7	10kΩ 5%	3kΩ3 5%
R8	100kΩ 5%	33kΩ 5%
R12	25kΩ5 1%	2kΩ49 1%
R13	2kΩ37 1%	390Ω 5%
R16	43kΩ2 1%	12kΩ1 1%
R17	680kΩ 5%	390kΩ 5%
R20	100kΩ 1%	10kΩ 1%
R21	16kΩ2 1%	10kΩ 1%
R26	7kΩ15 1%	5kΩ49 1%
R29	43kΩ2 1%	24kΩ3 1%
R30	680kΩ 5%	820kΩ 5%
R33	100kΩ 1%	47kΩ5 1%
R34	6kΩ8 5%	21kΩ5 1%
R43	6kΩ8 5%	5kΩ6 5%
R44	18kΩ 5%	15kΩ 5%
R45	18kΩ 5%	15kΩ 5%
R46	27kΩ 5%	22kΩ 5%
R52	27kΩ 5%	22kΩ 5%

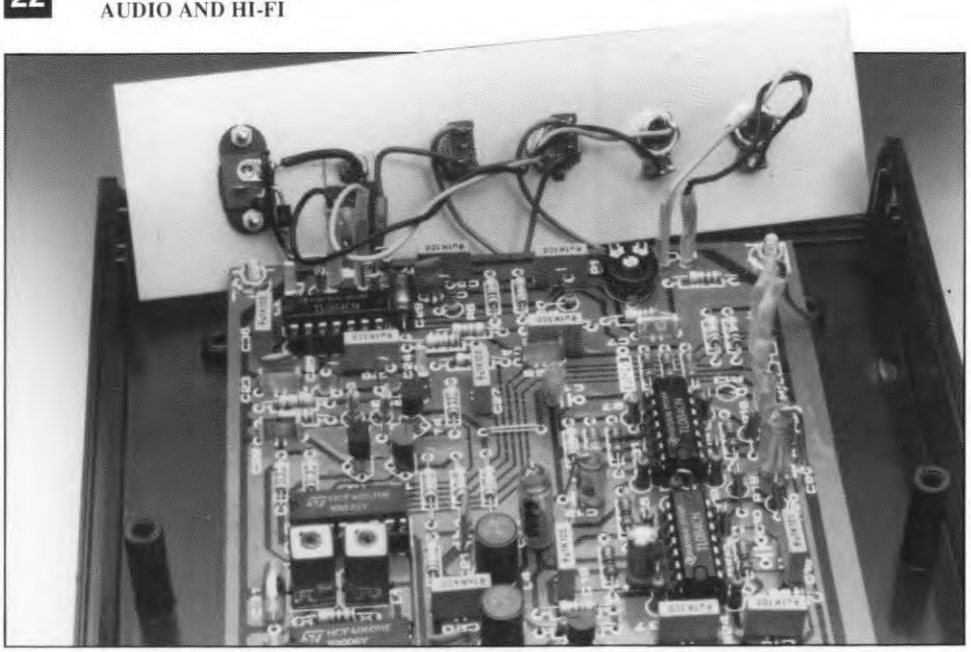


Fig. 8. The rear panel of the Teko 222 enclosure has on it the frequency select switch (FC H/L), the scrambling on/off switch (S. 0/1), the input and output sockets, a DC adaptor socket, and a 7812 voltage regulator with decoupling capacitors.

The circuit in detail

The following description applies to the encoder as well as to the decoder, which are electrically identical circuits. Where necessary, a distinction will be made.

The audio signal applied to the input of the circuit is fed to a low-pass filter built around IC1d and IC1c. The roll-off frequency is set to 10 kHz in the decoder, and to 6.6 kHz in the encoder. The previously mentioned high-pass filter is formed by capacitors C1, C4 and C5 and resistors R1, R4 and R7. The roll-off frequency is set to 500 Hz in the decoder, and to 150 Hz in the encoder. Apart from functioning in the active filter, IC1c amplifies the input signal about 10 times.

The phase filter is formed by the circuit around opamps IC3 and IC4. The indicated component values must be maintained to ensure the correct operation of the filters. The frequency-determining capacitors are types from the E12 series with a tolerance of 2.5% or better. Here, polystyrene types are used in combination with 1% resistors from the E96 series. The theoretical roll-off frequencies of every filter section are indicated in the circuit diagram. The equation

$$f_c = 1 / RC$$

applies for the roll-off frequency, and may be used to match capacitors and resistors if you can not secure the ones given in the circuit diagram and the parts list (consult Ref. 2). The resistors must not be made smaller than about 2 kΩ, or larger than about 200 kΩ. For instance, if you can not get hold of a 24nF 2.5% capacitor (C10 in the encoder), look at the pole of the filter section, which is 356 Hz. Verify this using the above equation and given that R=116.2 kΩ (R20+R21). Next, substitute C10 by, say, 27 nF 2.5%. The total resistance of R20+R21 works out at 103.2 kΩ. Since resistors are available in far more values than capacitors, this value is easier to create than 24 nF.

The multiplication with the 24-kHz signal is effected by electronic switches IC5b and IC5c. Two 1% resistors, R39 and R40, add the product signals. Transistor T1 functions as an impedance transformer to ensure that low-pass filter L3-L4-C19-C20-C21 is terminated correctly. Next, the filtered signal is fed to a second multiplier built around T2, T3 and IC2b. This circuit supplies the decoded signal in the decoder, and the encoded signal in the encoder. A low-pass filter fitted at the input of opamp IC1a suppresses the unwanted high-frequency products, and provides the drive signal for the output buffer, IC1b. An electronic switch ahead of IC1b allows you to disable the decoder/encoder by applying +12 V to the S.0/1 (scrambling on/off) terminal.

The central clock oscillator is contained in IC6, an HEF4060 CMOS oscillator/divider, which is used with an inductor, L1 or L2, instead of the rather more usual *R-C* combination, to determine the oscillator frequency. Properly constructed, the *L-C* oscillator meets the 0.33% accuracy requirement. Note that capacitors C30 and C31 must be types with a low temperature co-efficient (NP0- or COG-class devices; these are ceramic capacitors with a black band at the top side of the body). Alternatively, fit 2.5% tolerance polystyrene types, which are more readily available. The presence of two oscillator inductors enables you to switch between two different frequency shifts. The selection is effected via switch IC2c, which is controlled by applying +12 V or 0 V to the FC H/L (frequency control high/low) input of the board. This is also useful when a single scrambled source makes use of two encoding frequencies.

The HEF4060 combines the function of central oscillator and divide-by-64 prescaler to supply the 22.5-kHz switching signal. The divide-by-60 function is realized by a 4516, IC7, and 4013, IC8. The former is wired to divide by 15, the latter to divide by 4. The Q and $\overline{Q}$ outputs of the bistables in the 4013 supply the control signals for multipliers IC5b and IC5c.

Construction and adjustment

As already mentioned, the printed-circuit board (Fig. 6) can be used to build the decoder as well as the encoder. The construction itself is straightforward, and not expected to cause problems. Note that there is one wire link on the board, and that a fair number of parts are fitted vertically. Make sure you select the right component values from the parts list when building up the decoder and/or the encoder. The value coding of the 1% E96-series resistors may cause confusion if you have never used these devices before — when in doubt, use your digital multimeter to establish the value.

The inductors, L1 and L2, consist of 100 turns of 0.1-mm dia. (SWG40) enamelled copper wire on Type 7A1S formers from Neosid. After winding the inductor and soldering the wires to the appropriate pins (look on the component overlay!), fit the ferrite cup, and secure it with a few drops of wax or glue. This **must** be done to ensure the stability of the oscillator. Next, fit the screening can and the core, check the continuity of the inductor at the base pins, and mount the assembly on the PCB. Solder all the pins, including those of the screening can, quickly to prevent overheating. The self-inductance of L1 and L2 is about 100 µH.

The central oscillator may not work properly if you do not use an HEF4060. The HEF4060 is a Philips Components product, and a LOCMOS version of the ubiquitous CMOS 4060. We tested a few 'ordinary' 4060s of different makes in the circuit, and found that none of these achieved the reliability and stability of the HEF4060.

Apply 12 V to the decoder, and an encoded input signal. Connect an amplifier to the output. Adjust P1 for an input amplitude of about 200 mV_{pp}. Leave the FC H/L and S.0/1 pins unconnected, and adjust L1 until you measure 1.44 MHz at pin 9 of the HEF4060 (IC6). Remember, the central clock frequency determines the shift, which you are free to set between 1 kHz and 2 kHz. A shift of 1,500 Hz is obtained with a 1.44 MHz clock. Take the FC H/L pin logic high (+12 V), and adjust L2 similarly. Listen to the output signal, and fine-tune the inductor(s) for best results. Your ears are the best test instruments for this purpose.

The adjustment of the encoder is identical — the only difference in the set up is that the output is connected to a transmitter, a cassette recorder, etc., instead of to an amplifier for direct sound reproduction. ■

References:

1. "Feedback killer", *Elektor Electronics* February 1990.

2. "Wideband phase-shift networks", in *Electronic Filter Design Handbook* (1981), by Arthur B. Williams, McGraw-Hill Publishers.

3. "HF transmitters", in *Radio Communication Handbook*, published by the Radio Society of Great Britain (RSGB).

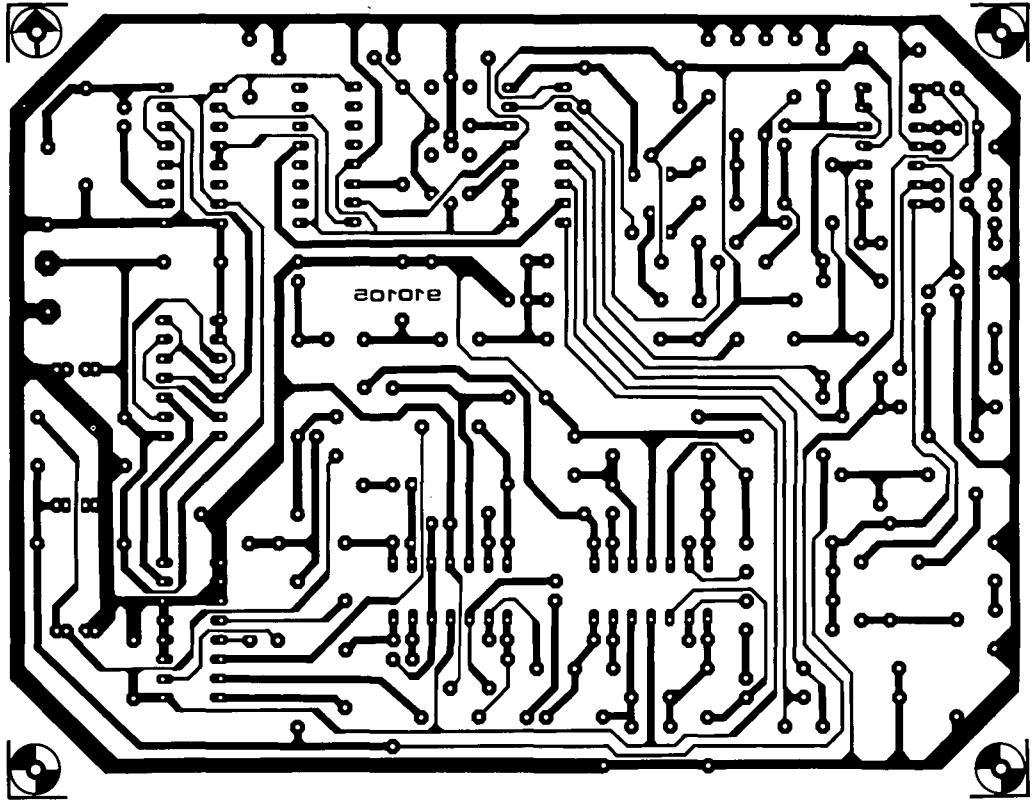


Fig. 6a. Track layout (mirror image) of the single-sided printed circuit board designed for the spectrum shift encoder/decoder.

INTERMEDIATE PROJECT

A series of projects for the not-so-experienced constructor. Although each article will describe in detail the operation, use, construction and, where relevant, the underlying theory of the project, constructors will, none the less, require an elementary knowledge of electronic engineering. Each project in the series will be based on inexpensive and commonly available parts

BUILD THE OPTICALOCK – PART 2

by Michael Swartzendruber

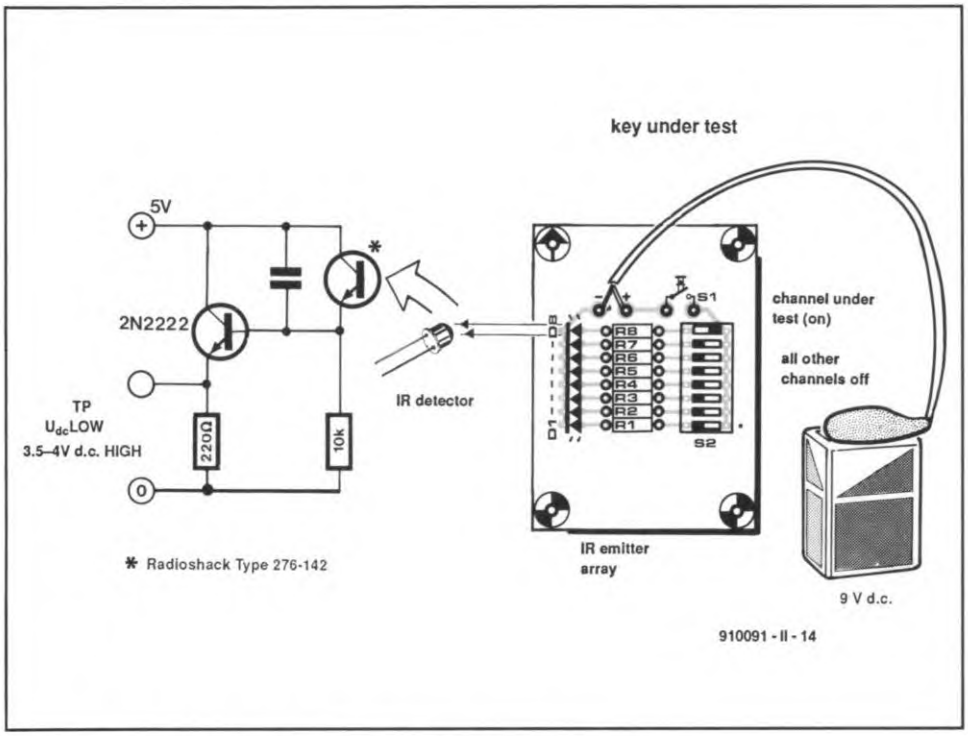


Fig. 7. Auxiliary circuit for testing the key assembly.

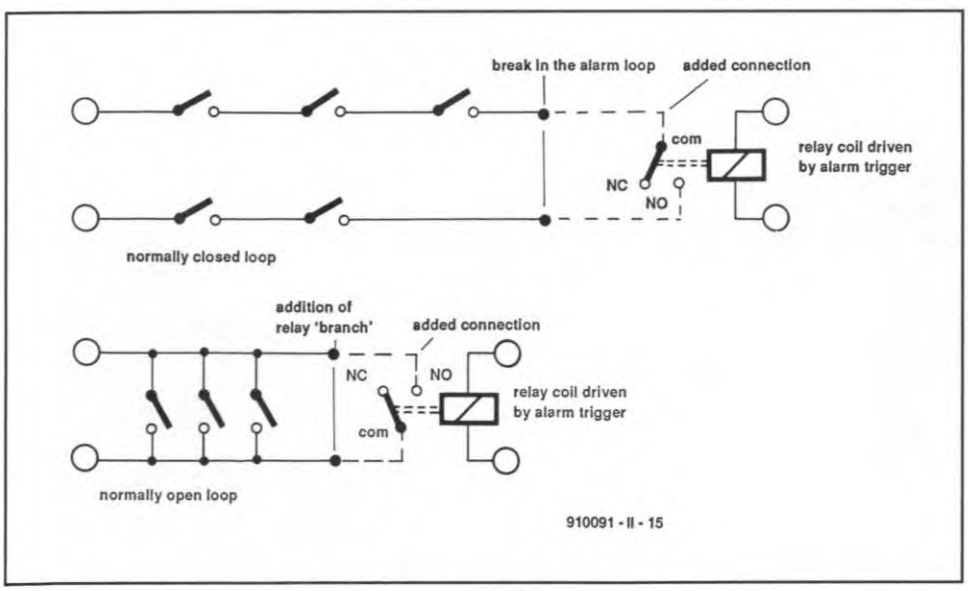


Fig. 8. How to use relays as trigger switches.

Test procedures for the key assembly

1) Construct the circuit shown in Fig. 7.
2) Enable one of the emitter circuits at a time by leaving all DIP switches off, except the channel currently on test: that channel's switch should be turned on.
3) Aim the emitter on test at the test circuit.
4) Look at the voltage or logic level of test point A.
5) If the emitter passes test, disable this channel and proceed to the next channel until all have been tested.
6) Trouble-shoot as necessary. If all emitters fail, double-check the battery connections, or check the device polarity for correct installation on to the board, or check the device with a junction tester.

Test procedures for the detector/amplifier array

1) After the assembly is completed and passes all visual inspections and voltage tests, install the two transistor arrays and connect the assembly to a 5-V d.c. supply.
2) The assembly is tested with the aid of the key assembly. Enable all emitters on the key by setting all DIP switches to 'on'.
3) Locate the test node for the first amplifier channel and place the probe of the voltmeter or logic probe at this node.
4) Aim the key at the completed detector array (these two assemblies should have been constructed so that the key array mates with the detector array) and enable the key by closing the push-button switch.
5) The voltage or logic level at the node of the channel on test should swing to logic level 1 or about 4 V and should remain at this level as long as the key is aimed at the detector. When the key is removed, the logic or voltage level should return to zero.
6) Proceed to the next channel of the array and repeat the test. Trouble-shoot as necessary.
7) If all detectors fail the test, check the wiring of the supply, all polarized devices for

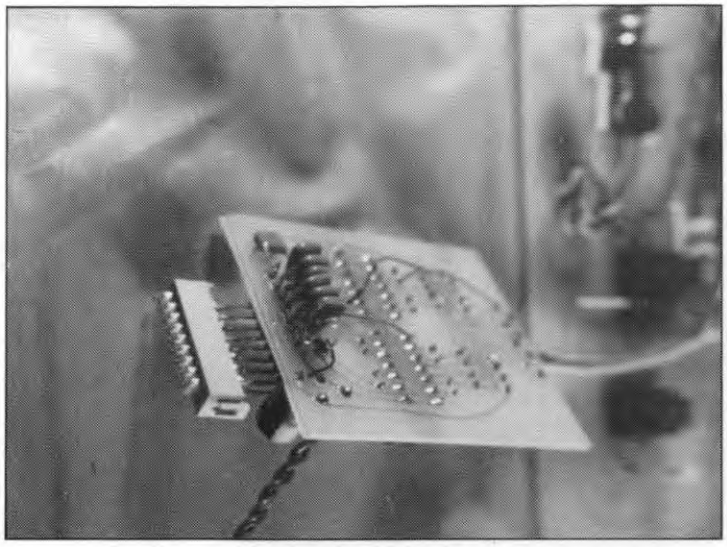


Fig. 9. Mounting method of the optional bypass capacitors connected across each photodetector of the infra-red detector array should they be required to remove fluorescent interference.

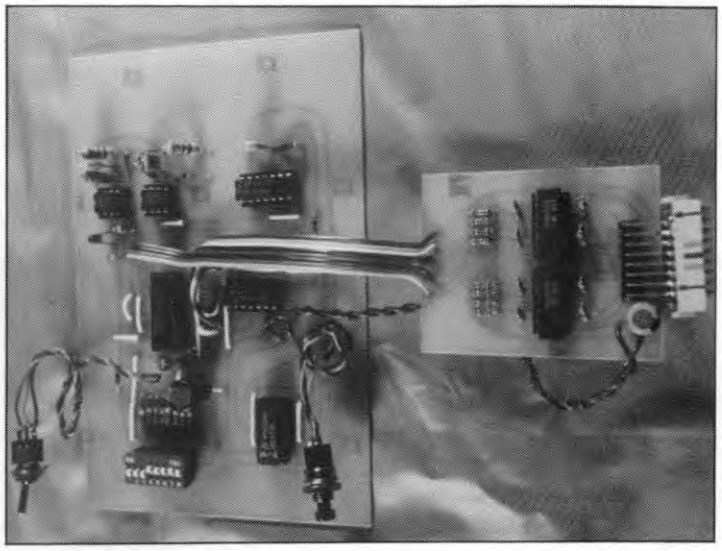
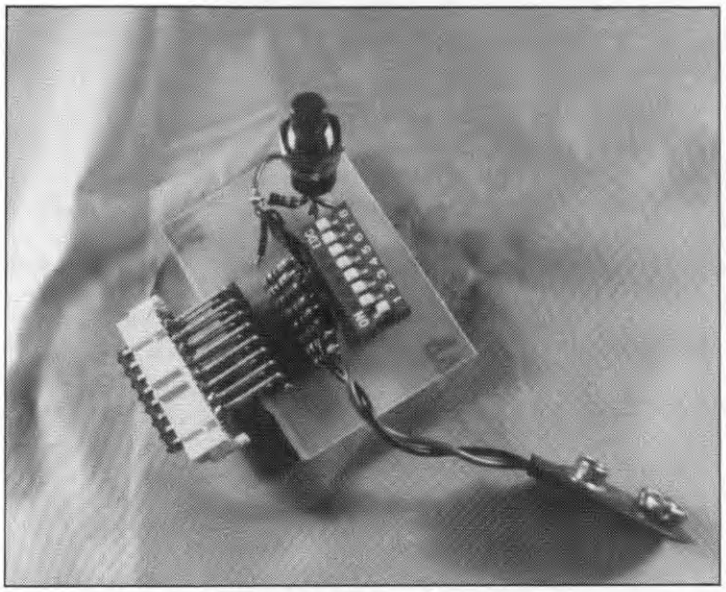


Fig. 10. View of the Opticalock's main logic board and the 'keyhole' connected by a ribbon cable.



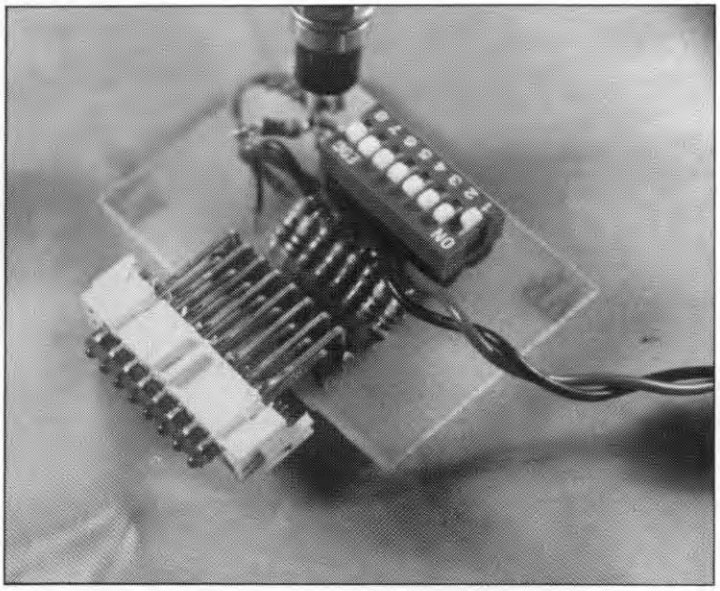


Fig. 11. Two views of the Opticalock's 'key'. The one on the right shows a detail of the method used to mount the infra-red array at a right angle to the board. A 20-pin, 0.1"×0.1" connector is soldered to a right-angle header assembly and the array is inserted into the socket.

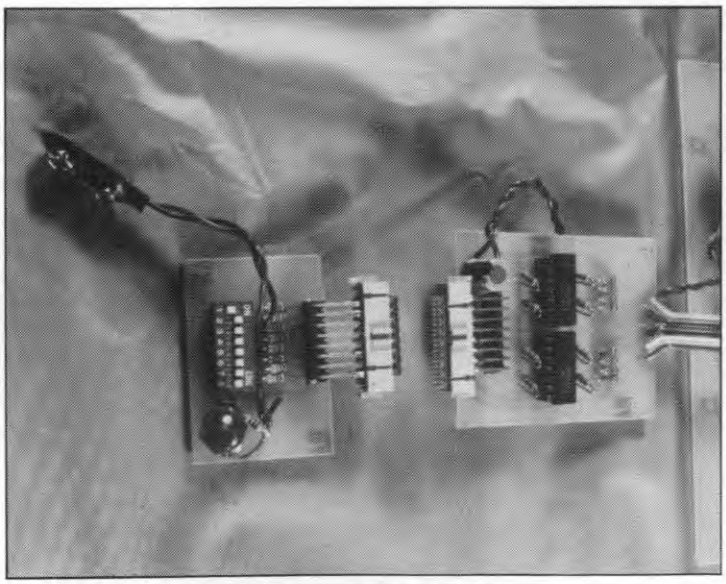


Fig. 12. General view of the key and keyhole.

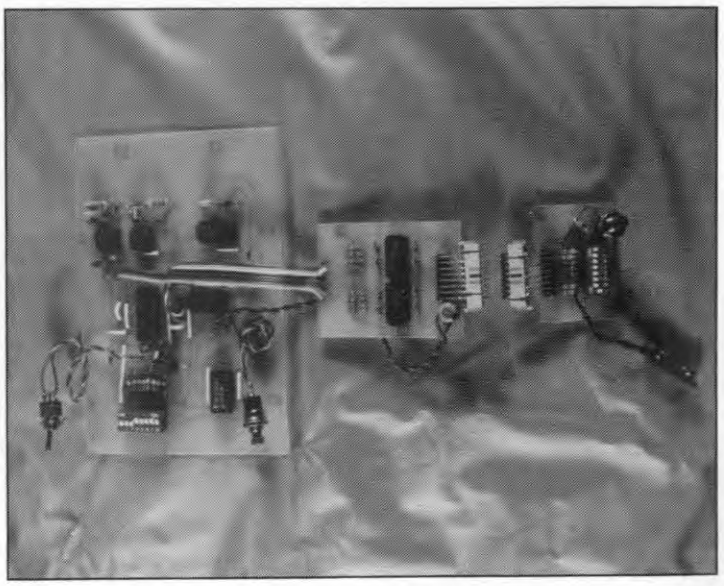


Fig. 13. General view of the entire Opticalock system.

correct placement with particular attention to the infra-red detector array.

Test points for the infra-red detector/amplifier can be found at the following points of the circuit.

Channel	IC	Pin
1	1	12
2	1	3
3	1	5
4	1	10
5	2	12
6	2	3
7	2	5
8	2	10

Interfacing the logic driver board

The relay driver outputs of the logic driver circuit can be used in a number of ways. Some simple examples are given to aid any custom applications.

Example 1. How to use the lock relay to drive a solenoid deadbolt.

A) Solenoid coil can be driven by 24 V d.c. at 250 mA. Directly power a solenoid coil with PwrDrv-1 by connecting the solenoid coil leads to 'load +' and 'load –' pads next to IC_8, which becomes actuated by the lock driver circuitry.

B) Load is high voltage or high current load. Use 'load +' and 'load –' pads adjacent to IC_8 or IC_7 to actuate a 24 V d.c. coil relay or triac. Use the switch contacts or device for load isolation and control.

Example 2. How to use the relays as trigger switches.

To interface the alarm relay into an existing alarm loop, use the appropriate relay contacts and insert them into the loop—see Fig. 8.

For instance, to hook the Opticalock's alarm relay into a parallel, normally closed loop, connect the the normally closed and common contacts of the relay into the loop and simply add the relay contacts as another parallel branch.

To interface the relay into a series (normally open in most cases), break the loop at some convenient point and splice the relay into the series by using the normally open and common contacts.

Example 3. How to use the unlock relay to provide an alarm enable/disarm switch.

To use this technique, a method must be developed to hold a voltage level high or low without having to excite the relay continuously. One very easy way of doing this is to use the normally open and common contacts of a relay to trigger a flip-flop (bistable)—see Fig. 9.

Programming and initial system testing

Set the combination of the DIPswitch on the key to match the combination of the DIPswitch on the main logic board. Align the key with keyhole and press the push-button key. The LED indicating that the lock is 'unlocked' should light.

Change the setting of one of the DIPswitches and press the push-button on the key three times, whereupon the alarm drive should become actuated. If you wish, you may elect to use a logic probe on the Q outputs of the JK bistables (flip-flops) while performing the three illegal-entry attempts to monitor the counting action of this chip.

These tests should give a pretty good idea as to how the system works. Switch S_5 may be used to disable the 'unlocked' output altogether to prevent the circuit from providing an unlock signal even when someone has the IR key with the correct combination.

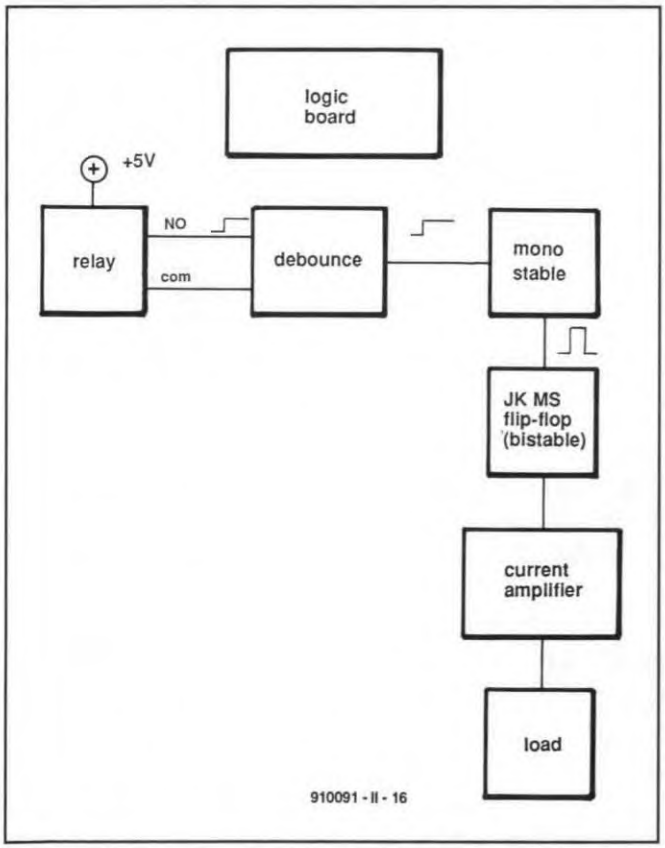


Fig. 14. How to use the unlock relay to provide an alarm enable/disarm switch.

Finishing touches and installation

Now the electronic part of the combination lock has been completed, we need to look at some installation ideas to make the project a useful addition to an existing security system or to help fortify some accessible doorway by providing a pick-proof lock system.

To augment a security system, the contacts of the alarm relay can be interfaced easily with an existing alarm loop, or its contacts can be used as a switch for an alarm horn circuit.

The lock relay may be used to drive a solenoid-driven lock system such as an electric car door lock or it may simply be connected to any solenoid (modified for use as a deadbolt) with a spring return latch assembly. With these approaches, the lock will always be in a locked mode unless the relay is actuated. The lock relay may also be tied to an existing alarm as the arm/disarm switch using the contacts of the relay to toggle a bistable (flop a flip-flop).

A word of caution: you may void any warranty of a professionally installed alarm system by attempting to connect the opticalock to it. Consult your professional installer if you are not sure or have any questions regarding connection of the device to this type of alarm.

Experienced computer enthusiasts may want to consider the following computer/opticalock interface possibilities. The lock driver output and the alarm driver outputs may be used to drive a bistable and this flip-flop can be monitored by a monitoring control system. In addition, the 8-position DIP switch on the main logic board can be driven by a parallel output port on a control system.

As for security, the detector array can be located at any remote location in relation to the rest of the system. If the distance between the detector and logic board exceeds more than a couple of feet (about a metre), additional pull-up and bypass circuitry may be required at each end of the connecting cable.

The detector should be well-armoured against vandals and attempted intrusions. However, its destruction will not give access to any unauthorized persons; bear in mind, though, that access may also be denied then to authorized persons. So you may consider a rather beefy metal plate assembly to guard it.

The lock disable switch should be located inside at a convenient position.

The alarm reset switch should be situated in a semi-secure area that is easy to get to should it have to be used by those who should know its whereabouts.

Should you opt to use a solenoid-driven dead-bolt, or indeed any electric lock, that may be located at a remote position also.

The system can easily be made to work in many different situations. You may wish to experiment with it on some inner-structure door, such as a closet, or your laboratory door, to find out the finer points of a possible application.

The opticalock makes an innovative use of opto-electronics, and with the numerous benefits of this type of lock system, it would not be surprising to see more designs like this very soon. Who knows, perhaps you have just built the key of the future. ■

50 MHz 8-BIT DAC

The digital-to-analogue converter described here is fast, based on discrete components, and comparable — in terms of performance — to pretty expensive integrated circuits.

by R. Shankar

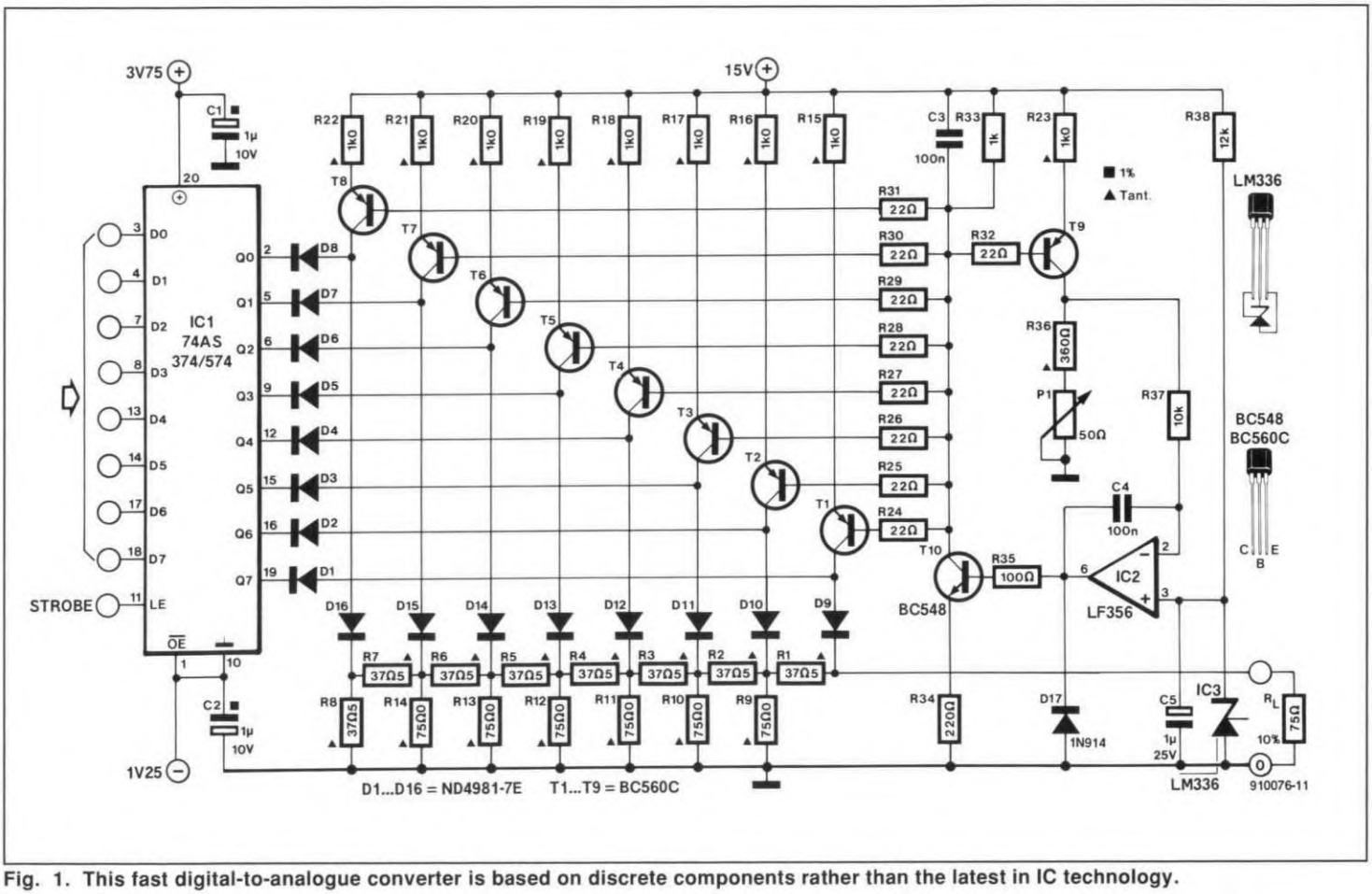


Fig. 1. This fast digital-to-analogue converter is based on discrete components rather than the latest in IC technology.

THE converter can directly drive a 75-Ω load at a full scale voltage of 0.510 V, and a settling time of less than 10 ns to 1 LSB. The temperature coefficient of the output voltage is 0.005%/K, which is significantly better than that of many integrated DACs.

The circuit is based on the conventional *R-2R* ladder network, with current switching accomplished by Schottky diodes and a high-speed Advanced Schottky (ALS) TTL latch. Resistors R1 to R14 form the ladder network, the output impedance of which is 75 Ω. This enables the DAC to drive a load via a 75-Ω coax cable.

Transistors T1 to T8 form the current sources that supply 6.67 mA each to the ladder network via D9 to D16. When any output of IC1 goes low, the resulting current is diverted to IC1 via D1 to D8. This arrangement requires the outputs of IC1 to go negative with respect to ground. Therefore, IC1 is operated with a dual supply of –1.25 V and +3.75 V. To make the inputs TTL compatible, a level translator like the one shown in Fig. 2 may be used at each input. The input loading ('fan-in') of the translator is equal to that of 3 standard TTL inputs. The circuitry around T9, T10, IC2 and IC3 is necessary to obtain a low temperature coefficient.

A suggested power supply circuit is shown in Fig. 3. The current consumption of the DAC is of the order of 170 mA.

Construction hints

Although no PCB design is available for this project, the construction is fairly simple even on a prototyping board. D1 to D16 can be any Schottky diode that meets the following specifications: $I_{Dmax} \geq 10$ mA; $V_{BR} \geq 5$ V; $V_f \leq 0.5$ V at 10 mA.

The connections of the ladder resistors, the Schottky diodes, IC1 and C2 should be kept as short as possible. A large ground

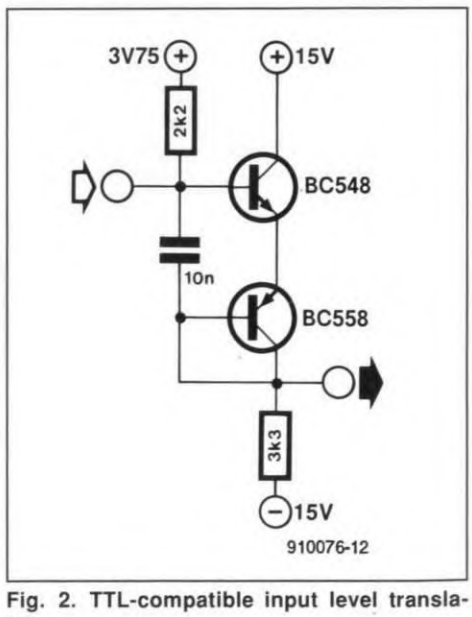


Fig. 2. TTL-compatible input level translator.

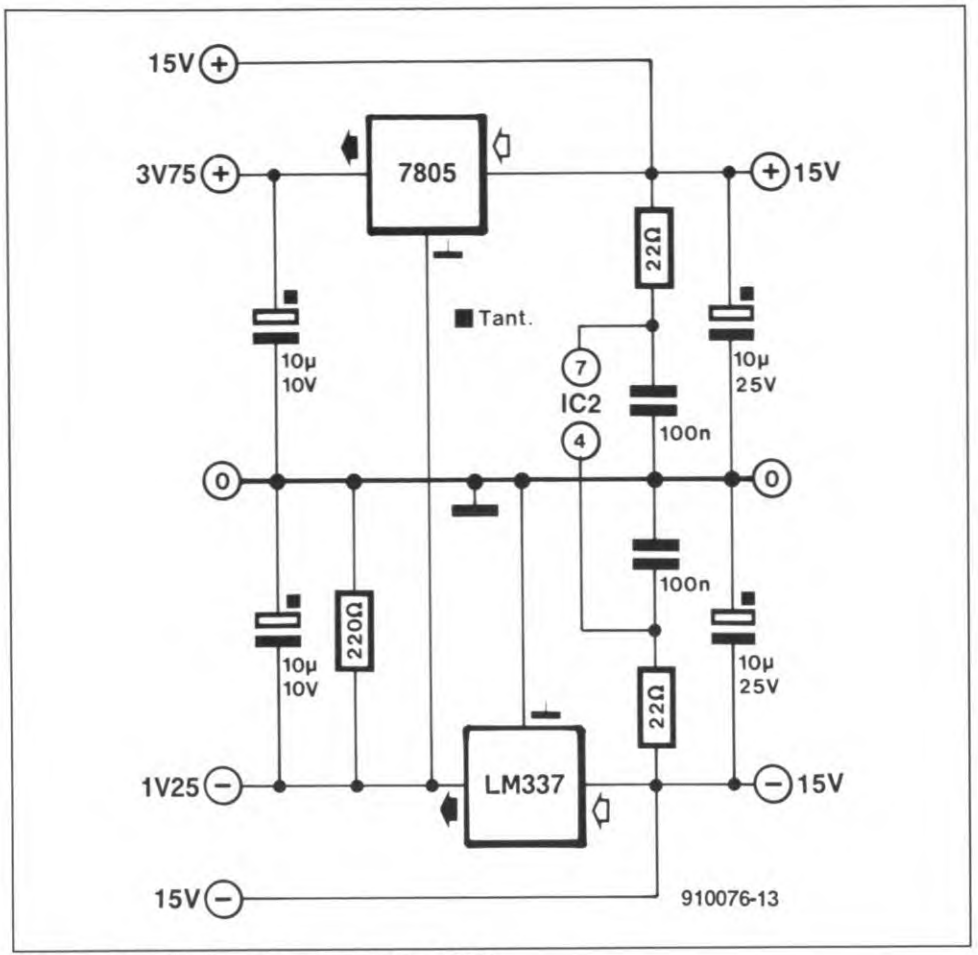


Fig. 3. Regulated power supply for the DAC.

plane should be provided, and the earthed ends of R8-R14, R1 and C2, should meet at one point.

Without component matching, the linearity is of the order of 2 LSB. However, by carefully matching the resistors with the aid of a digital multimeter, a linearity of better than 0.5 LSB can be achieved (tip: use the closely matched resistors towards the MSB). Similarly, the transistors can be matched by measuring the resistance of the base-emitter junction, and noting the values. This measurement should be carried out with the collector and the base connected via the test probe. Additionally, T1 to T9 may be mounted on a common heat-sink for close thermal tracking. ■

COMPUTER-CONTROLLED WEATHER STATION

PART 2: ELECTRONIC HYGROMETER

The electronic hygrometer, or relative humidity (RH) sensor, discussed here is remarkably accurate, and functions automatically as part of the PC-controlled weather station we set out to describe earlier this year.

by J. Ruffell

AMONG the basic measurements carried out daily by amateur meteorologists are those of air temperature, air pressure and relative humidity. In the first instalment of this series, Ref. 1, we described an electronic indoor/outdoor thermometer that works in conjunction with an advanced I/O card for PCs and compatibles (Ref. 2), and dedicated control software. This month we propose to extend the function of this PC-based system with a relative humidity (RH) sensor. Sensors to capture other meteorological data such as air pressure, wind speed and wind direction, will be described in future instalments.

The present circuit measures the relative humidity of ambient air with the aid of a special sensor manufactured by Philips Components. The sensor is basically a non-polarized variable capacitor whose capacitance is a function of the ambient air humidity. Unfortunately, the capacitance does not change linearly with humidity, but that can be corrected fairly easily with the aid of the computer, which is at the heart of our weather station. The computer also greatly facilitates the calibration of the sensor.

Relative humidity

Relative humidity, RH (or *U* in physics), is the ratio of the pressure of ambient air, *e*, to that of saturated air over a plane liquid water surface at the same temperature, *e'*, and is normally expressed as a percentage, that is, $RH = 100e/e'$. The relative humidity depends strongly on air temperature. A relative humidity of 0% indicates a total absence of water particles in the air. In the evening and at night, RH values can rise to 80-90%, while in dense fog the RH is virtually 100%. Typical daytime RH values are between 60% and 70% in sunny weather.

Professional meteorologists measure RH with a wet and dry bulb hygrometer. This consists of two mercury-in-glass thermometers mounted side by side. The dry bulb thermometer shows the air temperature in the usual way, while the other has its bulb kept moist continually (by means of a small reservoir mounted underneath it) and exposed to an adequate draught. The rate at which the air takes moisture from the reservoir depends on how far it is being saturated with water vapour. If the air is already saturated, it takes up no more and the two thermometers read the same (that is, RH=100%). The lower the humidity of the ambient air, the more it tends to take up moisture, and the lower the reading of the wet bulb thermometer compared with that of the dry one. The difference in temperature gives a highly reliable measure of relative humidity, which is read from tables.

Another, less accurate, method to measure the RH makes us of a hair. This based on the fact that the length of a strained hair depends on the humidity of the ambient air. This inexpensive way of measuring the RH is applied mainly in simple hygrometers of the type hung up in the living room.

RH measurement: the electronic way

Unlike most electronic hygrometers, which are based on a hygristor (a component whose electrical *resistance* varies with humidity), the one described here is based on a Philips Type H1 sensor whose *capacitance* changes with humidity. The capacitance of this device varies between about 110 pF and 145 pF for RH values of between 10% and 90%. The RH-dependent capacitance is used to vary the frequency of an oscillator based on the well-known 555 timer IC — see the circuit diagram in Fig. 1. The RH sensor and the external components around the 555 give the oscillator a nominal frequency range from 43.6 kHz to 33.1 kHz. Since the capacitance of the sensor has a maximum tolerance of 15%, the lower frequency may drop to 28 kHz, while the upper frequency may rise to about 52 kHz. Capacitance-to-frequency (*C*-to-*F*) conversion is applied here to allow the measured RH value to be carried over a considerable distance via a cable.

The software that runs on the PC takes care of the linearization of the sensor characteristic. The method developed for this purpose is remarkable because it reduces the number of adjustments in the hygrometer to one! By making use of the 6-digit frequency meter function provided by the measurement card for PCs, the electronic hygrometer achieves a resolution of 1%.

Two H1 sensors and associated interfaces are used since it is usually required to measure the indoor and outdoor RH values simultaneously. The two frequencies produced by the sensor interfaces are applied to the F6 (RH indoor) and F7 (RH outdoor) inputs of the PC measurement card. These two connections are located on connector K6.

MAIN SPECIFICATIONS

Electronic hygrometer

Range:	0% to 95%
Channels:	2 (indoor and outdoor)
Recording:	continuous (interval = 10 minutes)
Measurement:	once every minute
Measured values:	current, day maximum, day minimum
Indication:	percentage RH; low, normal, high
Software:	memory-resident data logger, calibration utilities and full-colour graph display programs

System requirements

IBM PC XT, AT, 386, or compatible
640 KByte RAM
EGA or VGA display adapter
DOS 3.3C or later

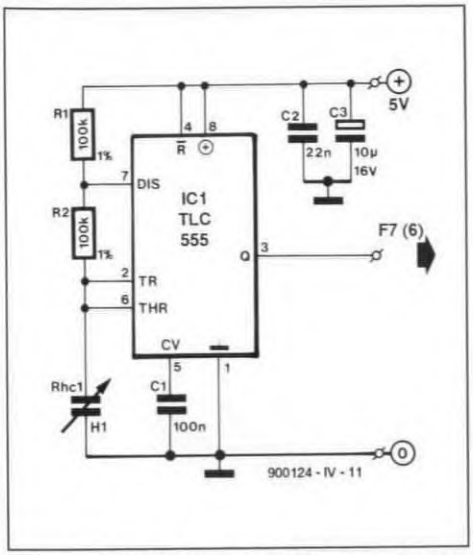


Fig. 1. Circuit diagram of the capacitance-to-frequency converter.

The two sensors may be connected to the computer via ordinary wire, e.g., a 4-wire telephone cable. Cable lengths up to 15 m (50 ft) should not present problems. Each interface requires a supply voltage of 5 V, which is conveniently taken from the stabilized +5-V supply rail brought out on connector K6 of the PC measurement card. Decoupling capacitors C1 and C2, and reservoir capacitor C3 compensate the impedance of the (long) supply wires, which would otherwise degrade the stability of the *C*-to-*F* converter. With C1, C2 and C3 fitted, the measured value, at RH=50%, varies no more than 0.004% per mV of variation in the supply voltage.

In some cases, it is not desirable that the sensor interfaces are powered by the PC. The alternative is to build a small, regulated 5-V supply based on, for example, an 7805, and fit this in the sensor enclosure. The stability of the *C*-to-*F* converter is best when the voltage regulator is located as close as possible to the sensor interface board. The enclosure used to house the sensor should offer plenty of space for the regulator and the usual decoupling capacitors at its input and output.

Construction

Figure 2 shows the printed circuit board on which the sensor is mounted. The construction of the interface is downright simple, and merits no further discussion. The PCB fits exactly in a Type E406 enclosure from Bopla. Other enclosures may be used, but may require some drilling and filing to fit the PCB.

Drill two 1.5-mm holes in the cover of the enclosure to allow the connecting terminals of the H1 sensor to pass. Extend the H1 terminals with 5-cm long wires before securing the device to the cover with a few drops of two-component glue. The two wires are soldered to the PCB holes marked Rhc1. Drill a hole in the side panel of the case to allow the cable to the PC to pass. This hole should be just large enough for the cable to pass. Fit a strain relief on the cable at the inside of the enclosure.

The interface must be protected against moisture. This is best achieved by applying a little silicone compound on the edges of the enclosure before this is closed. Also use the compound to seal the holes for the sensor wires and the cable to the PC. In both cases, apply the compound at the inside of the enclosure. Before the enclosure is closed, the holes in the bottom half must be sealed with compound. Like the thermometer module, the sensor is temporarily connected to the PC via a small piece of stripboard.

Software in action

The floppy disk supplied for this project, ESS 1561, contains the software required for the adjustment of the sensor, the measurement routines, and the automatic logging routines. In addition, it contains the first update (version 1.1) of the control software for the thermometer module. The new software is marked by an extended version of the memory-resident data logger. It integrates temperature and RH measurements, and provides links for the wind speed and wind direction modules sensors to be discussed in future instalments of this article.

The name of the background program has been changed from TLOGGER into XLOGGER. The update takes into account that XLOGGER is the only program that addresses the PC measurement card — i.e., XLOGGER now performs measurements continuously, and adds measured values to the log file every 10 minutes. The programs for the graphics display of temperature and relative humidity, TEMP and HUM respectively, translate the data contained in the log file into graphs on the PC screen, and at the same time open a RAM data path to XLOGGER. This path allows them to receive the current measurement results, i.e., maximum, minimum and current temperature, and maximum, minimum and current relative humidity. This data transfer avoids complex synchronization protocols between XLOGGER and other programs, and prevents the hardware on the PC measurement card being addressed simultaneously by two synchronous routines, which would cause the PC to crash.

The communication between XLOGGER and the calibration utilities, HADJUST and TADJUST, is similar to that described above. XLOGGER performs all its functions in the background, i.e., as a memory-resident utility. This allows you to use the PC as before for word processing or drawing. The relative humidity is measured once every minute, and the temperature once every 15 minutes. XLOGGER can be actuated by typing XLOGGER /I from the command level. The program takes about 50 KBytes of the base memory in the PC.

.LOG and .CFG files

Users of TLOGGER should note that log files produced by this program (Tyymmdd.LOG) are not compatible with the new log files produced by XLOGGER (Xyymmdd.LOG). The old files may still be viewed with version 1.0 of TEMP, however, provided XLOGGER is removed from memory. This is achieved by typing XLOGGER /U.

The new Xyymmdd.LOG files contain

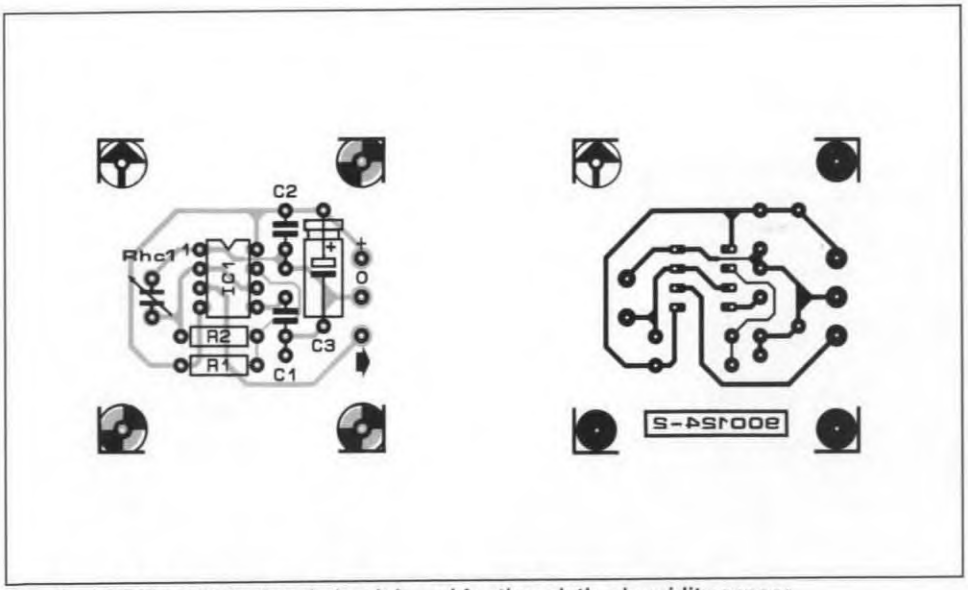


Fig. 2. Single-sided printed-circuit board for the relative humidity sensor.

COMPONENTS LIST

Resistors:

2	100kΩ 1%	R1;R2

Capacitors:

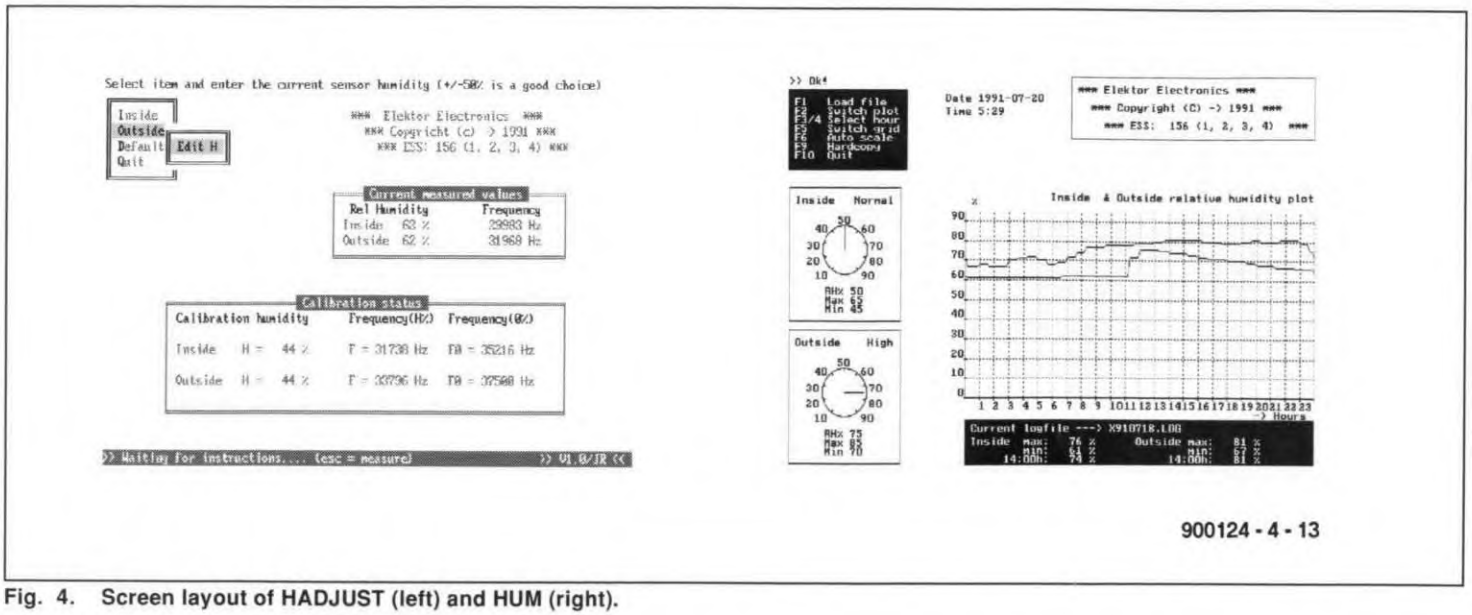
1	100nF	C1
1	22nF ceramic	C2
1	10µF 16V	C3

Semiconductors:

1	TLC555	IC1

Miscellaneous:

1	H1 humidity sensor (Philips Components 2322 691 90001)	Rhc1
1	ABS enclosure, size approx. 65×50×30 mm; e.g., Bopla EG406	
1	Printed circuit board	900124-2
1	Control software on disk	ESS 1561



900124 - 4 - 13

Fig. 4. Screen layout of HADJUST (left) and HUM (right).

voltages, frequencies and binary codes only. The calibration utilities create transfer files, HTRANS.CFG and TTRANS.CFG, that allow these quantities to be converted into relative humidity and temperature, respectively. TTRANS.CFG contains two linear functions, while HTRANS.CFG is used to store the arithmetically determined frequencies that correspond to a relative humidity of 0%. The status windows of TADJUST and HADJUST show the content of these transfer files. It should be noted that the transfer files must be saved before the sensors are recalibrated, which may be necessary after some time. Without the saved transfer files, it is impossible to convert the existing log files into the correct values.

The RAM channel conveys frequencies, voltages and codes only. The calibration points are recorded in the TCALI and HCALI configuration files.

The main program, XLOGGER.EXE, uses a configuration file, XLOGGER.CFG, from which it reads a number of initial parameters. XLOGGER.CFG is a text file which may be opened and modified with any ASCII-compatible word processor. The first text line defines the path and the directory where the log files are stored. The path is also used by TEMP and HUM, and may have to be modified depending on your requirements. The next four lines in XLOGGER.CFG are the labels, or names, assigned to the sensors. These labels are used by TADJUST, HADJUST, TEMP and HUM, and may be changed if necessary.

One line of text has been added to ADCF.CFG — this will be reverted to below.

XLOGGER — background to a complex and powerful program

A/D conversion

XLOGGER controls the A-D converter on board the PC measurement card in an interesting way. Each measurement result is based on an average of 500 measurements. The start-of-conversion command for the ADC is generated by the system clock of the PC. Unfortunately, the USER-$1C-TIMER-TICK interrupt can not be used without modifications because it would result in excessive wait times for 500 or so conversions. This interrupt is executed at a rate of only 18.2 Hz, which is so slow that it would cause the total conversion time to rise to 27.5 seconds. XLOGGER reduces the conversion time to 5 seconds by reprogramming the prescaler in the system clock. When XLOGGER is installed, the standard interrupt BIOS routine $08, which controls, among others, the real-time clock, and is started at the end of the $1C handler, is replaced by a new routine that takes care of the data acquisition operations. This interrupt-$08 routine calls the standard BIOS routine (and with it interrupt-$1C) every 55 ms on average, so that the real-time clock continues to function normally.

Pop-up handler and frequency meter

XLOGGER uses interrupt-$1C to pulse the PB0 line when an error condition occurs. Interrupt-$1C also starts a so-called pop-up handler every 15 seconds. The pop-up routine reads the results of the interrupt-controlled frequency measurement, starts the next measurement with a new channel number, and adds the captured values to the log file every ten minutes. The main advantage of the pop-up handler is that DOS operations such as disk I/O may continue to run asynchronously. By virtue of the pop-up handler, the fact that DOS is not 're-entrant' is no stumbling block any more in the development of interrupt routines and memory resident programs. When the pop-up handler is run, the PB0 line is briefly actuated.

Install/Uninstall

When XLOGGER is started, the presence of a valid parameter, the /I (install) or /U (uninstall) switch, is checked. The program terminates with an error message when the parameter is missing or incorrect. When the /I option is used, XLOGGER is installed only when it is not already present in the memory. The uninstall option, /U, works only if no other memory-resident programs were loaded after XLOGGER. Note that DOS PRINT is a memory resident program!

Interrupts

XLOGGER can not be used without the hardware on the PC measurement card. Since the hygrometer makes use of the frequency meter on the measurement card, one of the jumpers JP2-JP7 must be fitted to enable the interrupts to be handled correctly. First, determine which interrupt line is still free in your PC (in most cases, this is IRQ2). Fit the respective jumper on the X-row. Next, check that jumper JP8 is fitted in position E. The interrupt line used by XLOGGER is found in the last line of the configuration file ADCF.CFG, and must be the same as the line selected by the jumper. Make sure that this is the case by opening ADCFG.CFG with your word processor, or typing TYPE ADCF.CFG, and looking at the interrupt number. If necessary, change the number to match the hardware interrupt selection.

Software: features and applications

HADJUST

This is the calibration program for the hygrometer function. The layout of the screen, and the use of pull-down menus are similar to those of TADJUST, which may already be familiar from the earlier article on the thermometer module. Adjusting the RH sensor is much less complicated than the temperature sensor because it involves only one setting for each sensor. This is achieved by virtue of purposely written calibration routines contained in the HADJUST program. The principle of linearization is explained separately elsewhere in this article. Before the calibration is started, the two sensors are placed in an environment of which the relative humidity is known. It is best to use a

calibrated hygrometer for this purpose, and you may need to go to your local weather centre to ask for assistance.

Install XLOGGGER, and run the HADJUST utility. The menu appears, and two windows, one containing the calibration status, the other the currently measured values. When the program is run for the first time, the calibration window shows a default status, which is simply a rough estimate of the expected calibration data. The other window displays the frequencies produced by the two sensors. From the installation of XLOGGER, it takes about one minute before the two frequencies are measured for the first time. Until then, the frequency variables remain at nought. The Currently Measured Values (CMV) window also shows the relative humidity values derived from the frequencies and the calibration status.

When the first frequency measurement is complete, select the 'inside' option from the menu, after which the 'EDIT H' window is opened. Press the return key, and enter the actual humidity (the value read from the reference hygrometer) at the cursor location. You will notice that the calibration status and the calculated relative humidity are automatically updated. The other sensor (option 'outside') is calibrated likewise.

Leave the program via the 'quit' option, and subsequently select 'update' to save the transfer and calibration files. Do not select 'abandon' at this stage, as this will terminate the program without storing the previously obtained calibration data.

HUM

HUM (for humidity) handles the graphics presentation of currently measured as well as previously saved values. The screen layout of HUM is basically the same as that of TEMP. HUM displays two analogue hygrometers with an RH percentage scale in the left-hand bottom corner of the display. It should be noted that RH measurements are carried out only when XLOGGER has been installed. The current RH value is also shown numerically on the screen, together with the minimum and maximum values measured during the last 24 hours. The analogue RH meters have scale areas marked 'low' (for RH values smaller than 45%), 'normal' (for RH values between 45% and 65%), and 'high' (for RH values greater than 65%).

The control menu appears in the top left-hand corner of the screen, while the auto-ranging 24-hour graphs take up most of the right-hand side of the screen. Below the graphs you find the name of the log file and the minimum and maximum values in this.

Finally, function keys F3 and F4 allow you to request the RH value logged at every full hour. The other function keys serve to select certain program options: a graphics or non-graphics graticule; inside RH or outside RH graph or both; producing hardcopy (Epson FX-85 mode only); loading another log file; and leaving the program. ❑

References:

1. "Computer-controlled weather station — Part 1: indoor/outdoor thermometer", *Elektor Electronics* March 1991.

2. "Multifunction measurement card for PCs", *Elektor Electronics* January and February 1991.

HUMIDITY SENSOR CALIBRATION — WITH A SINGLE ADJUSTMENT

The Type H1 humidity sensor from Philips Components is basically a non-polarized capacitor whose capacitance depends on the humidity of the air around it. The dielectric consists of a non-conductive foil with a thin gold layer at both sides. The two gold layers form the electrodes in this sensor. The dielectric constant, ε, of the foil is a function of the atmospheric humidity. The absolute capacitance of the capacitor is calculated from Eq. (1). In the equation, r is the relative humidity, A the surface of the electrodes, and d the thickness of the foil. Unfortunately, the characteristic of the sensor is non-linear (see the curves below). This means that converting a measured capacitance (or a frequency which is inversely related to the capacitance) into a relative humidity value is not so simple. A further complicating factor is that the real characteristics of the two sensors you have lie somewhere between the C_{max} and C_{min} curves. This spread is caused by production tolerances on the thickness of the dielectric. The problem with the spread can be solved by basing all calculations on the capacitance factor, $K_{(r)}$, calculated from Eq. (2), instead of on the absolute capacitance. Substituting Eq. (1) in Eq. (2) proves that $K_{(r)}$ is independent of the thickness of the dielectric. Eq. (3) shows that the dielectric constant, $\varepsilon_{(r)}$, is the single factor that determines the actual capacitance. This means that $K_{(r)}$ is independent of the spread on the sensors, so that it can be used as a repeatable starting point. Table 1 shows a few discrete values of $K_{(r)}$. Remember, $K_{(r)}$ is expressed as a percentage.

In effect, the oscillator frequency is measured instead of the sensor capacitance. The frequency is an inverse function of the capacitance. In Eq. (4), parameters a and b are oscillator-dependent constants which have no further significance here. Combining Eq. (1) with Eq. (4) yields Eq. (5). Finally, substitution of Eq. (5) in Eq. (3) yields the output frequency formula, Eq. (7).

Calculations

Assuming that $f_{(0)}$ is known, Eq. (6) enables us to calculate the capacitance factor on the basis of a measured frequency $f_{(r)}$. The relative humidity (RH) associated with the value of $K_{(r)}$ may then be deduced, by interpolation, from Table 1.

The frequency at 0% RH, $f_{(0)}$, is also determined arithmetically. This is done by performing a frequency measurement at one, known, RH value. The capacitance factor follows from Table 1 via inverse interpolation. Next, the measured frequency and the result of the interpolation are entered into Eq. (7). The result is a frequency, $f_{(0)}$, expressed in Hertz.

$$C_{(r)} = \frac{\varepsilon_{(r)} \times A}{d} \quad [F] \quad \text{Eq. (1).}$$

$$K_{(r)} = \frac{C_{(r)} - C_{(0)}}{C_{(r)}} \times 100 \quad [\%] \quad \text{Eq. (2).}$$

Substituting Eq. (1) in Eq. (2) yields:

$$K_{(r)} = \frac{\varepsilon_{(r)} - \varepsilon_{(0)}}{\varepsilon_{(r)}} \times 100 \quad [\%] \quad \text{Eq. (3).}$$

The oscillator frequency is a function of the humidity as expressed by

$$f_{(r)} = \frac{a}{b \times C_{(r)}} \quad [Hz] \quad \text{Eq. (4).}$$

Substituting Eq. (1) in Eq. (4) yields:

$$\varepsilon_{(r)} = \frac{a \times d}{A \times b \times f_{(r)}} \quad [F/m] \quad \text{Eq. (5).}$$

Substituting Eq. (5) in Eq. (3) yields:

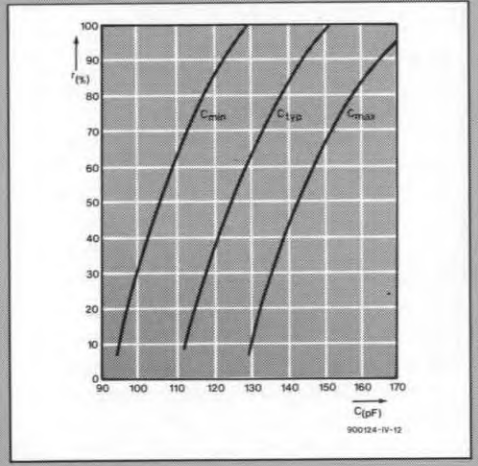
$$K_{(r)} = \left(1 - \frac{f_{(r)}}{f_{(0)}}\right) \times 100 \quad [\%] \quad \text{Eq. (6).}$$

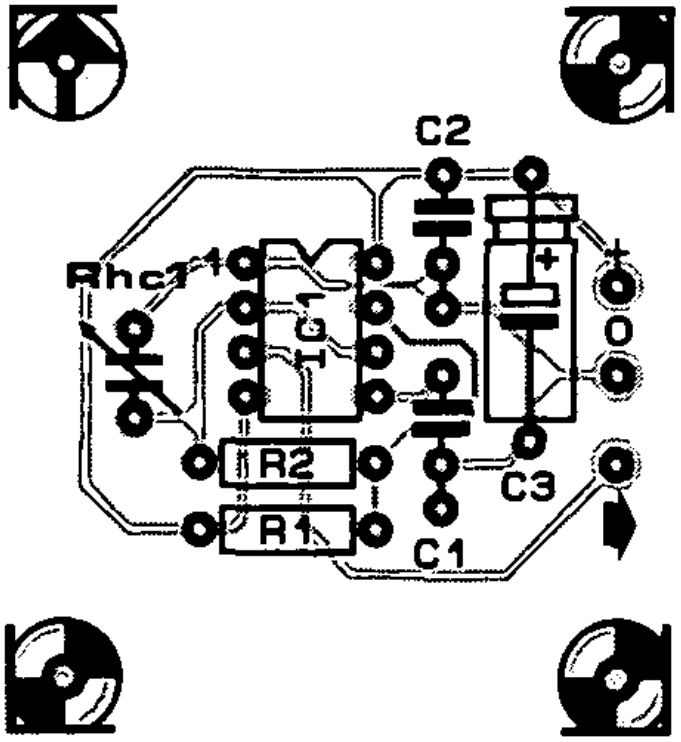
After calibration, Eq. (6) can re rewritten as

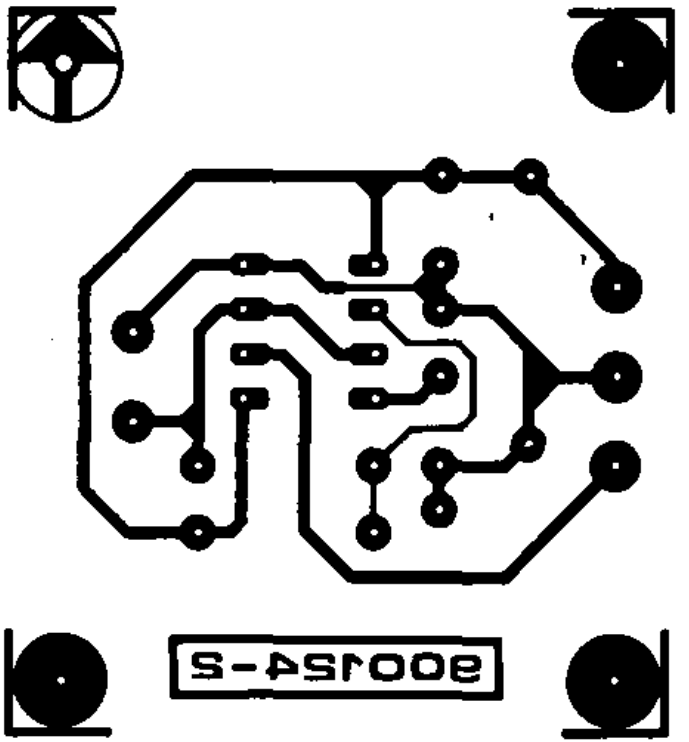
$$f_{(0)} = \frac{f_{(r)}}{1 - K_{(r)} / 100} \quad [Hz] \quad \text{Eq. (7).}$$

Table 1. RH - K correlation

r (%)	K (%)
10	2.135
20	4.265
30	6.463
40	8.865
50	11.433
60	14.130
70	16.918
80	19.767
90	23.238




C2
Rhc1
IC1
R2
R1
C3
C1


90024-S

MEASUREMENT TECHNIQUES – PART 8

Measurements in digital circuits

by F.P. Zantis

MEASUREMENTS in digital circuits fall into two categories: (1) voltage, current and resistance; and (2) data. The first category presents no problems other than those already discussed in previous instalments of this article. The measurement of data and associated circuits will be described in this part in the series.

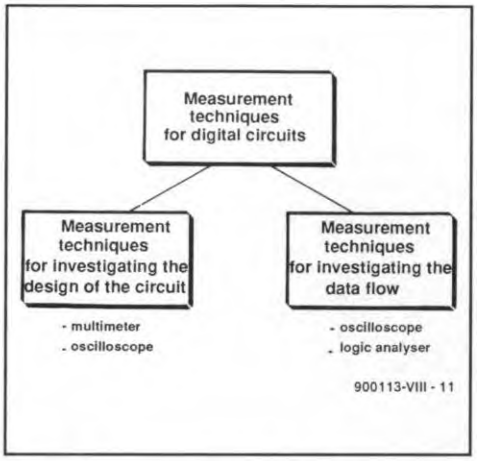


Fig. 1. Measurements in digital circuits fall into two categories.

Techniques for examining circuit design

Typical of digital circuit designs is the use of densely populated printed-circuit boards, whose tracks, owing to space limitation, are extremely thin. This makes faultfinding particularly difficult and it is, therefore, advisable to carry out a thorough inspection of the board before any work on it is begun. For this, an ohmmeter, multimeter, oscilloscope with component tester or a simple continuity tester is perfectly adequate.

Once the board has been populated, a multimeter is useful only for checking the various supply voltages. A component tester is also very useful: this quickly shows whether a component is all right or defect. Voltage levels and edges of digital signals can be checked with an oscilloscope.

Data pulses can cause problems with the triggering of an oscilloscope, however. Some oscilloscopes are, therefore, provided with special facilities, such as pre-trigger or hold-off. The pre-trigger function allows it to examine the trigger pulse just prior to its first mesial (see p. 41). In other words, the signal behaviour before triggering is displayed. The trigger pulse is then not shown at the left-hand of the screen, but in the centre, or even at the right-hand.

The hold-off function allows the continuous lengthening of the delay between two time-base sweep voltages as shown in Fig. 2. Pulses or other spurious signals occurring during the hold-off time cannot influence the triggering, which is particularly useful in the case of aperiodic pulse bursts of constant amplitude. The oscilloscope can be triggered on different points of the waveform.

A digital storage oscilloscope is particularly useful for measuring an aperiodic signal. Note, however, that in moderately priced models short spurious pulses or glitches may not all be identified The scanning rate and storage capacity limit the use of these oscilloscopes.

Examining the data flow

As is well-known, logic signals in digital engineering can only be 1 (high) or 0 (low). Errors caused by insufficient pulse voltage levels or resulting from too long a first transition duration of the pulse can be traced with standard test methods.

A typical problem encountered in measurements in digital circuits is that the examination of only one signal usually does not yield enough information about the functioning of the circuit. Generally, two or, in complex circuits, even more signals must be examined simultaneously.

For analysis of two signals, a dual-beam oscilloscope is required (not a dual-trace type, because that cannot capture two fast transient events). Examination of the data bus of an 8-bit microprocessor system requires an 8-channel oscilloscope. Even that is not sufficient if a control signal such as 'write' or 'read' must be triggered.

Some, rather expensive, oscilloscopes permit the simultaneous display of four signals—see Fig. 3. A dual-beam oscilloscope may be converted with a multi-channel adaptor to make relatively simple, qualitative test on a number of channels possible. The major disadvantage of such adaptors is their low chopping rate, which allows examination of relatively slow signals only.

In contrast to oscilloscopes, which measure voltage as a function of time, logic analysers are designed primarily for examining the data flow in complex digital circuits and microprocessor systems.

A logic analyser is suitable for testing hard-

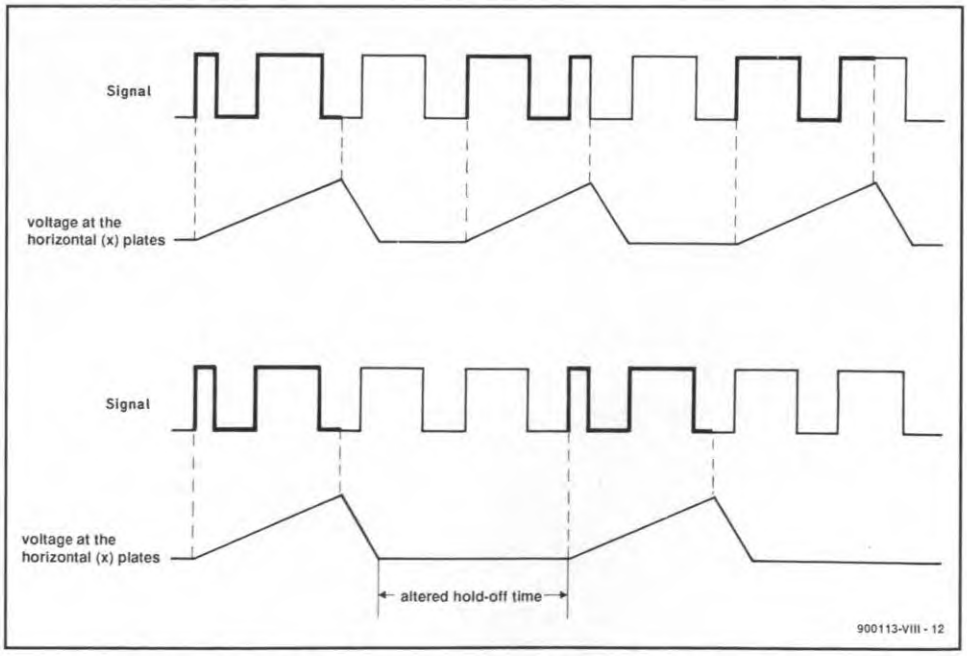


Fig. 2. Illustrating the hold-off function. At the top, this facility is not operating. The oscilloscope is then triggered when the electron beam is in its output position and a leading edge appears in the signal. The waveform in bold print is then shown on the screen; the picture runs across the screen. Analysis is not really possible. In the lower half, the hold-off time is altered in a manner that ensures that triggering always occurs at identical leading edges, so that the picture stands still. It is then possible to identify any glitches or other short, spurious pulses clearly.

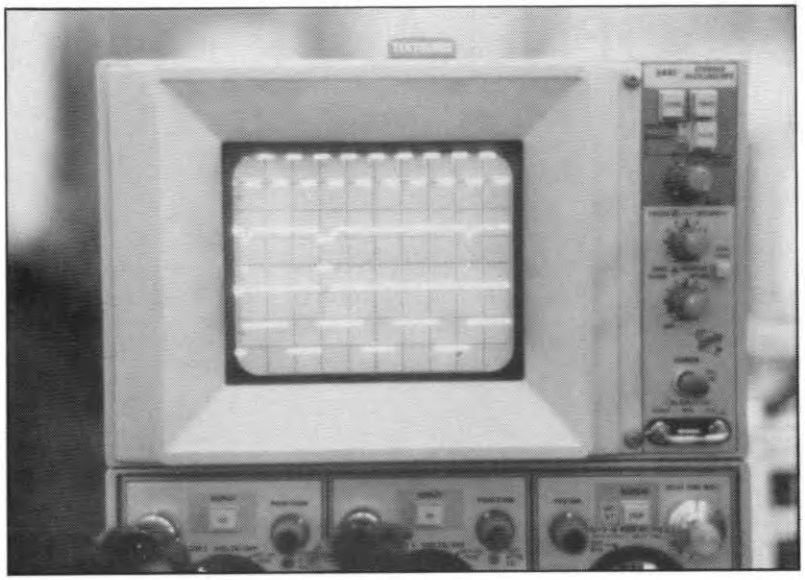


Fig. 3. Display of four logic signals on a 4-channel oscilloscope.

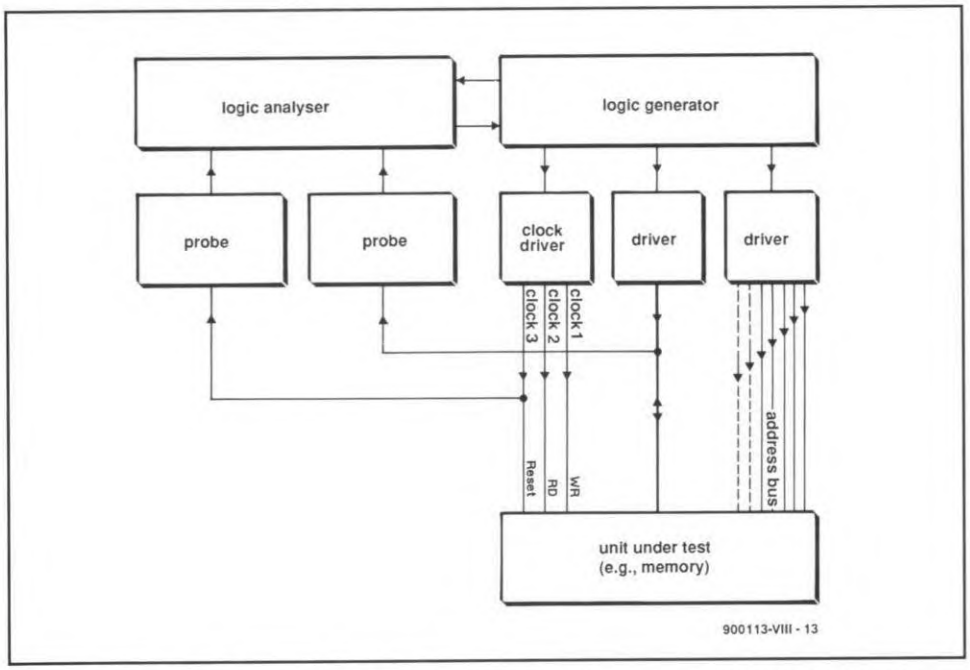


Fig. 4. Typical application of a combination of a logic analyser and a logic generator.

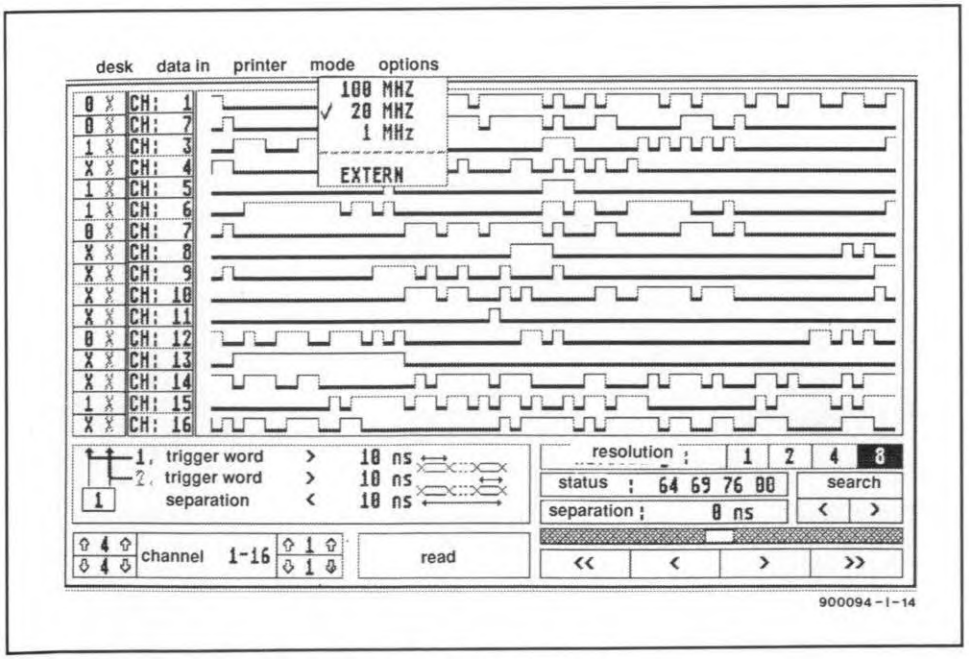


Fig. 5. Timing diagram displayed on a logic analyser.

ware as well as software; it measures data, that is, a pulse is considered as a bit with a value of 1 or 0, and not as a rectangular waveform with a certain amplitude. At the time of writing, logic analysers with up to 72 input channels are commercially available.

A combination of a logic analyser and a logic generator, which provides sample bits that can be fed to the circuit under test, is especially useful. It allows the analysis of circuit sections even when these are driven by external sources.

A typical application of such a combination is shown in Fig. 4, in which the unit on test is controlled via a data bus and an address bus. The logic generator provides the addresses, data input and clock, while the analyser records the test data.

The most important part of a logic analyser is the control computer, whose software serves to operate the analyser and to analyse and display the test results. In that process, it is necessary for the relevant information to be extracted from a vast mass of data. In the simplest case, it is only necessary for a given sample bit to be triggered. More complex systems permit triggering even when, for example, the operating conditions are normally set by one or a number of data buses.

Differences between various logic analysers, reflected in their price, are manifested particularly in the possibilities of data extraction and the way this is carried out.

The data display on the screen may take various forms. Apart from the usual timing diagram as shown in Fig. 5, which shows the behaviour of signals as a function of time, it is also possible for the status or the disassembler to be shown.

The status diagram shows the input channels in a circuit section that were determined previously by a configuration menu.

The disassembler diagram is particularly useful for an analysis of the interaction between hardware and software. The recorded data show the converted machine code of the computer system on test in mnemonic form. This process is called disassembly.

Every logic analyser is provided with special probes that facilitate proper connection of the many inputs to the unit under test. It is, of course, essential that fast pulse bursts be transmitted only via short lines that introduce no or negligible losses.

The probes contain drivers that amplify the signal prior to transmission to the appropriate analyser input. ■

Reference:

'Logic analyser' (5 parts), *Elektor Electronics*, January, February, April, June, July 1991.

DIGITAL FUNCTION GENERATOR – PART 1

by T. Giffard

The outstanding properties of this function generator are its accurate frequency setting and very low distortion of the sine-wave output. It is, therefore, eminently suitable for use as an a.f. generator for test and design.

FRONT COVER PROJECT

Fig. 1. General view of the digital function generator.

UNLIKE many function generators produced over the past 15 years or so, the one described here is NOT based on the ubiquitous XR2206, but on discrete components. Although this does not necessarily make it easier to construct, digital ICs are fairly inexpensive and a digital circuit has fewer calibration points.

Basic design

The block diagram in Fig. 1. shows that the design is based on four printed-circuit boards, indicated by the dashed lines. The heart of the frequency-synthesis board at the top left is the phase-locked loop (PLL). A phase comparator compares the output frequency of a crystal oscillator with that of a voltage-controlled oscillator (VCO). Any phase difference between the two signals causes the comparator to generate a voltage that is used to synchronize the VCO with the reference oscillator.

The wanted output frequency of the generator is determined by the divisor of a divider in the PLL, which is set by means of a keyboard on the front panel. The divider also provides the information for the four-digit LED display.

The VCO and the display share a common power supply, although the supply voltage to the VCO is 6.6 V, whereas that to the remainder of the circuit is 6.0 V.

The divider provides a rectangular signal, which is directly proportional to the output frequency of the generator, to a digital-to-analogue (D-A) converter. This converter consists of a shift register and an appropriate resistance network.

Each period of the wanted sine-wave signal is built up of 32 clock pulses provided by the divider. This means that the clock signal for the D-A conversion must lie between 32 Hz and 3.2 MHz if the generator is to provide an output frequency of 1 Hz to 100 kHz.

Since, in addition to the high precision of the wave shaping, the D-A converter is followed by a phase-locked filter, the distortion factor is even smaller than that of a Wien-bridge type sine-wave generator commonly used for a.f. design and test measurements. Usually, function generators produce a sine wave that is converted into rectangular and triangular signals of the same frequency: this technique does not result in low distortion factors.

Technical data

	Sine wave	Rectangular	Triangular
Frequency	1 Hz–99.99 kHz	1 Hz–99.99 kHz	10 Hz–99.99 kHz
Output voltage	1 V_{rms}, e.m.f.	0–12 V e.m.f.	0–12 V e.m.f.
	0.9–1 V_{rms} into 50Ω	0–6 V_{pp} in to 50Ω	0–6 V_{pp} into 50Ω
Offset	±2.5 V (U_o=0 V)	±12 V	±12 V
	±4.75 V (U_o=1 V_{rms})		
Output impedance	600Ω	50Ω	50Ω
Frequency stability	Three digits better than shown on display		
Frequency resolution	0.01% of full-scale reading		
Third-harmonic distortion	0.03% typical		
Slew rate	130 V/µs		

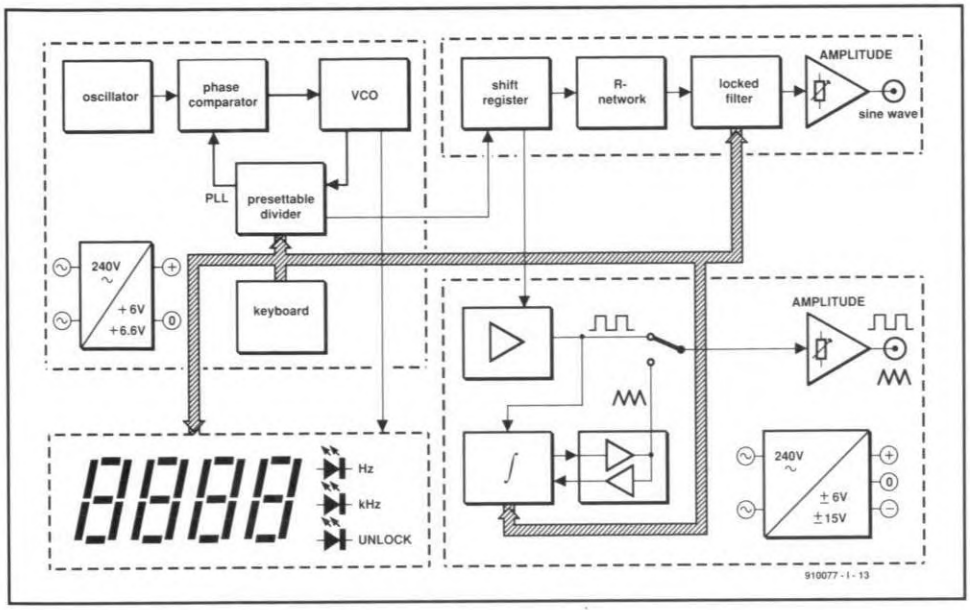


Fig. 2. Block diagram of the digital function generator.

After the amplitude has been set, the sine-wave signal is fed to the output socket via a buffer amplifier.

For the other two wave forms, the shift register in the sine-wave section is used as a divider to ensure that the frequency remains the same as that of the sine-wave signal. The signal is fed to the rectangular/triangular output socket via a voltage amplifier, change-over selector and power amplifier. At the same time, it is used to drive an operational transconductance amplifier (OTA) which converts the rectangular signal into a triangular one.

Divisors

An expanded, more detailed block diagram of the PLL frequency synthesis section is given in Fig. 5. Apart from the crystal oscillator, the VCO and the presettable divider, the section contains several more dividers and

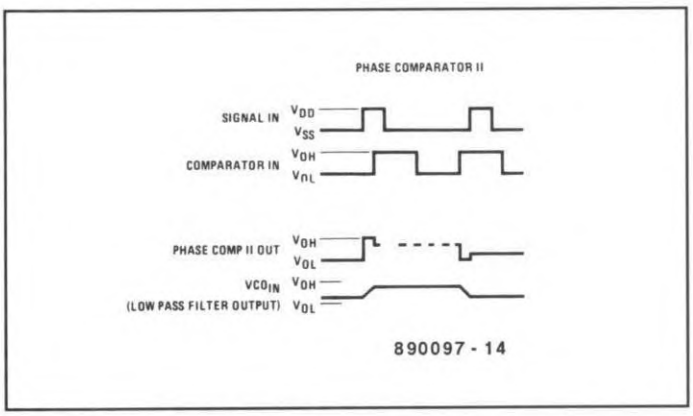


Fig. 3. Mode of operation of phase comparator IC_4.

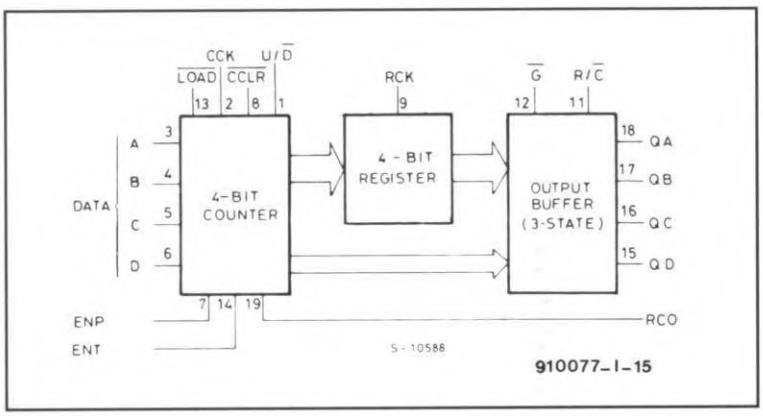


Fig. 4. Internal circuit diagram of the Type 74HC696 counter.

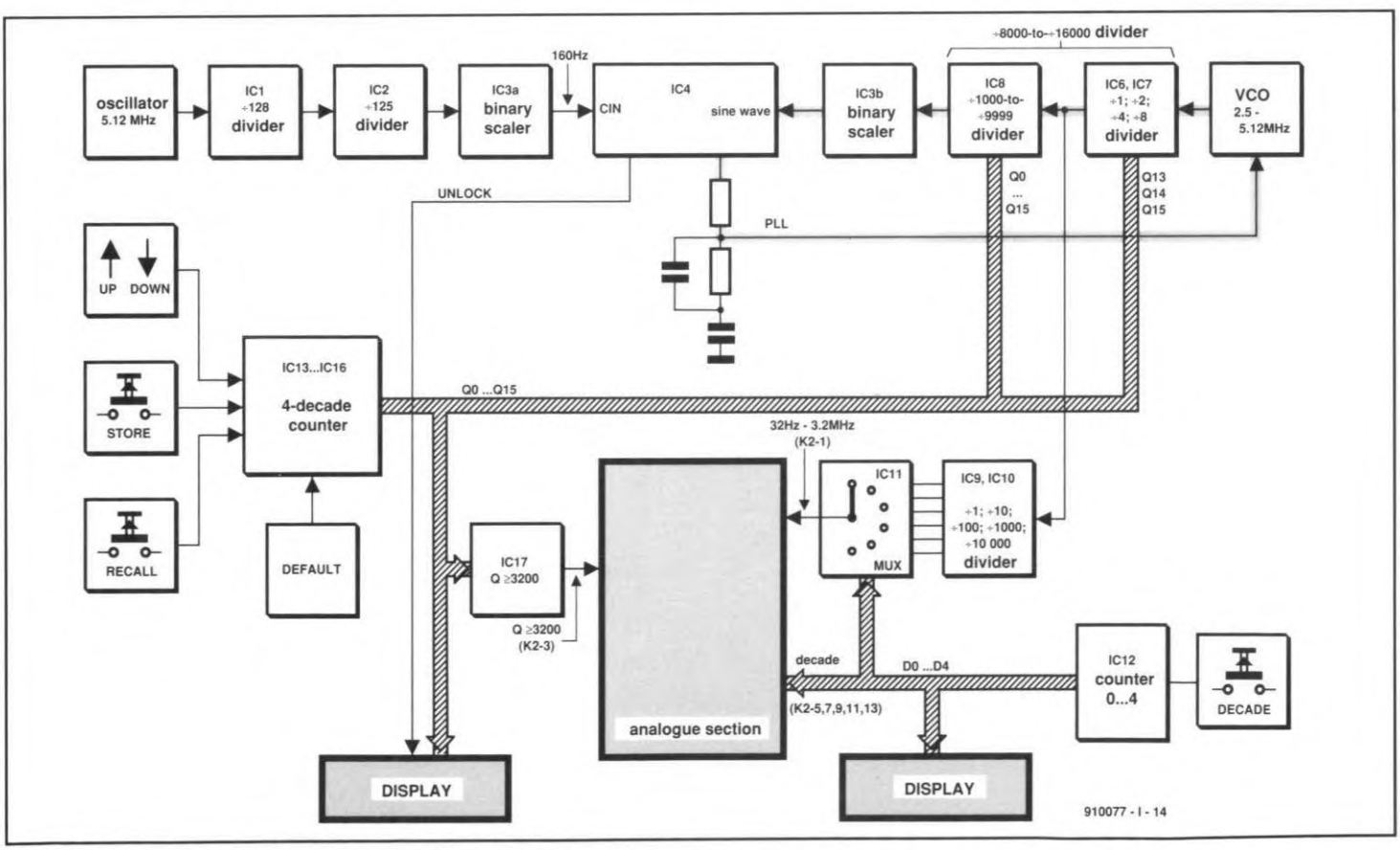


Fig. 5. Block diagram of the phase-locked loop (PLL) section.

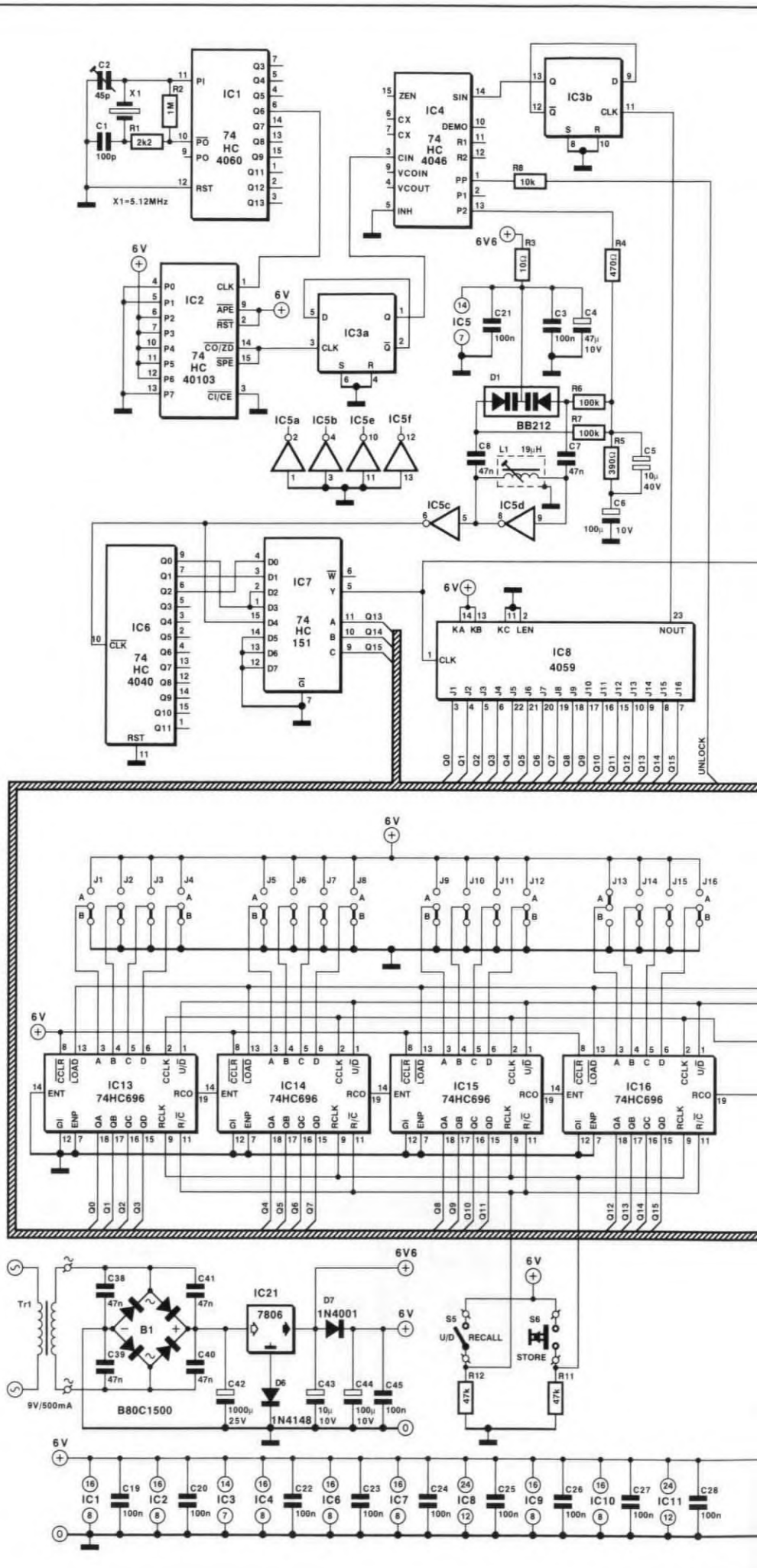
PARTS LIST

Resistors:
R1 = 2.2 kΩ
R2,R14,R15,R17,R19 =1 MΩ
R3 = 10 Ω
R4 = 470 Ω
R5 = 390 Ω
R6,R7,R10,R20,R22 = 100 kΩ
R8,R9,R13,R18,R21,R23 = 10 kΩ
R11,R12,R16,R61–R65,R67, R68 = 47 kΩ
R30–R60 = 470 Ω
R66,R69 = 1 kΩ

Capacitors:
C1 = 100 pF
C2 = 45 pF trimmer
C3,C51–C53 = 100 nF ceramic
C4 = 47 µF, 10 V, tantalum
C5 = 10 µF, 40 V, bipolar, radial
C6,C44 = 100 µF, 10 V, radial
C7,C8 = 47 nF
C9,C10,C12,C14,C15,C17, C18 = 100 nF
C11 = 470 nF
C13,C16 = 3.3 nF
C19–C37,C45 = 100 nF, ceramic
C38–C41,C50 = 47 nF, ceramic
C42 = 1000 µF, 25 V, radial
C43 = 10 µF, 10 V, radial

Semiconductors:
D1 = BB 212
D2–D6,D10 = 1N4148
D7 = 1N4001
D8,D9,D11 = LED, 3 mm
B1 = B80C1500
LD1–LD4 = HD 11310
T1–T4 = BC 547 B
T5 = BC 557 B
IC1 = 74 HC 4060
IC2 = 74 HC 40103
IC3 = 4013
IC4 = 74 HC 4046
IC5 = 74 HCU 04
IC6 = 74 HC 4040
IC7 = 74 HC 151
IC8 = 4059
IC9,IC10 = 74 HC 4518
IC11 = 4067
IC12 = 74 HC 4017
IC13–IC16 = 74 HC 696
IC17 = 74 HC 682
IC18,IC20 = 4093
IC19 = 4001
IC21 = 7806
IC22–IC25 = 74 HC 4543

Miscellaneous:
X1 = quartz crystal 5.12 MHz
S1,S2,S4,S6,S7 = push-to-make switch
S3 = double change-over switch
S5 = SPST switch
Tr1 = mains transformer, secondary 9 V, 500 mA
PCB 910077-1
PCB910077-2
Front panel foil 910077-F
K1, K3 = 34-way D-type plug for PCB mounting
K2 = 14-way D-type plug for PCB mounting
L1 = 45 turns 0.2 mm dia. enamelled copper wire on Neosid Type 7A1S former (= 19µH)



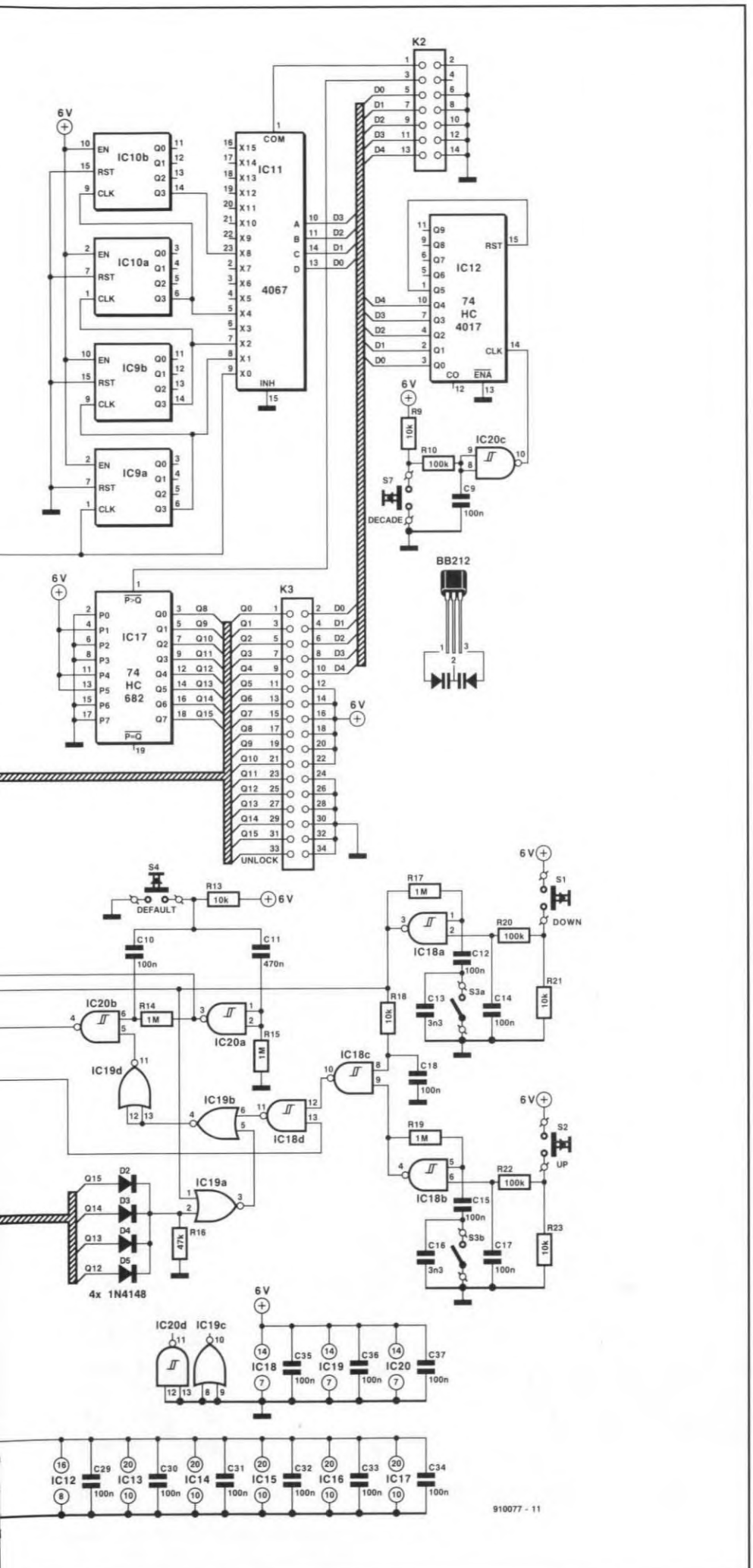


Fig. 6. Circuit diagram of the digital section of the function generator.

counters.

The heart of the section is IC_4, which contains the phase comparator and a VCO that is not used here. The comparator compares the two input signals, C_{in} (the reference frequency) and S_{in} (the signal frequency). Its output goes high when the phase of the signal frequency leads that of the oscillator frequency, and low in the opposite situation (see also Fig. 3). For this purpose, it uses the first transitions, not the duty factor, of the signals. After both transitions, irrespective of in which order, have been received, the output of the comparator has a high impedance until the next first transition arrives. The duration of the 'low' and 'high' periods is, therefore, directly proportional to the phase shift.

The low-pass loop filter at the output of IC_4 determines the properties of the loop as regards capture range, bandwidth, transient response and stability. After passing through this filter, the output voltage of the comparator controls the VCO

The state when the two signals are in perfect synchronization, that is, are locked, is indicated by the LED at the LOCK output of IC_4 going out.

The reference frequency of 160 Hz is obtained by dividing the 5.12 MHz oscillator frequency by 32 000.

The required frequency range of 32 Hz to 3.2 MHz is, however, not so easily obtained, because a large bandwidth leads to a long transient response and poor stability. That is why the range, outside the loop, is split into five by decadic divider IC_9–IC_{10}: the wanted range is selected by multiplexer IC_{11}. This means that the PLL is only active when the highest sub-range, that is, 320 kHz – 3.2 MHz, is selected.

The highest sub-range is itself further divided into four, so that the frequency of the VCO must be a multiple of two, which makes the PLL very stable. Depending on the set frequency, IC_6–IC_7 divides the VCO frequency by one, two, four or eight. The combined divisor of IC_{3b} and IC_8, which counts down from a preset number (max. 9 999) to 1 000, and IC_6–IC_7 lies, therefore, between 16 000 and 32 000. Since the frequency of the signal fed by IC_{3b} into IC_4 must always be 160 Hz (reference!), the frequency range of the VCO is 2.5–5.12 MHz.

The setting of the divisor within the PLL, and thus the frequency within one of the sub-ranges, is effected by four cascaded up-down counters, IC_{13}–IC_{16}. These counters have been set to a predetermined start value by jump leads. The start value may be overridden at any time by pressing the 'up' or 'down' push-button key.

The set divisor is strobed by comparator IC_{17}, whose output is active as long as the ratio is lower than 3 200. This is of importance for the locked filter in the analogue section.

The signal for the analogue section is taken from junction IC_7–IC_8 (pins 5 and 1 respectively). The frequency of this signal lies be-

tween 320 kHz and 3.2 MHz.

The signal is applied to divider IC_9–IC_{10}, whose outputs have frequencies of 10^0, 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} times that of the input.

The sub-range is selected by multiplexer IC_{11}, whose setting is determined by a 0–4 counter.

The analogue section thus receives a signal whose frequency lies between 32 Hz and 3.2 MHz and information, from D0–D4 of IC_{12}, as to which sub-range has been selected.

The display board is fed with data representing the frequency setting via bus lines Q0–Q15, D0–D4. It is also connected to the 'unlock' pin (PP) of IC_4.

Circuit description

Digital section

The reference oscillator is formed by fourteen-stage shift register IC_1. The 5.12 MHz clock is divided internally by 2^7. The resulting 40 kHz signals is taken from pin 6 and fed to a synchronous, binary down-counter, IC_2. A start count, related to the divisor (here :125) can be set at inputs P0–P7. From that start, the IC counts to zero, resets itself via pin 15 and at the same time triggers flip-flop) IC_{3a}.

The bistable functions as a binary scaler, which provides a pure signal with short tran-

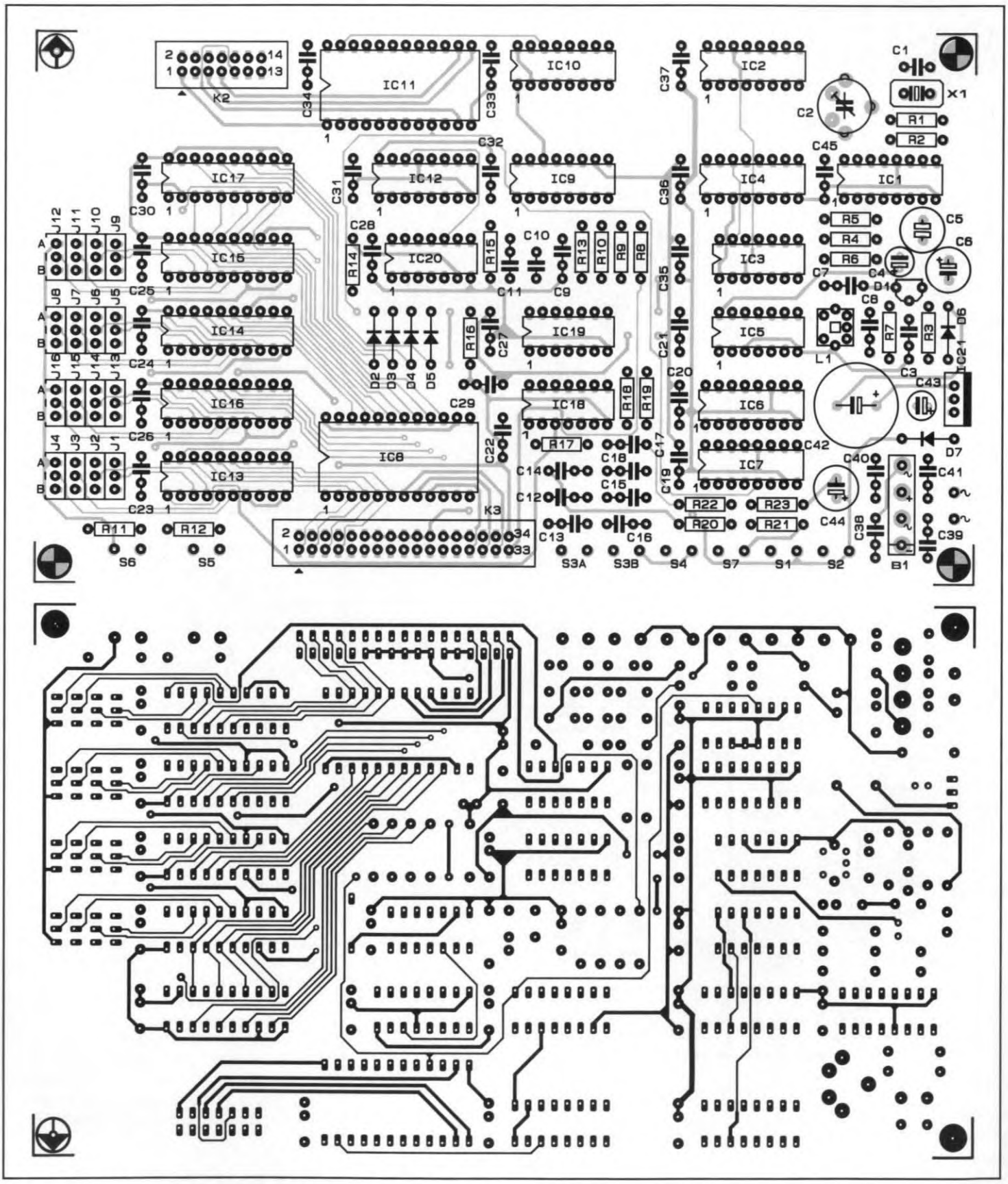


Fig. 7. Printed-circuit board for the digital section of the function generator.

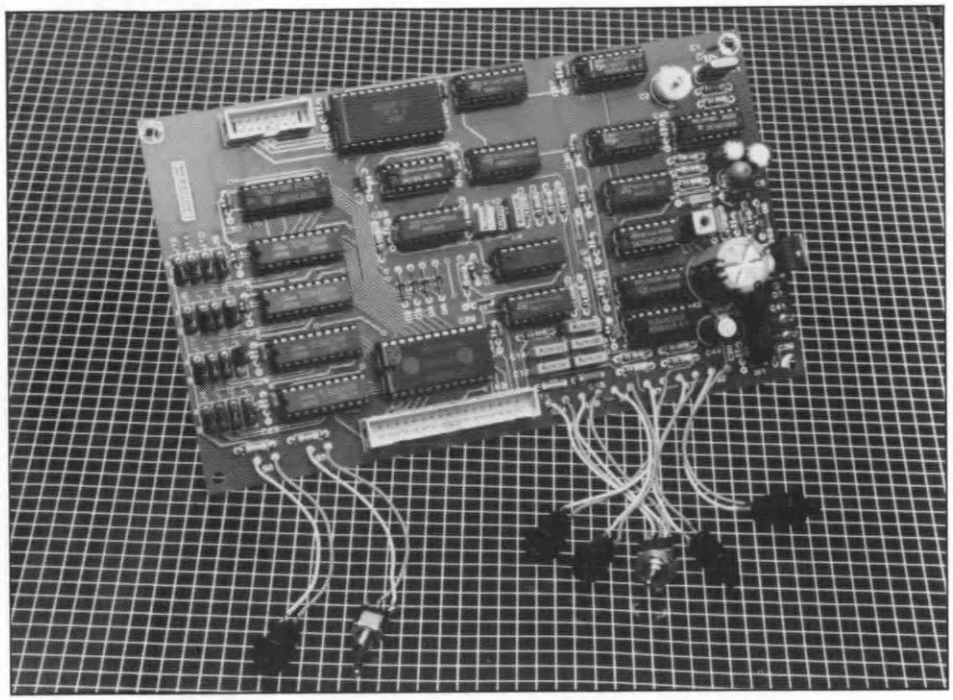


Fig. 8. Photo of the completed digital section board.

sition durations as required by IC_4. The loop filter consists of R_4–R_5–C_5–C_6.

The VCO consists of inverter IC_{5d}, which functions as an amplifier, and an *LC* circuit that is tuned by dual capacitance diode D_1. The control voltage for that diode is derived from potential divider R_6–R_7. The VCO signal is taken from across IC_{5c}.

The setting of divider IC_6–IC_8 is rather complex. The fundamental VCO frequency exists at pin 15 of IC_7; scaled by two at pin 9 of IC_6; divided by four at pin 7 of IC_6; and divided by eight at pin 6 of IC_6. Which of these signals is passed on depends on the three bits at inputs A–C of multiplexer IC_7.

The output of IC_7 (pin 5) is fed to IC_9-IC_{10}–IC_{11} and thence to the analogue section and is also used as the clock for programmable *n*–divider IC_8. This chip consists of four cascaded four-bit counters, which operate as decade scalers. The decade range of the three that determine the thousands, hundreds and tens is determined by appropriate bits at each of the groups of inputs J13–J16, J9–J12 and J5–J8 respectively. That of the fourth depends on the level at inputs KA, KB and KC. This counter determines the units of the overall divisor. If, for instance, *n* is to be 1507, J13–J16 must be set to 1000; J9–J12 to 1010; J5–J8 to 0000; and J1–J4 to 1110.

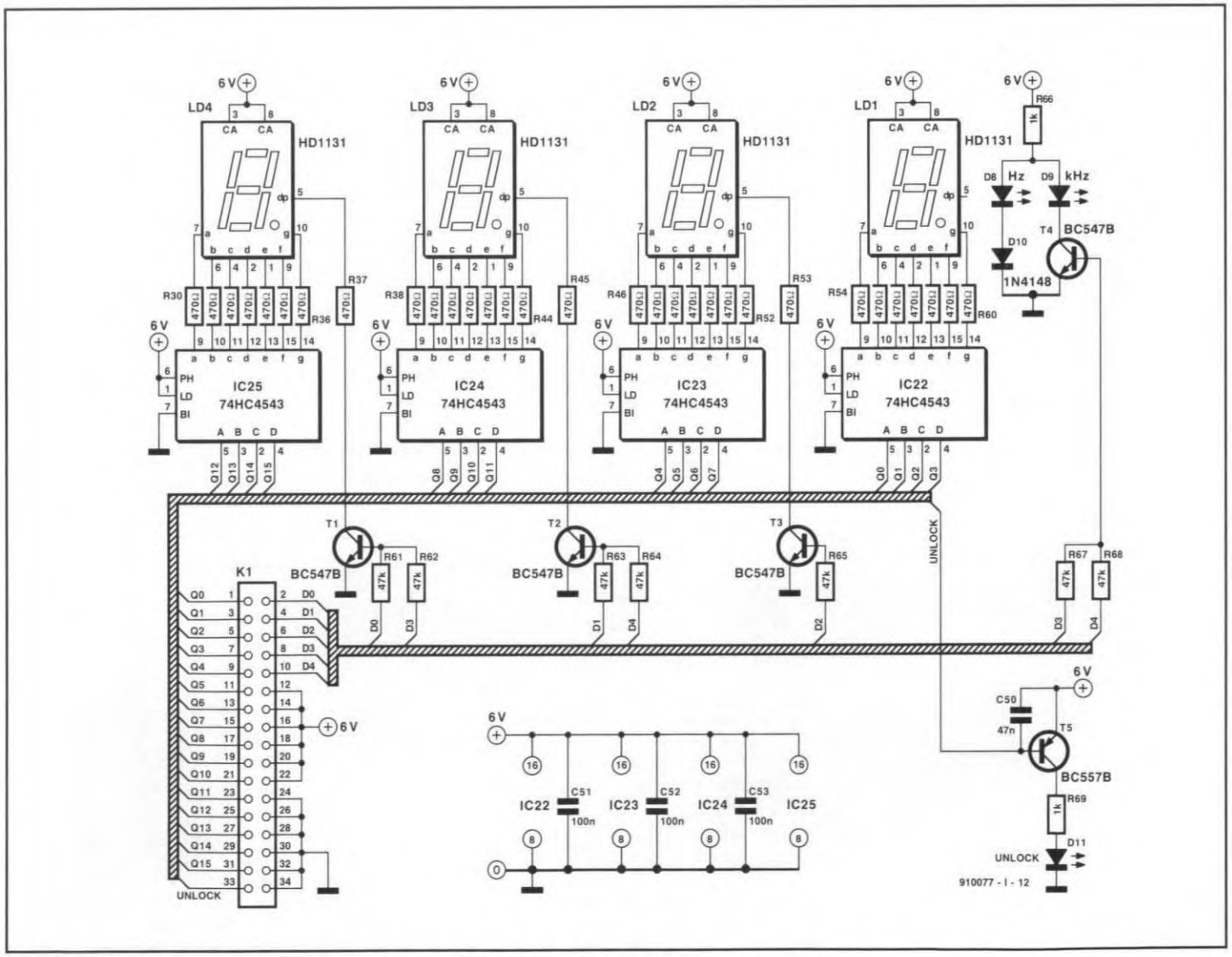


Fig. 9. Circuit diagram of the analogue (display) section of the digital function generator.

The first three are also applied to IC_7 in such a manner that when the divisor lies between 1 000 and 1 999, the VCO frequency divided by eight (pin 4 of IC_7) is internally coupled to the output (pin 5) of IC_7. When the divisor is between 2 000 and 5 999, pin 3 (VCO frequency divided by four) is connected to pin 5; when it lies between 6 000 and 7 999, pins 1–2 are connected to pin 5; and when it is between 8 000 and 9 999, pin 15 (VCO frequency) is coupled to pin 5. The overall divisor can, therefore, be set between 8 000 and 16 000.

The signal is applied to IC_9–IC_{10}, where it is divided by powers of 10 (10^1–10^4) as determined by the last transition of the clock. The four resulting signals, as well as the direct output of IC_7, are applied to pin 9 of multiplexer IC_{11}. This makes it possible to cover five decadic ranges with the same overall divisor, that is, 32–320 Hz, 320–3200 Hz, 3.2–32 kHz, 32–320 kHz and 0.32–3.2 MHz.

The multiplexer connects one of inputs X0–X4 or X8 to pin 1 (COM) and thence, via K_2, to the analogue section.

The multiplexer is controlled by DECADE switch S_7. Combination R_{10}–C_9 forms an anti-bounce network. When S_7 is pressed, a pulse is generated by Schmitt trigger IC_{20c} and used to clock Johnson counter IC_{12}. Successive outputs of IC_{12} go high at each clock input.

To avoid elaborate settings of the divisor by means of binary coded decimal switches and to make presetting possible, UP and DOWN switches, S_2 and S_1 respectively, are used in combination with counters IC_{13}–IC_{16}. These circuits function as synchronous decade scalers with presettable output latches. Figure 6 shows that these circuits consist of three stages: the counter/divider proper; a memory (4-bit register); and a three-state output. The circuits are cascaded in such a manner that IC_{13} processes the lowest decade, and IC_{16} the highest. Preset inputs A–D on each of the scalers are connected to ground or the positive supply line by wire bridges or jump leads, except pin 3 of IC_{16}, which is connected permanently to +6 V to ensure that the start-up value is always 1 000. It is, of course, possible to set this value to a higher number.

The counters may operate in either of two modes, determined by RECALL/U-D switch S_5. If their pin 11 is high, the output buffer writes the content of the register, which in its turn stores the data of the counters when a first transition appears at pin 9 (RCLK). This transition is generated by switch S_6 (STORE). In practice, this means that, when S_5 is set to RECALL, S_6 must be pressed briefly to ensure that the counter status is stored and passed to the output of the IC. When S_5 is set to U/D, pressing S_6 causes the counter status to be written into the register when the ICs receive the clock signal at their pin 2. Depending on the level at pin 1, the counters count up or down at every first transition.

The correct timing of the above process is ensured by the networks associated with

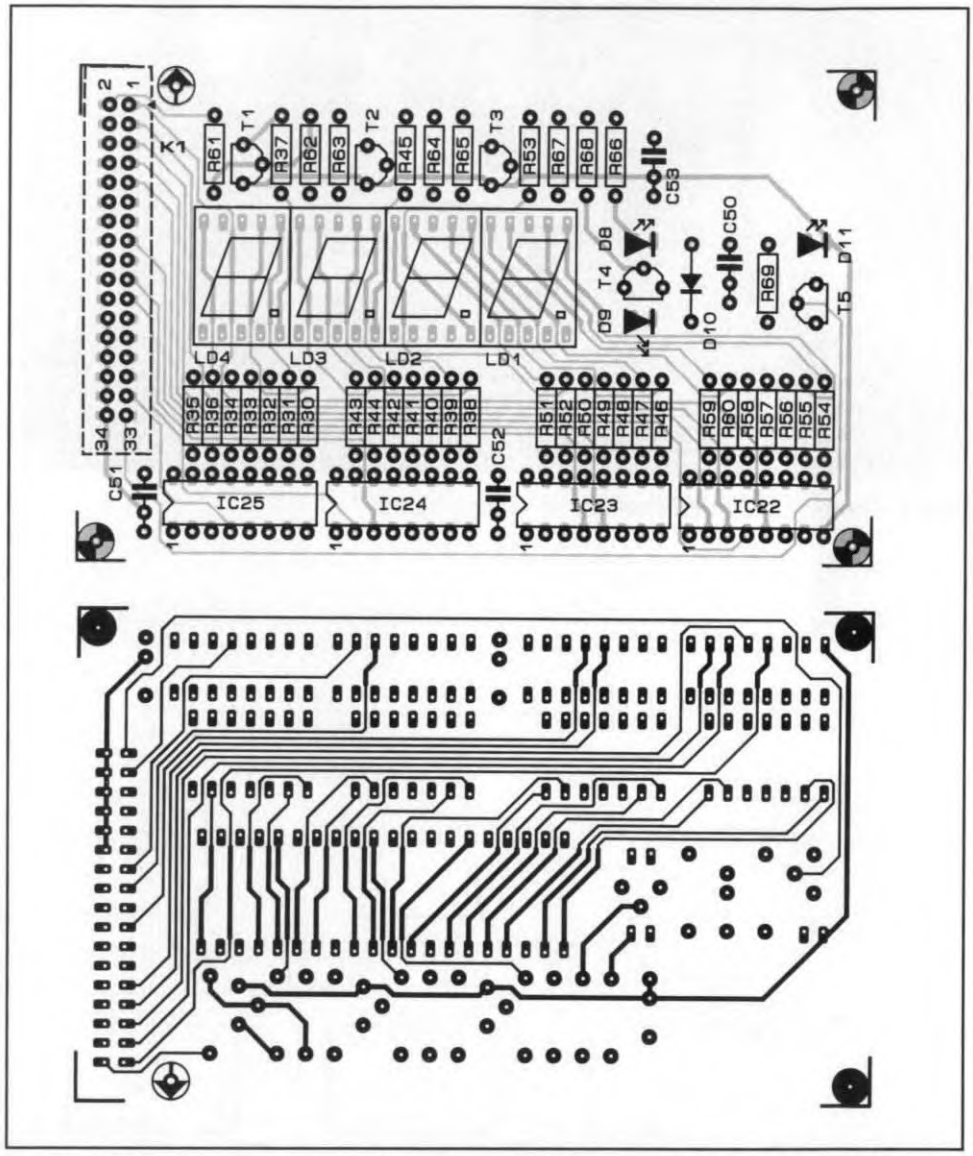


Fig. 10. Printed circuit board for the analogue (display) section of the generator.

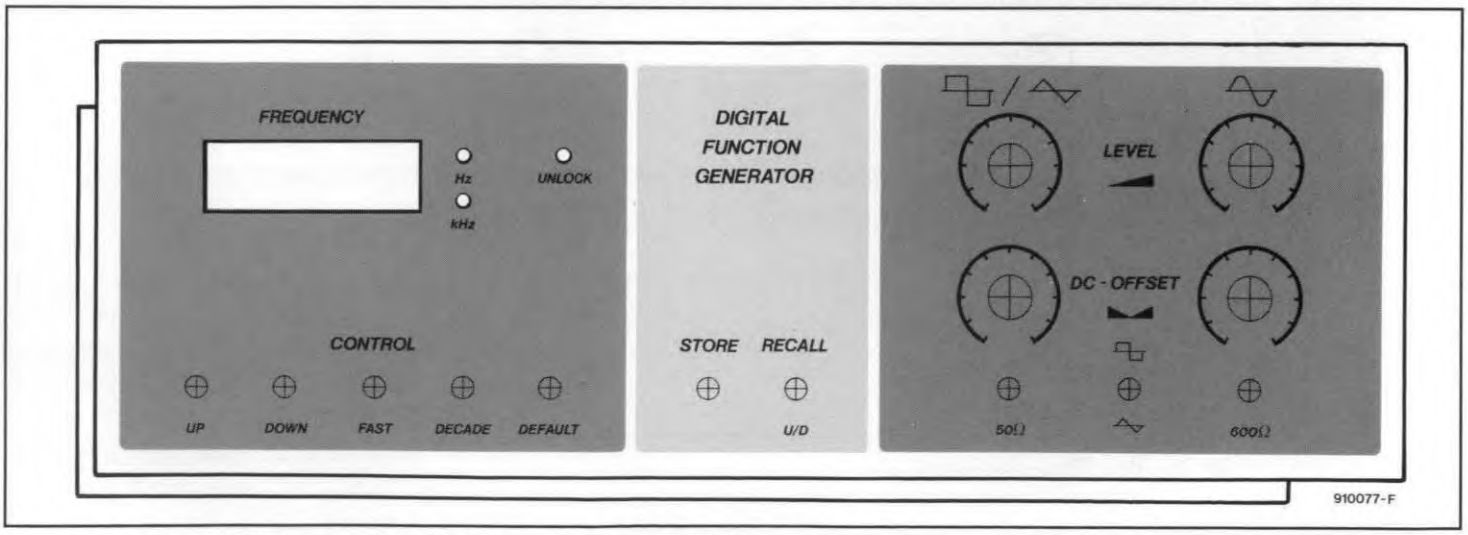


Fig. 11. The foil for the front panel is available through our Readers' services.

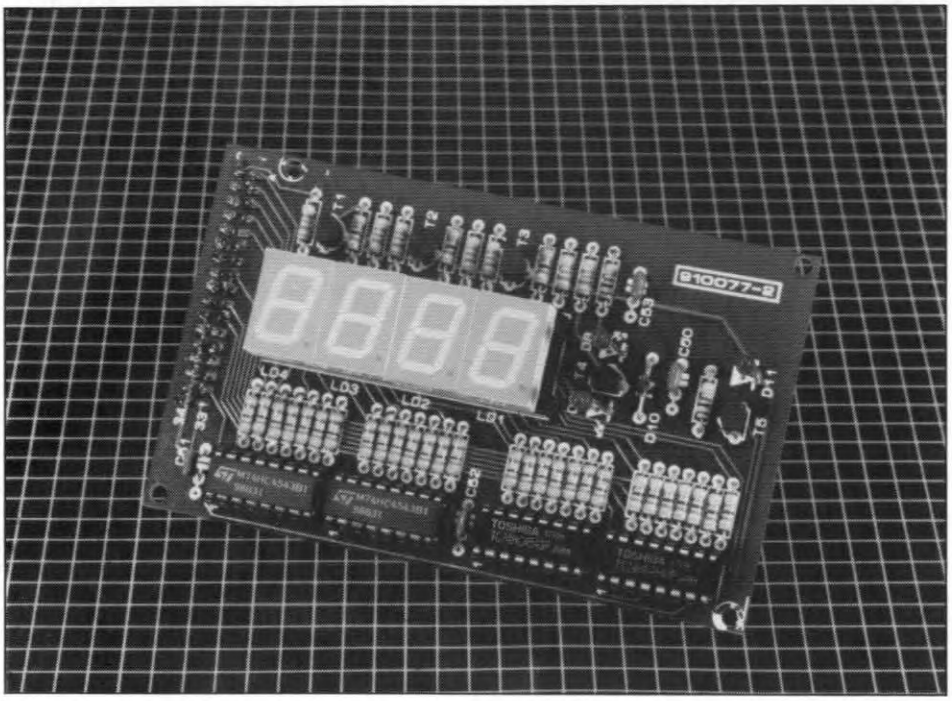


Fig. 12. Photo of completed analogue (display) section board.

switches S_1 and S_2.

When S_1 is pressed, a rectangular pulse is generated by IC_{18a} and applied directly to the U/D input of IC_{13}–IC_{16}. The frequency of the pulse is determined by C_{12} and C_{13}. The clock is delayed slightly by network R_{18}–C_{18} and inverted by IC_{18c}–IC_{18d}–IC_{19b}–IC_{19d}–IC_{20b} and then applied to pins 2. This arrangement ensures that the level at U/D is always low before the first transition arrives.

When S_2 is pressed, a pulse is generated by IC_{18b} which ensures that pins 1 remain high so that the ICs can count upwards.

Switch S_3 enables selection of a low or fast clock. The clock speed can be altered by changing the values of C_{13} and C_{16} (fast: S_3 open) or C_{12} and C_{15} (slow: S_3 closed).

Although the clock is connected to all four pins 2, the ICs do not count identically, because the RCO output of IC_{13} enables the ENT input of IC_{14} only when in counter position nine the next first transition arrives, that is, at every tenth input pulse. The same happens between IC_{14} and IC_{15} and between IC_{15} and IC_{16}.

Gates $IC_{19a\text{-}b\text{-}d}$, IC_{18d} and a (discrete) OR gate disable the counters when position 1000 or 9999 is reached. For instance, when the 10 000th pulse is received, the RCO output of IC_{16} is applied to NAND gate IC_{18b} so that the clock is disabled. The same happens when the four highest bits, Q12–Q15, are low simultaneously (divisor <1 000) and the output of the wired-OR gate, D_2–D_5, is also low.

The counters can be reset to the start value by pressing the DEFAULT switch, S_4. Network IC_{20a}–C_{11}–R_{15} generates the required pulse, which is applied to the LOAD input.

Analogue (display) section

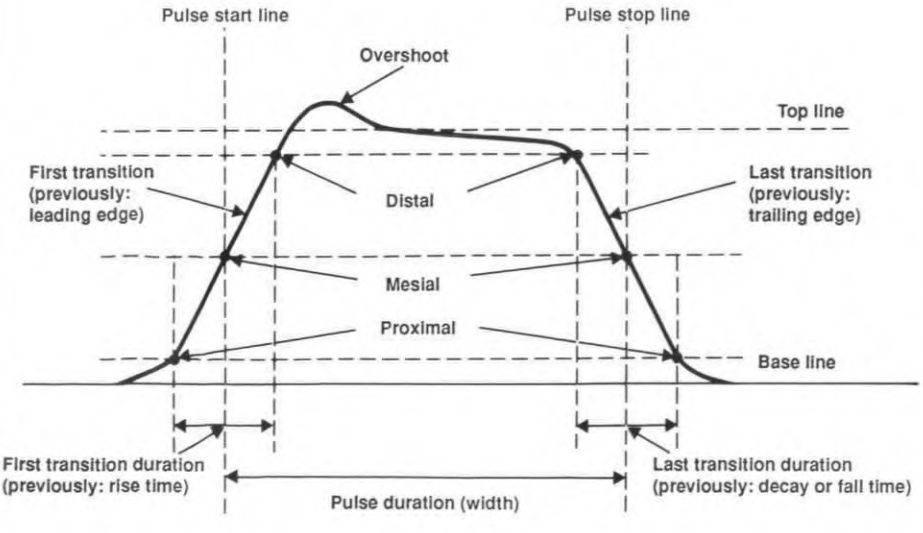
The display section (circuit diagram in Fig. 9 is constructed on a separate board, shown in Fig. 11 (see also Fig. 10). All inputs to the section are via plug and socket (K_1).

The Q-bus is split into four. Each of the four groups is coupled to a BCD-to-seven-bit converter, IC_{22}-IC_{25}, which drive the four display segments. The decimal points and the Hz/kHz LEDs are controlled via the D0–D4 lines.

Diode D_{11} lights when the PLL is not locked.

The second instalment of this three-part article will be published in our November 1991 issue.

Anatomy of a practical pulse



Anatomy of a practical pulse

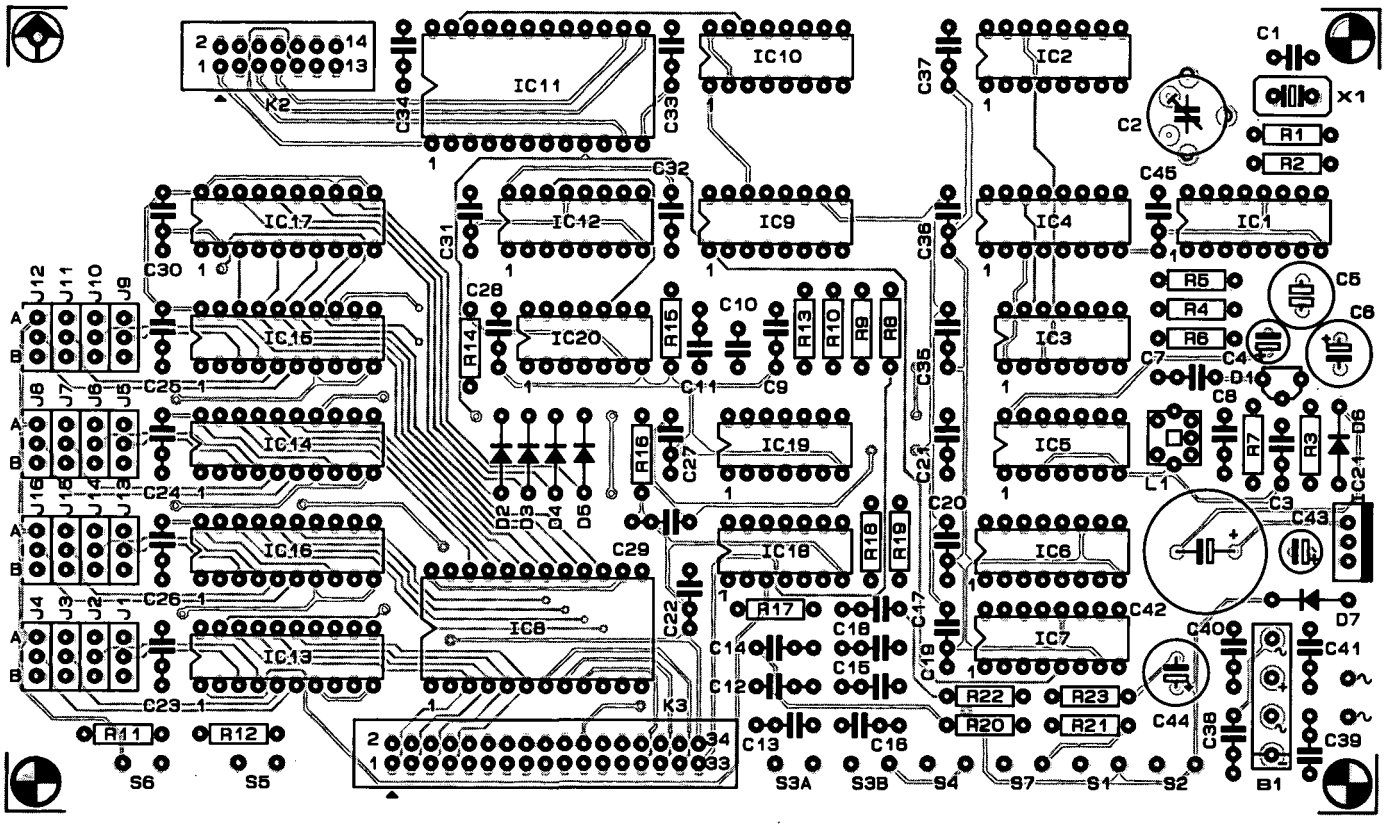
915081-11

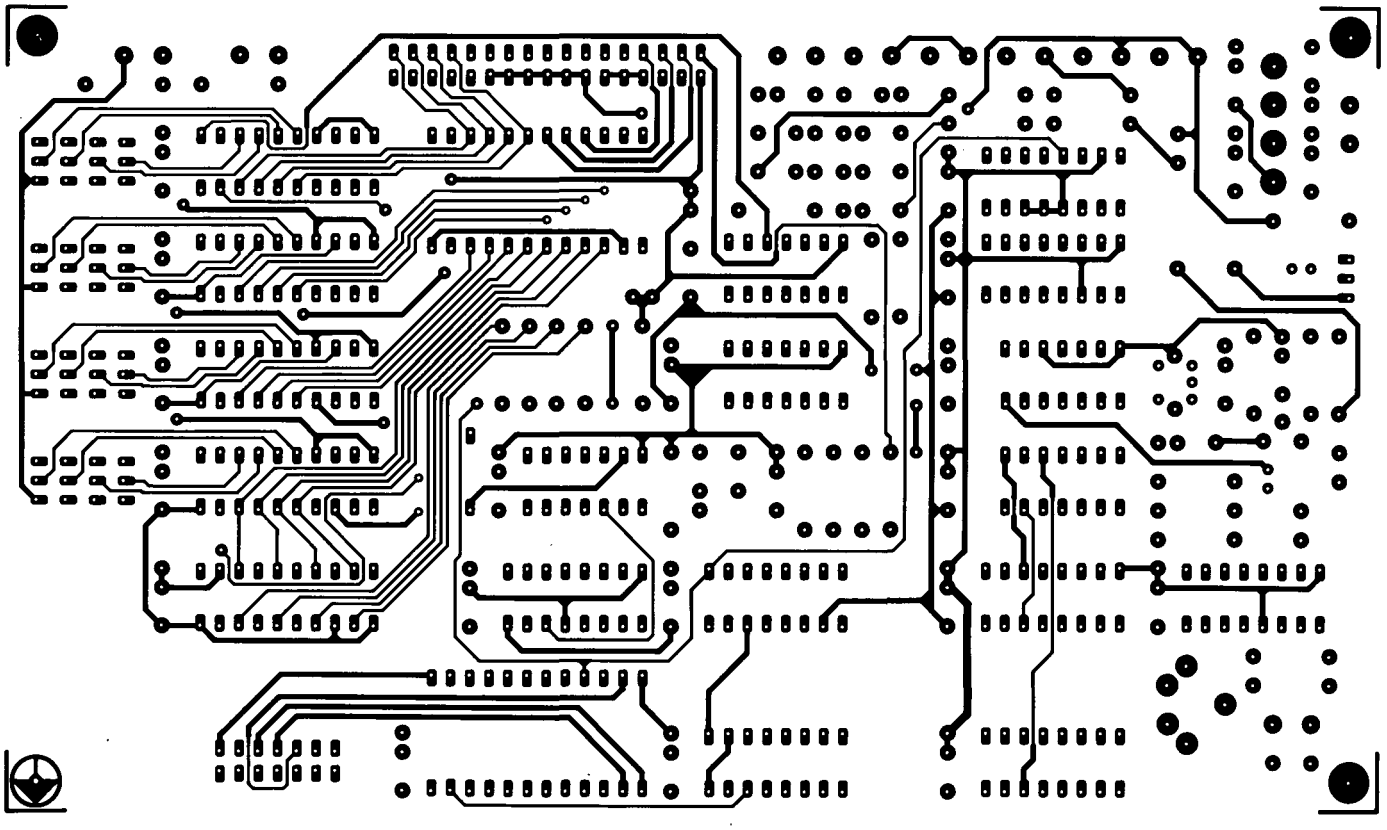
TECHNICAL literature is sprinkled with vague, misleading and ambiguous terms applied to a pulse. What does the term 'positive edge' mean when applied to a negative pulse? Is it the 'positive-going' edge, that is, in this case, the last transition, previously called 'decay' or 'fall' time, or is it the 'first transition', previously 'leading edge'?

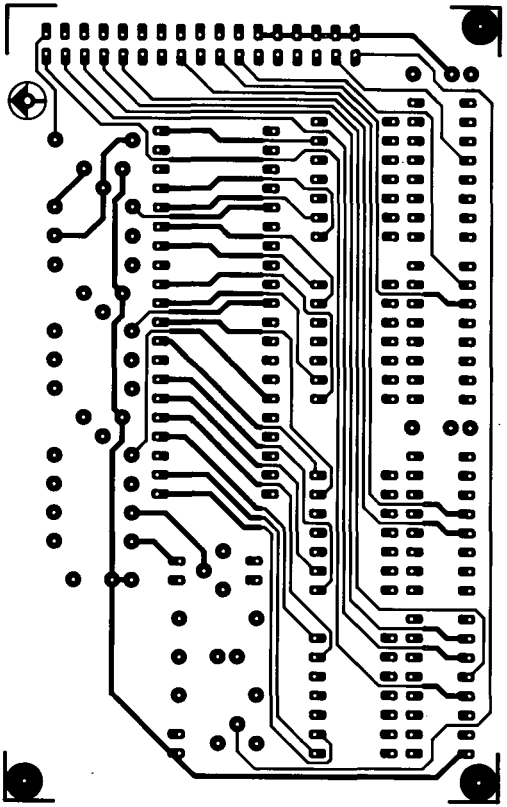
Another common error is to assume that the 'pulse duration (or width)' is the duration between the first and last proximals, whereas, in fact, it is the duration between the pulse start and stop lines.

It should also be noted that the term 'duty cycle' should not be used in connection with pulses; the proper term for the ratio of the pulse duration to the pulse repetition period of a periodic pulse train is 'duty factor'.

Fortunately, there are national and international standards (among them British Standards and IEC) that normalize the terms describing properties of a pulse waveform: the most important are shown in the drawing. ■


K2
IC11
IC10
IC2
C1
X1
C2
R1
R2
C45
IC17
IC12
IC9
IC4
IC1
C30
C31
C32
C36
C37
R5
R4
R6
C5
C6
C7
C4
D1
C8
IC16
IC20
IC3
R14
R15
R13
R10
R9
R8
C28
C10
C11
C9
C35
J12
J11
J10
J9
J8
J7
J6
J5
J16
J15
J14
J13
J4
J3
J2
J1
C25
C24
C26
C23
IC14
IC19
IC5
R16
C27
C21
R7
R3
D6
IC24
C3
C43
D2
D3
D4
D5
IC16
IC18
IC6
R18
R19
C20
C29
C22
R17
C18
C47
C19
C42
D7
C40
C41
IC8
IC13
IC7
C14
C15
C12
R22
R23
R20
R21
C44
C38
C39
B1
K3
C13
C16
C34
C33
R11
R12
S6
S5
S3A
S3B
S4
S7
S1
S2




C51
34
33
2
1
K1
IC25
R35
R36
R34
R33
R32
R31
R30
LD4
LD3
LD2
LD1
R61
T1
R37
R62
R63
T2
IC24
R43
R44
R42
R41
R40
R39
R38
C52
R45
R64
R65
T3
IC23
R51
R52
R50
R49
R48
R47
R46
R53
R67
R68
R66
D9
T4
D8
C53
D10
C50
IC22
R59
R60
R58
R57
R56
R55
R54
R69
T5
D11

A review of coding theory

by Brian P. McArdle

1. Introduction

The general area of Coding Theory for telecommunications and computer applications is reviewed to provide a simple introduction to the subject. For further information, the reader should consult the books in the reference section.

There is no formal definition of a code. Essentially, messages are represented in some form more easily transmitted than normal written language. In this article, a code is a digital electronic signal that represents a message symbol, such as a letter or number. For example, a teleprinter code would have to have a signal for every possible symbol (26 letters, 10 numerals and other symbols) and signals for every operation (that is, space, carriage return and line feed controls). Figure 1 shows the arrangement.

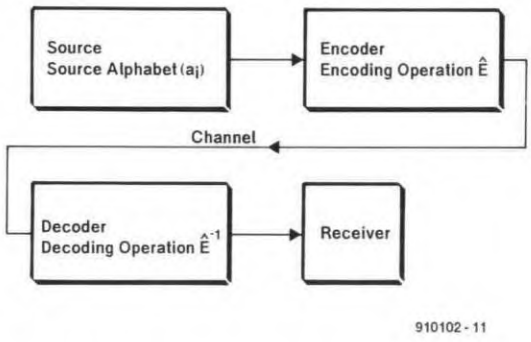


Fig. 1. Encoding/decoding operation.

An encoding operation Ê turns a message symbol a_j into coded form for transmission over a channel. The set $\{a_j\}$ is the source alphabet and $\{\Pr(a_j)\}$ is the set of probabilities associated with this alphabet: $\Pr(a_j)$ is the probability that a_j occurs. In normal language, this is the probability of occurrence of letters. The word 'channel' has a general meaning. It could be a cable, radio link or storage medium where the receiver is retrieving the messages at some later time. Obviously, the receiver must be able to apply a decoding operation $\hat{E}^{-1}$. Hence, the principal requirement for a satisfactory code is that the coded symbols be uniquely decodable. In mathematical terms, $\hat{E}[a_j]$ cannot represent more than one symbol. $\hat{E}[a_i]$ cannot equal $\hat{E}[a_j]$ unless a_i and a_j are effectively the same symbol. For example, Ê might not distinguish between upper and lower case letters. The decoded messages may be printed in upper case letters only. Thus, apart from small variations that should not affect normal understanding, the encoding operation, irrespective of its complexity or purpose, must be exactly and uniquely reversible.

A more formal mathematical definition is that the set of coded symbols $\{\hat{E}[a_j]\}$ must be uniquely partitioned (that is, can be divided into subsets that do not overlap) such that each partition can be associated uniquely with a source symbol. Figure 2 shows the arrangement.

The remainder of this article attempts to explain the meaning of Ê in different applications, such as error-detection-correction and encryption. It is always assumed that the encoding operation is uniquely reversible unless otherwise stated. Another important assumption is that a *memory-less source* is involved. The probabilities $\Pr(a_j)$s are the probabilities of the general occurrence of these symbols in normal language. In reality, letters occur in groups (digraphs and trigraphs). *i* before *e*, except before *c* is a well-known expression. Consequently, the probability of occurrence of a particular letter could be influenced by preceding letters. It is also assumed that a memory-less source is being considered unless otherwise stated. The analysis is mostly confined to digital coding except for Section 6, which deals with coding for analogue signalling.

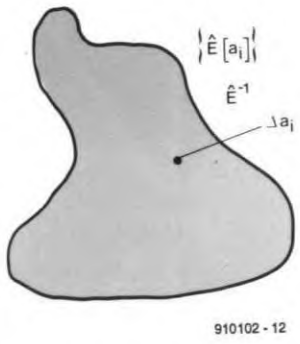


Fig. 2. Partition of the set of coded symbols.

2. Different codes

Codes can be analysed from many different viewpoints, but engineers are generally concerned only with two main categories.

(a) Fixed length codes

Every character of such a code is represented by a block of bits with every block having the same length. A typical example would be a computer code, such as ASCII or EBCDIC. Both of these use blocks of eight bits. Thus, there is a total of 2^8 or 256 difference blocks. Any two blocks of the same code would have to differ in at least one bit. The blocks need not be symbols (letters, numbers, punctuation marks) but can be controls (carriage return, line feed, etc.).

(b) Variable length codes

Consider an alphabet of five symbols $\{a, b, c, d, e\}$. In (**a**), this would require a code of three bits per block or symbol with $2^3-5 = 3$ redundant blocks. However, if the following arrangement of three blocks of two bits per block and two blocks of three bits per block is used,

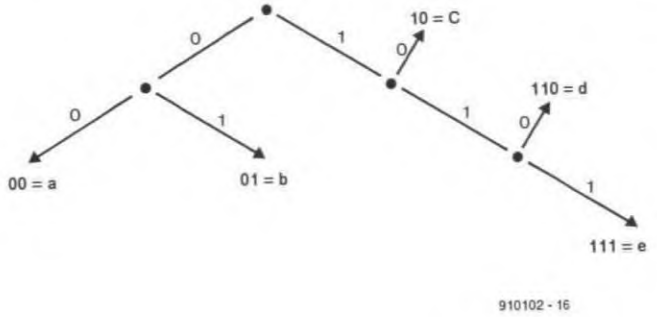
$$a = 00, b = 01, c = 10, d = 110, e = 111,$$

the average length of a message would be reduced. Blocks still have a specific length, but it is no longer the same fixed length. The basic requirement for unique decoding must still be maintained. For example, the bit sequence **011100011110** is easily decoded to ***bdaec*** with no errors. This must apply for all combinations of the symbols. An important quantity is the average length *L* of a block given by

$$L = \sum_j p_j \tau_j \qquad \text{[Eq. 1]}$$

where p_j is the probability of occurrence of block type j with τ_j bits. Ideally, this should be as small as possible to minimize the total number of bits per message. To ensure this requirement, the large probabilities would be paired with the smaller blocks. Morse code is another example where the common letter *e* is a dot, but *z* is two dashes followed by two dots.

This particular example has a special significance in addition to variable length blocks. If it is rewritten in the form of a diagram as below, it seems to have a tree-type structure with different branches.



Each branch is terminated by a symbol. The branches join together at nodal points which do not in themselves represent symbols. This arrangement indicates an *instantaneous* code. This means that the decoding operation does not require a 'memory', that is, it does not refer to blocks before or after any block that is being decoded. In the decoding of ***bdaec***, it was not necessary to test the 3rd bit before deciding that the first two bits, 01, represented *b*. This property remained true for the full operation and for all decoding operations irrespective of the combinations of symbols. (This should not be confused with a memory-less source, defined earlier, where there is no relationship between the occurrence of different symbols.). In an instantaneous code, no block can be a prefix or suffix for another block. Huffman codes, which are too involved to be considered in a simple overview, come into this category. However, it must be emphasized that any collection of blocks of varying lengths does not make an instantaneous code. There is a specific requirement given by the *Kraft inequality*

$$\sum_j \left(\frac{1}{2}\right)^{\tau_j} \leq 1$$

to form such a code. Further analysis is outside the scope of this paper and the reader is referred to the Reference Section for further study.

3. Information theory

Information theory has steadily increased its profile over the past few years and it is no longer possible to study telecommunications, especially coding, without touching on it somewhere. At first glance, the ideas behind it can appear too general and abstract for simple, direct applications. The fundamental fact is that the basic concepts of entropy, equivocation and channel capacity come from information theory, which, in turn, has influenced coding theory, and require some explanation.

There is a fundamental difference between an electronic signal and its value as information. In sound broadcasting, un unmodulated carrier would not convey any programme content to a listener. Therefore, there is a need to be able to quantify the value of a signal as information. In the 1920s, Hartley put forward the idea that the *logarithmic function* could be used as a measure of information. This was one of the landmarks in information theory. If two messages, a_i and a_j, are independent,

$$\log\{\Pr(a_i) \text{ and } \Pr(a_j)\} = \log\{\Pr(a_i)\} + \log\{\Pr(a_j)\} \qquad \text{[Eq. 2]}$$

and the base 2 is normally used. Remember that 'log' is not a linear function. The idea that the information contents of two independent messages is simply the sum of the information of each separate message seems instinctively correct. However, this method of measuring information has no connection with an actual signalling system. The *entropy* for $A=\{a_j\}$ is given by

$$H(A) = -\sum_j \Pr(a_j)\log\Pr(a_j) \qquad \text{[Eq. 3]}$$

and is a measure of the average information. An alternative explanation, which has become more common in recent years, is that it is a measure of the uncertainty in the information. $H(A)=0$ means that $\Pr(a_j)=1$ and $\Pr(a_i)=0$ for all other messages. Consequently, there is no doubt about the message. The maximum value occurs when the probabilities are the same and all messages are equally likely. If there are *n* possible messages, $H(A)$ is between 0 and $\log(n)$. The dimension is *information bits per symbol*.

To apply information theory to coding theory, consider Fig. 3 where there is a noisy channel between sender A and receiver B. The joint entropy is given by the equation

$$H(A,B)=H(A)+H(B/A) \qquad \text{[Eq. 4]}$$

where $H(B/A)$ is known as the conditional entropy or *equivocation*. This in turn is defined as

$$H(B/A) = \sum_j H(B/a_j)\Pr(a_j). \qquad \text{[Eq. 5]}$$

In non-mathematical terms, $H(B/A)$ is a measure of the information loss in transmission. The *channel capacity* is given by

$$C(A,B)=\text{maximum } H(A)-H(A/B). \qquad \text{[Eq. 6]}$$

This appears correct because the limit on the information conveyed over a channel is determined by the original uncertainty of that information (before reception) reduced by the uncertainty after reception. Essential capacity is limited only by noise and the *Hartley-Shannon law* sets an upper limit of $W\log(1+S/N)$, where *W* is the information bandwidth, *S* is the signal power and *N* is the noise power. For technical reasons, present-day systems operate well below this limit. The reader is referred to the Reference Section for further study.

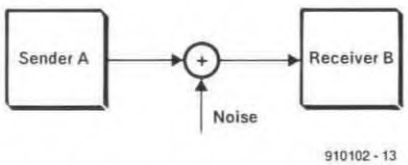


Fig. 3. Communications channel.

From the point of coding and electronic engineering, Eq. 6 can be simplified for normal use. Consider the case for a binary channel where *A* and *B* represent the input and output respectively. In general, the probability of a '0' or '1' is $^1/_2$, which gives $H(A)=\log(2)=1$. (The entropy of the source alphabet could be computed from the probability of occurrence of the various symbols, but it is the channel that is now under consideration.). If *p* is the probability of an error where a '0' is received as a '1', or vice versa, the channel conditional probabilities are

A	B	0	1
0		$(1-p)$	p
1		p	$(1-p)$

From Eq. 5:

$$H(A,B)=\Pr(0)[-(1-p)\log(1-p)-p\log(p)]+ \Pr(1)[-(1-p)\log(1-p)-p\log(p)] \qquad \text{[Eq. 7]}$$

which gives a new expression for the *channel capacity*:

$$C(A,B)=1+p\log(p)+(1-p)\log(1-p) \qquad \text{[Eq. 8]}$$

in *bits per symbol*. This is the usual expression in most textbooks on telecommunications. If the signalling rate is R symbols per second, the right-hand side is multiplied by R to give bits per second. Thus, information theory can be useful in the analysis of codes. The entire area has become extensive and has been treated only superficially here.

4. Error detection and correction

Error detection and correction is one of the main applications of coding theory and paralleled its development. Figure 3 showed the problems with errors where a '0' can be received as a '1' or vice versa. The use of the word 'receiver' is general in that it could represent a storage medium, and so on. It suffices to say that data is corrupted, which limits its value upon reproduction. Section 3 demonstrated that channel capacity is limited only by noise. To reduce the effects of errors, and therefore noise, extra bits are added to a block of data bits to create new and larger blocks, which in turn allow errors to be identified.

Consider the (7,4) *Hamming code* as follows:

Position: 7 6 5 4 3 2 1 $\quad c_4=(d_7+d_6+d_5) \bmod 2$

Bit: $d_7\ d_6\ d_5\ c_4\ d_3\ c_2\ c_1 \quad c_2=(d_7+d_6+d_3) \bmod 2$

$c_1=(d_7+d_5+d_3) \bmod 2$

There are three check bits in positions 1, 2 and 4 which have been derived from the data bits in the other four positions. The code is *linear* in the sense that the check bits are linear combinations of the data bits and the encoding operation is simply the application of the three linear equations. Since every block will have a total of seven bits without exception, the code is in the fixed length category. The position of the check bits within the block is very significant. A receiver generates the check bits from the received data bits and applies the decoding rule in Appendix 1. For example, if d_3 has been altered, c_1 and c_2 will not be validated and so on. The arrangement to check five data bits is given in Appendix 2. In both these examples, the set of coded blocks is such that the minimum variation between any two blocks in the same set is three bits. This is known as the *Hamming distance*. The reader is referred to Reference 2 for a more detailed explanation. The main point to note is that the method identifies only one error per block. In general, r check bits have (2^r-1) possible combinations and thus r bits in a total size of n bits must satisfy the condition $(2^r-1)\geq n$ in order to identify and therefore correct one error. To correct two or more errors per block, a code with a larger Hamming distance and more complicated arrangement would be needed.

Cyclic codes are the most commonly used for error detection and correction. These are also of the fixed length variety. For a block of total size n, the check bits are produced by a generator polynomial which is a factor of (x^n+1). A typical example is the specification MPT 1317 for the transmission of data over radio links. The format is as follows:

	DATA				*CHECK*		*PARITY*
Bit	64	63	62	17	16	15....2	1

with a block size of 64 bits. However, the first bit is for parity and is generated by the other 63 bits. The 15 check bits are generated from the 48 data bits using the generator polynomial

$$g(x)=x^{15}+x^{14}+x^{13}+x^{11}+x^4+x^2+1 \qquad \text{[Eq. 9]}$$

which is a factor of $(x^{63}+1)$. Refer to Appendix 3 for an exact breakdown. The data bits are the coefficients of the terms x^{62} down to x^{15} inclusive. Some books write the data bits on the right-hand side of the format, but this is not important provided they represent the high power terms of the polynomial. The polynomial consisting of only the data bits is divided by $g(x)$. The remainder is then added back to produce a new polynomial such that $g(x)$ is now a factor of the new polynomial. Since the check bits are essentially the original remainder, they represent the terms x^{14} down to x^0. Then a parity bit is added in order to detect odd numbers of 1s and the full 64-bit block should have even parity. Refer to Appendix 4 for the generation of a parity bit. The overall result is that the code can identify and correct up to four errors per block. This is a considerable improvement on the (7,4) Hamming code, but the operation is much more involved and the block size nine times larger. To check for errors, the receiver divides the polynomial by $g(x)$ and there should be no remainder..

Another example is the *POCSAG* code for paging, which uses the format

	DATA				*CHECK*		*PARITY*
Bit	32	31	30	12	11	2	1

and the generator polynomial

$$g(x)=x^{10}+x^9+x^8+x^6+x^5+x^3+1 \qquad \text{[Eq. 10]}$$

which is a factor of $(x^{31}+1)$. Refer to Appendix 5 for an exact breakdown. The overall method is the same with the 21 data bits generating the 10 check bits to produce a 31-bit block plus an extra parity bit.

5. Encryption

In encryption, the $\hat{E}$ operation, defined in Section 1, represents a secrecy operation and is usually written as $\hat{E}_K$ in most textbooks. The parameter K is known as the key and its purpose is to vary the operation. This is in complete contrast to error-detection-correction where exactly the same operation is performed on all blocks without exception. The importance of K is that it is generally the part of $\hat{E}_K$ that is kept secret. In a publicly known algorithm, such as the *Data Encryption Standard* (DES), the complete algorithm is known. A user chooses a key from the set of possible keys $\{K\}$ and encrypts the data. Thus, only encrypted data appears on the channel of Fig. 1. The data can be recovered by the inverse or decryption operation $\hat{E}_K^{-1}$ which also requires the correct key. If the key in use is kept secret and only known to authorized receivers, the data is kept secret. Obviously, $\{K\}$ must be sufficiently large to prevent an unauthorized user from trying each key in turn. There are a number of other requirements that are outside the scope of this paper.

There are three main methods of encryption.

(a) Stream encryption

In Fig. 4, each bit of the data is added modulo 2 using an XOR logic operation. A sequence of key bits is produced by the key generator such that each data bit is encrypted by its own particular key bit.

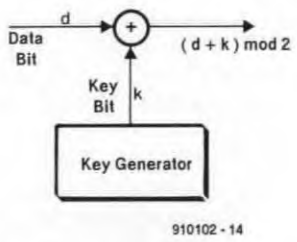


Fig. 4. Stream encryption.

The authorized receiver must know the method of key generation to reproduce the exact same sequence. The inverse operation is simply to apply the key sequence in the correct order to the sequence of encrypted bits. It would be too complicated to discuss the various techniques of key generation, but the most common method uses shift registers to generate a pseudo-random binary sequence. Generally, part of this process must be kept secret, such as the number of stages and feedback arrangements. The current proposals to provide en-

cryption facilities on the cellular system GSM or for digital short-range radio DSRR are believed to use a form of stream encryption. However, the information is confidential and it is very likely that the exact method will not be made public.

(b) Block encryption

Block encryption—see Fig. 5—differs from stream encryption in that a block is encrypted as a single unit. The most widely known method is the US *Data Encryption Standard*, which uses blocks of 64 bits for input and output. The algorithm was published in 1977 and the exact method is for public information. The actual key is a 56-bit block, so that the number of possible keys is 2^{56}. In operation, the authorized receiver would know the particular key in use and apply the inverse or decryption algorithm. Controversy has always surrounded the key size and recent articles have suggested methods for an improved DES.

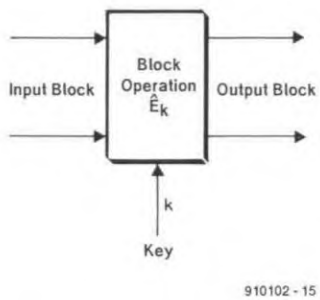


Fig. 5. Block encryption.

The main advantage over (a) is that a satisfactory block operation creates interdependence between the bits of a block. If one bit of an input block is varied, a number of bits in the output block are altered. However, it is generally much slower than stream encryption and cannot be used for high-speed telecommunications applications.

(c) Public key encryption

Public key encryption differs from the methods in (a) and (b) in that part of the key is made public, whence its name. The main requirement is that the part which must remain secret should not be easily deducible from the public part. A typical example is the RSA method introduced in 1978. Each user publishes two numbers, N and e. N is very large, of the order of 80 digits, and the product of two primes, P and Q, while e and d satisfy the equation

$$1 = ed \bmod (P-1)(Q-1). \qquad \text{[Eq. 11]}$$

Only e is made public; d, P and Q remain secret. If user A wishes to forward the message 'a' to user B, A looks up the parameters N and e for B and transmits

$$b = a^e \bmod N. \qquad \text{[Eq. 12]}$$

User B recovers the original message from

$$a = b^d \bmod N; \qquad \text{[Eq. 13]}$$

since d is one of the secret parameters, this cannot be done by any other user. From a secrecy point of view, an unauthorized user would have to factor N into P and Q to calculate d. Thus, as long as N is sufficiently large, this is impractical and the method is secure. There are other methods, such as the *Merkle-Hellman-Knapsack Method*, but they all follow the same principle of a public and a private key. Equations 12 and 13 are the equivalents of the encoding and decoding operations.

6. Coding for analogue signalling

In the preceding four sections, it was assumed that digital signal processing was in use. However, codes are also used in analogue electronics, but their application is rather limited. For example, in the PMR service, CTCSS (continuous tone controlled signalling system) has been around for a number of years. During transmission, an encoder generates a specific audio tone that modulates the radio frequency carrier. This tone is continuous for the duration of a message. In the absence of a CTCSS signal, the decoder at the receiving end is deactuated.

Another example is tone selective calling, such as EEA and ZVEI. In these methods, a sequence of five tones is used to form an address for a receiver. Both EEA and ZVEI have a total of 12 possible tones. Each possible address consists of a set of five, which actuates the receiver from the point of the user. However, despite these examples, coding has remained almost exclusively digital and the cellular GSM standard actually prohibits the use of tones.

On the secrecy side, there are voice scramblers that use frequency inversion, but increasingly the trend has been to digitize speech (for instance, ADPCM—Appendix 6) and to apply the techniques of Section 5. Coding in analogue signal processing is very restricted and need not be considered seriously.

7. References

(1) *Information Theory and its Engineering Applications*
D.A. Bell; Pitman (1968)

(2) *Coding and Information Theory*
Richard W. Hamming; Prentice-Hall (1980)

(3) *Information and Coding*
J.A. Llewllyn; Chartwell-Bratt (1987)

(4) *A First Course in Coding Theory*
Raymond Hill; Oxford University Press (1986)

(5) *Codes and Cryptography*
Dominic Welsh; Oxford University Press (1988)

(6) *Coding for Digital Recording*
John Watkinson; Focal Press (1990)

Appendix 1

The receiver re-calculates the check bits and validates them against the received values.

c_1 and c_2 are not validated $\Rightarrow$ d_3 is incorrect
c_1 and c_4 are not validated $\Rightarrow$ d_5 is incorrect
c_2 and c_4 are not validated $\Rightarrow$ d_6 is incorrect
c_1, c_2 and c_4 are not validated $\Rightarrow$ d_7 is incorrect

The sum of the indices of the check bits indicate the location of the erroneous bit. The correction process replaces a '0' by a '1' or vice versa. The principal difficulty is that two errors can cause a correct bit to be changed.

Appendix 2

The Hamming code for five data bits is $d_9\ c_8\ d_7\ d_6\ d_5\ c_4\ d_3\ c_2\ c_1$ and requires four check bits with the same procedure as in the (7,4) code. r check bits can test up to 2^r-1 locations. For $r=3$, this gives $n=2^3-1$, which leaves four data bits. For $r=4$, there is a block size $n=15$ which allows for 11 data bits and four check bits in the order:

$$d_{15}\ d_{14}\ d_{13}\ d_{12}\ d_{11}\ d_{10}\ d_9\ c_8\ d_7\ d_6\ d_5\ c_4\ d_3\ c_2\ c_1.$$

Appendix 3

For a 63-bit block, the factors of the modulus are:

$(x^{63}+1)=(x+1)(x^2+x+1)(x^3+x+1)(x^3+x^2+1)(x^6+x+1)(x^6+x^3+1)$
$(x^6+x^4+x^2+x+1)(x^6+x^4+x^3+x+1)(x^6+x^5+1)(x^6+x^5+x^2+x+1)$
$(x^6+x^5+x^3+x^2+1)(x^6+x^5+x^4+x+1)(x^6+x^5+x^4+x^2+1)$.

Each of these factors is irreducible in that it cannot be factored further while keeping real coefficients that are 0 or 1. In MPT 1317, the generator polynomial has the following factors:

$$x^{15}+x^{14}+x^{13}+x^{11}+x^4+x^2+1=(x^3+x^2+1)(x^6+x+1)(x^6+x^4+x^2+x+1)$$

and is a factor of $(x^{63}+1)$ as per the mathematical conditions.

Appendix 4

Consider the generation of a parity bit in the following table

d3	0 0 1 1
d2	0 1 0 1
c1	0 1 1 0

In mathematical terms, the appropriate generator polynomial, $g(x)$, divides the polynomial $(d_3x^2+d_2x+1)$. Let

$$g(x)=(x+1) \quad \text{and} \quad d_3x^2+d_2x+c_1=(w_1x+w_2)(x+1)$$

Equating terms on each side gives

$$w_1=d_3 \qquad w_1+w_2=d_2 \qquad w_2=c_1$$
$$w_2=w_1+d_2=d_3+d_2=c_1$$

as per the table. In Appendix 3, the generator polynomial is not a factor of $(x^{64}+1)$, but of $(x^{63}+1)$. The 64th parity bit has the generator polynomial $(x+1)$.

Appendix 5

For the *POCSAG*, the modulus has the following factors

$(x^{31}+1)=(x+1)(x^5+x^2+1)(x^5+x^3+1)(x^5+x^3+x^2+x+1)$
$(x^5+x^4+x^2+x+1)(x^5+x^4+x^3+x+1)(x^5+x^4+x^3+x^2+1)$

and the generator polynomial consists of the factors

$$x^{10}+x^9+x^8+x^6+x^5+x^3+1=(x^5+x^2+1)(x^5+x^4+x^3+x^2+1)$$

and thus divides $(x^{31}+1)$.

Appendix 6

ADPCM (adaptive differential pulse code modulation) was accepted by the CCITT in 1984 for encoding speech. Each sampled speech signal is originally encoded into a 12-bit block. This is compared with 16 quantizing levels and the nearest level chosen. This means that a block of 12 bits can now be replaced by a block of only four bits. The actual mechanism is quite complicated and is considerably different from PCM. The final result is a transmission rate of 12 kbit/sec. ■

FOUR-TERMINAL NETWORKS - PART 1

Getting to grips with attenuators

by Steve Knight, BSc

Most electronic equipment has an attenuator or attenuators of one sort or another. The object of these is to reduce to manageable levels a signal we have elsewhere worked like mad to build up. The usual objective is to turn out more than we want, then cut it down to the size we do want. This might sound like an easy option but, like many other things electronic, what we want and what we get aren't always identical. So, apart from the intrusion of Sod's Law which reigns universally, I hope that what follows will cast a ray of light on the often neglected subject of attenuator systems.

WE might define transmission networks in general terms as circuits that have two input terminals and two output terminals which are introduced between a generator and its load impedance. These networks, which are referred to as four-terminal networks, have properties that depend on the work they have to do in the transmission system; in the case of an attenuator, which is our main concern, it must enable us to obtain as output some desired fraction of the input which is entirely independent of the signal frequency. Clearly, such an attenuator system must be built up from purely resistive elements, since reactive components will lead to frequency discrimination over all or certain parts of the band. Further, existing impedance conditions in the system into which the attenuator is introduced must not be disturbed.

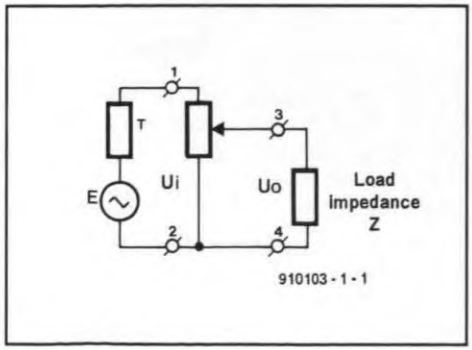


Fig. 1. Elementary attenuator.

The most elementary attenuator of all is the potential divider network that usually turns up in the form of a variable resistor. Figure 1 shows the system; our input goes between terminals 1 and 2 and we get the output from terminals 3 and 4. This is a four-terminal network, two of whose terminals are commoned, and we obtain an output that can be

a fraction of the input lying between the limits of 0 and 1. This is, no doubt, quite a satisfactory arrangement for turning the volume up and down on a radio receiver, but it fails dismally if we harbour ambitions for making precise quantitative measurements. Outside of the simplest applications to which such an attenuator might be put, the disadvantages are not hard to find.

The output has to feed into some kind of load impedance; assume for a moment that this is resistive. What the input terminals 'see', therefore, is not a constant resistance, but one that is made up of a non-linear combination of two resistances in parallel. Hence, the generator loading is variable and the potential difference—p.d.—at the input terminals is itself changing as the attenuator is operated. Not really the sort of thing we want if we are concerned with exactly how much attenuation we are getting and how it might affect the characteristics of the transmission path. What do we require from an attenuator if it is to do a worthwhile job? Well, we want it to introduce any needed degree of attenuation, but at the same time we want the input and output resistances to be such that the impedance conditions existing in the circuit are not upset in any way: if I want to insert an attenuator into a 300 Ω line, the attenuator impedance must also be 300 Ω.

Let us see how well-defined characteristics can be applied to attenuator networks so that we can get the quantitative results we want for our own particular requirements.

Basic characteristics

Four-terminal networks may come in two formats: symmetrical, in which we can interchange the input and output terminals without affecting in any way the electrical characteristics of the circuit; and asymmetrical, in which this last condition does not hold. The simple attenuator of Fig. 1 is clearly asymmetrical. Each of the two formats may be balanced or unbalanced, definitions to which we shall revert later on.

Symmetrical systems have two fundamental characteristics that are essential to our understanding of their function in life: the characteristic impedance, symbolized Z_0 in the complex case, and the attenuation constant α. For purely resistive networks, we can talk about characteristic resistance, R_0, and attenuation factor N.

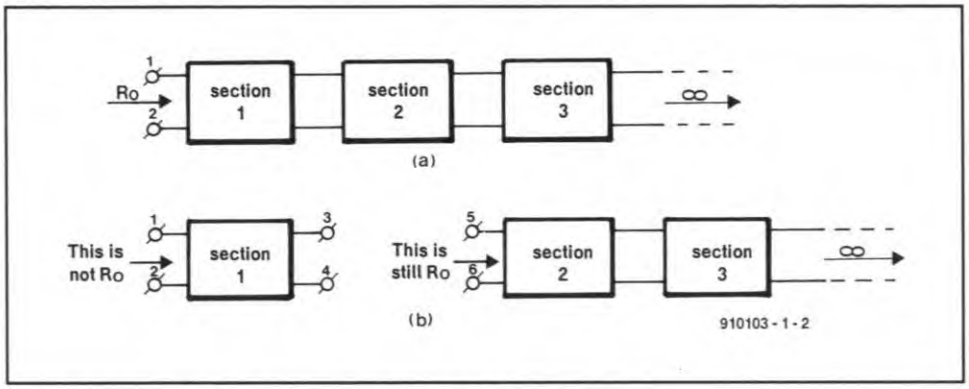


Fig. 2. Defining the characteristic resistance of a network.

Characteristic resistance

Let us use our imagination for a moment. Suppose we have a network made up of an infinite number of identical repetitive sections as in Fig. 2a. Each section contains a number of resistances, but how these are actually arranged is not important at this stage. Suffice it to say that the resistance measured at the input terminals, 1 and 2, will have a certain magnitude that will depend only on the nature and circuit arrangement of the individual sections. Suppose this resistance to be R_0.

Now, it might appear that we are getting into deep water by this approach: as the network is infinitely long, what chance do we have of calculating R_0 for a specific case where the contents of only one (or of a finite number) of the sections is known? Fortunately, there is a way out of the impasse by a relatively simple dodge: suppose we remove the first section of the infinite chain as in Fig. 2b. The input resistance of the remainder of the array as measured at terminals 5 and 6 will be the same as that measured originally at terminals 1 and 2, because the infinite nature of the assembly is unaffected by removing the first section (or any finite number of sections, come to that); so we could still measure R_0. The input resistance of the section we have removed, however, is very unlikely to be R_0, but we can make it so in one of two ways: by replacing the rest of the infinite chain (not a very practical way, to be sure), or by putting across terminals 3 and 4 a *single resistance* of a value equal to R_0.

Now, this second method gives us the clue we want: no matter how many finite sections we remove from the infinite chain, if we terminate them with a resistor of value R_0, the input resistance will also be equal to R_0 since these sections are effectively terminated by R_0 when reconnected to the rest of the infinite array. With this proper value of termination, that is, the characteristic resistance of each section, therefore, the input conditions are such that anything connected to the input terminals *cannot distinguish* between an infinite network or a finite network so terminated.

Hence, we can define the characteristic resistance, R_0, of an attenuator network: *any symmetrical network terminated in R_0 will have an input resistance also equal to R_0*. And this, of course, must also be true when working from output to input.

Attenuation factor

What about the attenuation constant, or the attenuation factor as it is more generally known for resistive networks? Well, this simply tells us the loss sustained in each section of a network. This loss may be expressed as a fraction of the input or as a reduction in decibels (dB); if the output is one-half the input, for example, the voltage (or current) attenuation is 6 dB.

Each section of an attenuator network will attenuate at identical rates, but the actual amount of attenuation is different as we proceed along the chain. Suppose, for instance, that we start off with unit input and lose half of this input in each section. The output of section 1 will be $^1/_2$ and this becomes the input to section 2. Here again, half the input is lost, so the output of section 2 will be $^1/_2 \times ^1/_2$ or $^1/_4$. After the following section, the output will be $^1/_4 \times ^1/_2 = ^1/_8$. Hence, the *amount* of loss differs for each section because the

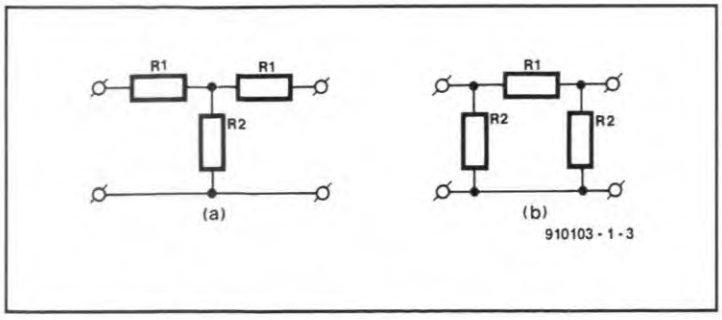


Fig. 3. The forms of the T-section and the π-section attenuators.

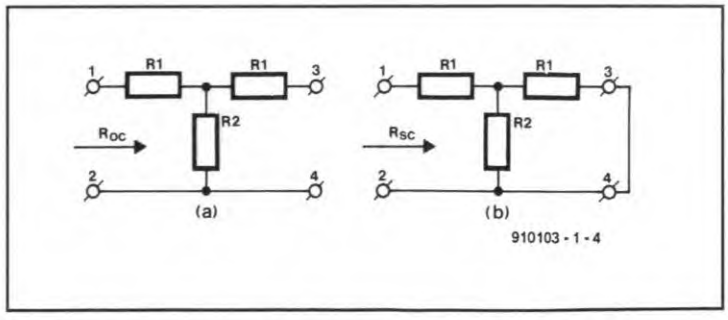


Fig. 4. Extreme termination conditions enable *Ro* to be calculated.

magnitude remaining to be attenuated at any stage is becoming progressively smaller. The *rate* of loss, however, remains constant.

For n sections, each reducing the input by some fraction p, the output will be p^n; if the attenuation is expressed in dB, n sections, each having a loss of p dB, will have an overall loss of np dB.

Finding R_o

Let us get down to the job of calculating the characteristic resistance of an attenuator section from a knowledge of its component parts.

The forms taken by individual sections of the general attenuator array are T–sections or π–sections; these are illustrated respectively in Fig. 3a and Fig. 3b. The T–section, as its configuration clearly implies, is made up of a divided series-arm and a central shunt-arm. Used between equal impedances, the section is symmetrical when the series-arm contains two equal resistances. The π–section, again as its form implies, consists of a single series-arm and two shunt-arms. This section is symmetrical when the shunt-arms are equal resistances. Both sections, though symmetrical in the way described, are unbalanced in the sense that the series-arm members are on one side of the section 'through' wires. It is essential not to get 'symmetry' confused with balance when talking about attenuator sections.

Working on the T–section for convenience (the π–section will lead us to exactly the same end-result), suppose we terminate the section at points 3 and 4 with the extreme conditions of, first, an open-circuit, and secondly, a short-circuit. Figures 4a and 4b show these cases. Clearly, in the first case, the input resistance seen at terminals 1 and 2 will be

$$R_{oc}=R_1+R_2,$$

and in the second case, it will be

$$R_{sc}=R_1+R_1R_2/(R_1+R_2).$$

Now, between these open-circuited and short-circuited conditions at the termination, there can be an infinite range of resistance values; as the termination changes through this range, the input resistance will change also. It seems reasonable, therefore, that there will be *some* value of the terminating resistance that will make the input resistance *also* equal to this value. This value

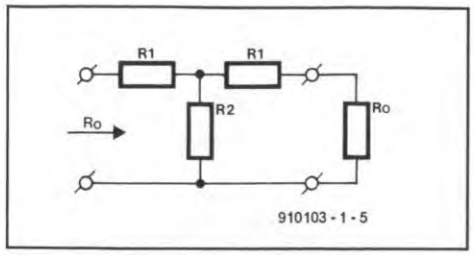


Fig. 5. A correctly terminated section.

must be, in accordance with our earlier stated definition, the characteristic resistance of the attenuator section. So, from Fig. 5 we have

$$R_o=R_1+R_2(R_1+R_o)/(R_1+R_2+R_o).$$

Solving this for R_o, we get

$$R_o=\sqrt{(R_1^2+2R_1R_2)}.$$

This enables us to find the characteristic resistance of a section from a knowledge of the resistor values making up the section.

We can now go one step further and find R_o without necessarily knowing the values of the elements used in a section. All we need to know is the input resistance (which can easily be measured) when the output terminals are either open-circuited or short-circuited. For, looking back a few lines, we have the product $R_{oc}R_{sc}$ expressible as

$$R_{oc}R_{sc}=(R_1+R_2)[R_1+R_1R_2/(R_1+R_2)],$$

and multiplying this out, we get

$$R_{oc}R_{sc}=R_1^2+2R_1R_2.$$

The right-hand side of this is R_o^2, so that

$$R_o=\sqrt{(R_{oc}R_{sc})},$$

which provides us with a very neat way of calculating the characteristic resistance of any section.

Cascaded sections

Although a single section will operate successfully as an attenuator, it is usual to have a number of sections in cascade or tandem so that a range of attenuation is provided. Once the R_o of a particular section has been found, another section may be added to it without affecting the overall characteristic resistance. For, if in Fig. 6a section A is terminated correctly by R_o, an identical section connected in place of R_o, will in turn correctly terminate A, since section A will be unable to distinguish between the presence of section B or the presence of a single terminating resistor equal in value to R_o.

It is plain that no matter how many such sections are wired in cascade, the input resistance will remain at R_o. What will change as we progress along the chain is the total attenuation: each section will introduce the same attenuation, but the desired overall attenuation can be achieved by using the required number of sections.

Finding the attenuation

Knowing the characteristic resistance of a circuit into which an attenuator is to be inserted, the problem of design boils down to finding suitable values for R_1 and R_2, given R_o and the required attenuation.

A desired value of R_o can be obtained with numerous combinations of R_1 and R_2; looking at Fig. 7, for instance, both sections shown have a characteristic resistance of 30 Ω (check on this for yourself!), but the network on the right will provide a greater degree of attenuation than the one on the left.

As we have already mentioned, the attenuation may be expressed as a fraction, that is, as a ratio of output voltage (U_o) to input voltage (U_i). Expressed in decibels,

$$\text{attenuation}=20\log(U_i/U_o).$$

If the ratio of the input power, P_i, and the output power, P_o, is taken,

$$\text{attenuation}=10\log(P_i/P_o) \quad [\text{dB}]$$

$$=20\log\sqrt{(P_i/P_o)} \quad [\text{dB}],$$

whence $U_i/U_o=\sqrt{(P_i/P_o)}=N$, the attenuation factor. Notice that with this notation the attenuation is expressed as a whole number, not a proper fraction. This often makes calculations easier.

In next month's concluding part of this article, we shall see how the attenuation can be provided, and how practical attenuators can be built to suit a variety of occasions.

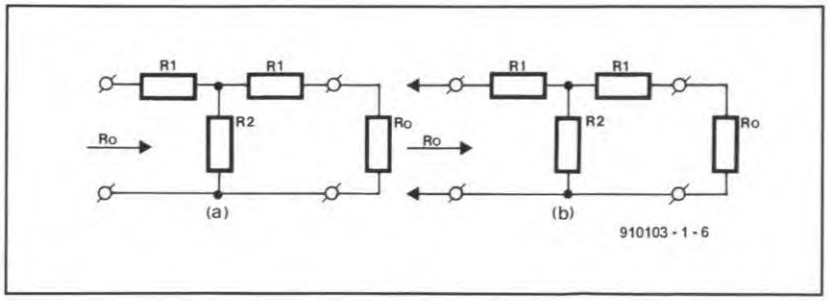


Fig. 6. How sections can be cascaded.

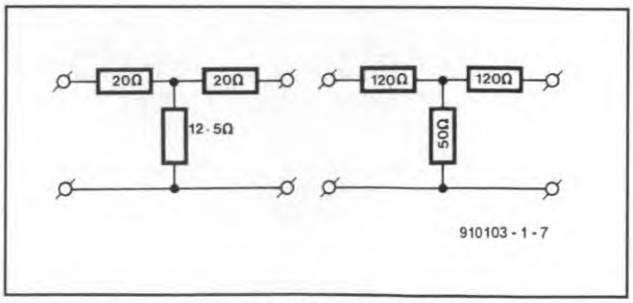


Fig. 7. Networks with identical R_o but different attenuations.

Fiber Optics — Part 2

Last month we discussed the basic theory behind optical fiber systems. It was learned that fiber optics are the light equivalent of microwave waveguides, and that they are capable of immense bandwidth communications. In this month's final instalment we will look at fiber communications and some of the circuits that can be used with experimenter grade optical fiber kits.

by Joseph J. Carr

Losses in optical fiber systems

Decibel notation can be used for optical fiber systems and refers to the gain or loss of each stage or component in the system. The dBkm scale uses decibels of gain or loss relative to the attenuation of a standard optical fiber section over a one kilometre (1 km) length. Alternatively, either dBmi (dB loss relative to attenuation over 1 mile) or dBl (a normalized unit length) can be used.

The light power at the output end of a optical fiber (P_o) is reduced compared with the input light power (P_{in}) because of losses in the system. As in many natural systems, light loss in the fiber material tends to be decaying exponentially (Fig. 1a), so obeys an equation of the form:

$$P_o = P_{in}\, e^{(-\Lambda / L)} \qquad [1]$$

Where:
Λ is the length of the optical fiber being considered;
L is the unit length, i.e., the length for which $e^{-\Lambda/L} = e^{-1}$.

There are several mechanisms for loss in fiber optics systems. Some of these are inherent in any light-based system, while others are a function of the design of the specific system being considered.

Defect losses

Figure 1b shows several possible sources of loss owing to defects in the fiber itself. First, in unclad fibers, surface defects (nicks or scratches) that breech the integrity of the surface will allow light to escape. In other words, not all of the light is propagated along the fiber. Second (also in unclad fibers), grease, oil or other contaminants on the surface of the fiber may form an area with an index of refraction different from what is expected and cause the light direction to change. And if the contaminant has an index of refraction similar to glass, then it may act as if it were glass and cause loss of light to the outside world. Finally, there is always the possibility of inclusions, i.e., objects, specks or voids in the material making up the optical fiber. Inclusions can affect both clad and unclad fibers. When light hits the inclusion it tends to scatter in all directions, causing a loss. Some of the light rays scattered from the inclusion may recombine either destructively or constructively with the main ray, but most do not.

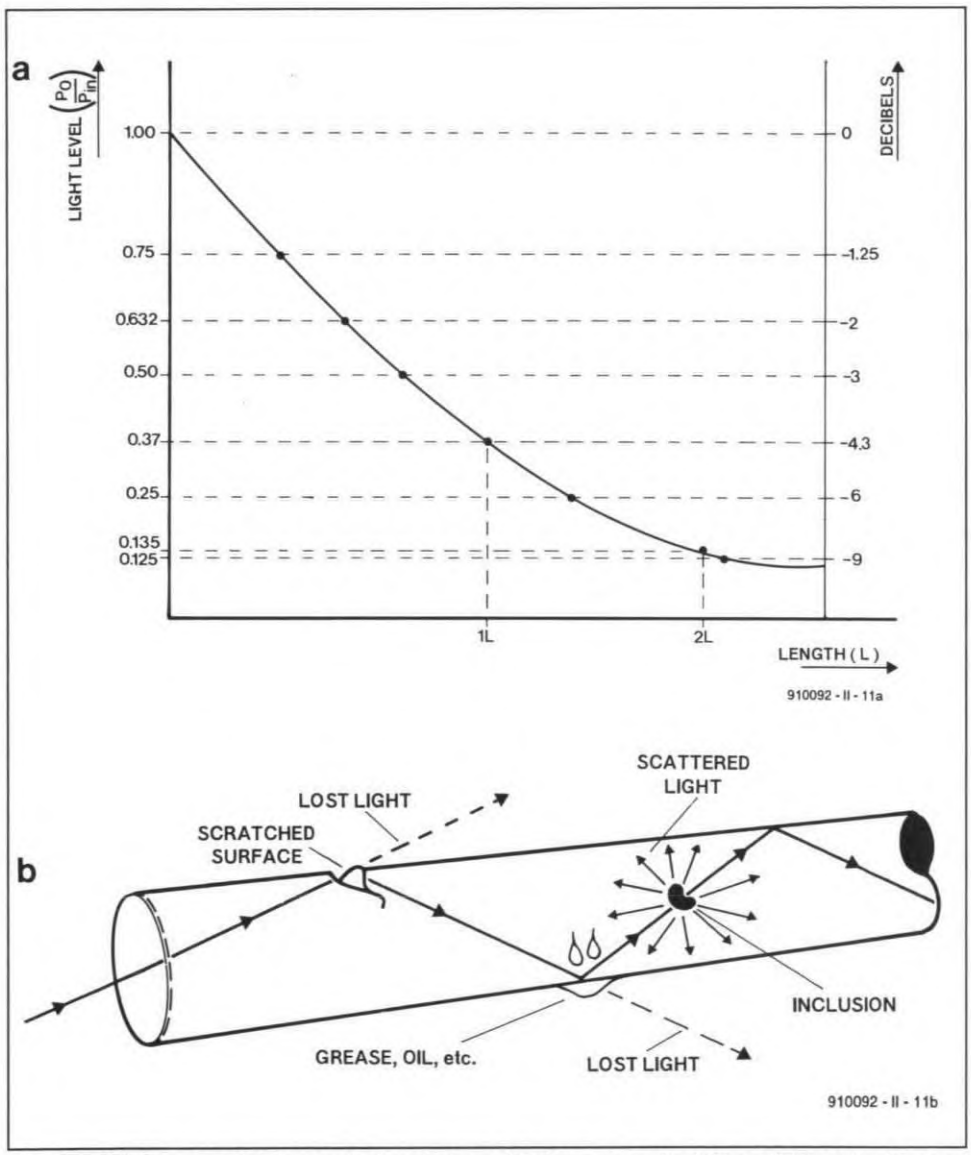


Fig. 1. (a) Exponential decay of signal strength along an optical fiber; (b) mechanisms of loss in fibers.

Inverse square law losses

In all light systems there is the possibility of losses caused by spreading (divergence) of the beam. If you take a flashlight and point it at a wall, and measure the illuminance per unit area at the wall at a distance of, say, one meter, and then back off to twice the distance (two meters) and then measure again, you will find that the illuminance has dropped to one-fourth. In other words, the illuminance per unit area is inversely proportional to the square of the distance ($1/d^2$).

Transmission losses

These losses are caused by light that is caught in the cladding material of clad optical fibers. This

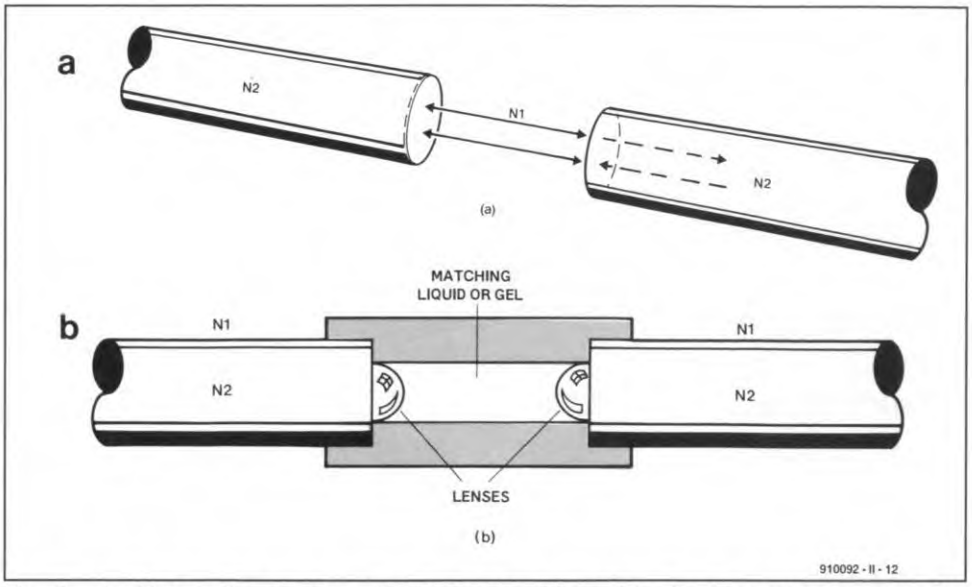


Fig. 2. (a) Air coupling creates reflection losses; (b) matching liquid coupling improves the junction.

light is either lost to the outside, or is trapped in the cladding layer and is thus not available for propagation in the core.

Absorption losses
This form of loss is caused by the nature of the core material, and is inversely proportional to the transparency of the material. In addition, in some materials absorption losses are not uniform across the light spectrum, but are thought to be wavelength sensitive.

Coupling losses
Another form of loss is caused by coupling systems. All couplings (of which more is said later) have loss associated with them. Several different losses of this sort are identified.

Mismatched fiber diameters
This form of loss is caused by coupling a large diameter fiber (D_L) to a small diameter fiber (D_S); i.e., the larger diameter fiber transmits to the lesser diameter fiber. As a ratio, this loss is expressed by:

$$\text{loss} = -10 \log \left(\frac{D_S}{D_L} \right) \quad \text{(dB)} \qquad [2]$$

Numerical aperture (NA) coupling losses
Another form of coupling loss occurs when the numerical apertures of the two fibers are mismatched. If NA_R is the numerical aperture of the receiving fiber, and NA_T that of the transmitting fiber, then the loss is expressed as:

$$\text{loss} = -10 \log \left(\frac{NA_R}{NA_T} \right) \quad \text{(dB)} \qquad [3]$$

Fresnel reflection losses
These losses occur at the interface between the optical fiber and air—see Fig. 2a—and are caused by the large change of refractive index at the glass-air barrier. There are actually two losses to consider. First is the loss caused by internal reflection from the inner surface of the interface, while the second is caused by reflection from the opposite surface across the air gap in the coupling. Typically, the internal reflection loss is of the order of 4 per cent, while the external reflection is about 8 per cent.

Any kind of reflection in an optical transmission system may be compared to reflections in a radio transmission line (standing-wave ratio). Studying standing waves and related subjects in books on RF systems can yield some understanding of these problems. The amount of reflection in coupled optical systems uses similar arithmetic:

$$\Gamma = \left(\frac{\rho_1 - \rho_2}{\rho_1 + \rho_2} \right)^2 \qquad [4]$$

Where:
Γ is the coefficient of reflection;
ρ_1 is the coefficient of reflection for the receiving material;
ρ_2 is the coefficient of reflection for the transmitting material.

Mismatched reflection coefficients are analogous to a mismatch of impedances in a radio transmission system. Whereas the cure in a transmission line system is to use an impedance matching device, in fiber optics, a coupler that matches the 'optical impedances', that is, the coefficients of reflection, is used. Figure 2b shows a coupling between the two ends of fibers (lenses may or may not be used depending on the system). The gel or liquid used to seal the couplings must have a coefficient of reflection similar to that of the fiber. The reflection losses are thereby reduced or even eliminated.

Optical fiber communications systems

A communications system requires an information signal source (e.g., voice, music, digital data, or an analogue voltage representing a physical parameter), a transmitter, a propagation media (in this case optical fibers), a receiver preamplifier, a receiver, and an output. In addition, the transmitter may include any of several different forms of encoder or modulator, and the receiver may contain a decoder or demodulator.

Figure 3 shows two main forms of communications link. The simplex system is shown in Fig. 3a. In this system a single transmitter sends light (information) over the path in only one direction to a receiver set at the other end. The receiver can not reply or otherwise send data back the other way. The simplex system requires only a single transmitter and a single receiver module per channel.

A duplex system (Fig. 3b) is able to simultaneously send data in both directions, allowing both send and receive capability at both ends. The duplex system requires a receiver and a transmitter module at both ends, plus two-way beam splitting Y-couplers at each end.

There is also a half-duplex system known in communications, but this is of little interest here. A half-duplex system can transmit in both directions, but not at the same time.

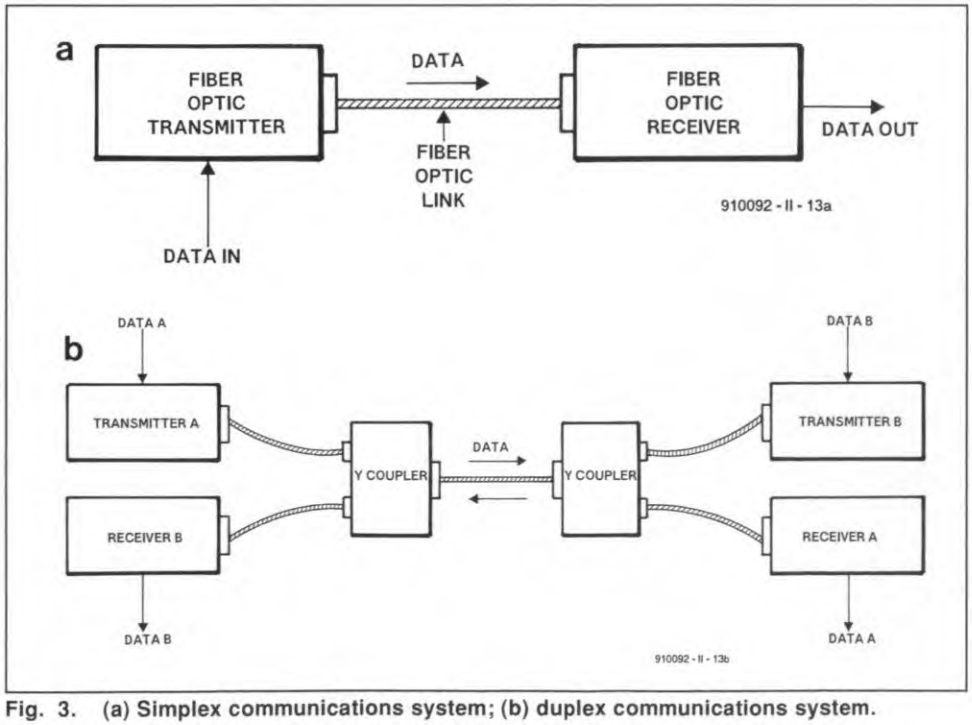


Fig. 3. (a) Simplex communications system; (b) duplex communications system.

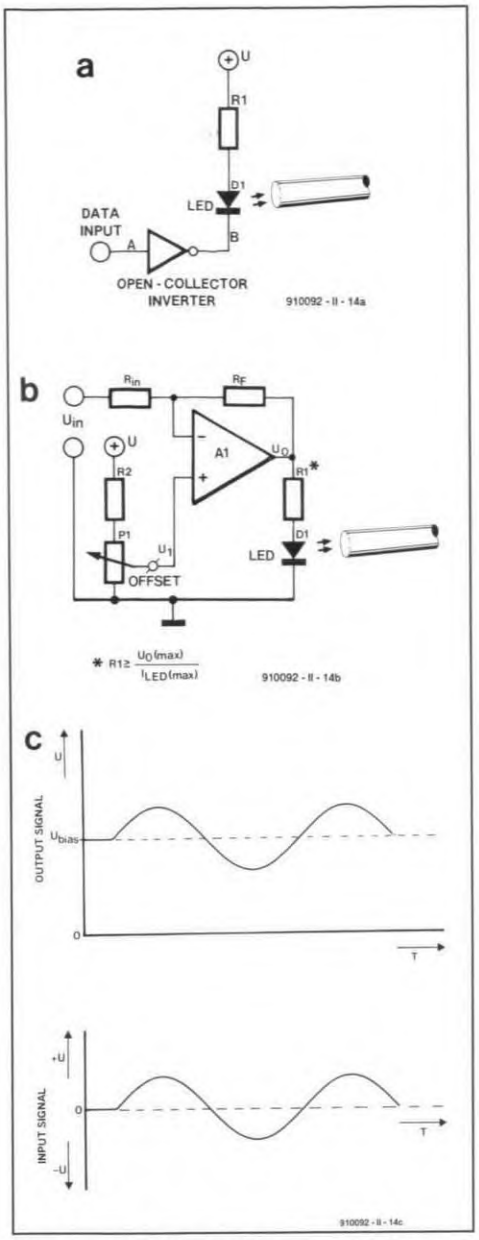


Fig. 4. (a) Digital driver circuit; (b) analogue driver circuit; (c) d.c. offset bias is needed to prevent distortion of the analogue signal.

Receiver amplifier and transmitter driver circuits

Before the fiber optics system is useful for communications, a means must be provided for converting electrical (analogue or digital) signals into light beams. Also necessary is a means for converting the light beams from the optical fibers back into electrical signals. These jobs are done by driver and receiver preamplifier circuits respectively.

Figure 4 shows two possible driver circuits. Both circuits use light emitting diodes (LEDs) as the light source. The circuit in Fig. 4a is useful for digital data communications. The signals are characterized by on/off (high/low or 1/0) states in which the light emitting diode is either on or off, indicating which of the two possible binary digits is required at the moment.

The driver circuit consists of an open-collector digital inverter. These devices obey a very simple rule: if the input (A) is high, the output (B) is low; if the input (A) is low, the output (B) is high. Thus, when the input data signal is high, point B is low, so the cathode of the LED is grounded. The LED turns on and sends a light beam down the optical fiber line. But when the input data line is low, point B is high, so the LED is ungrounded (and therefore turned off) — no light enters the fiber. The resistor (R_1) is used to limit the current flowing in the LED to a safe value. Its resistance is found from Ohm's law and the maximum allowable LED current:

$$R_1 = \frac{(U+) - 0.7}{I_{max}} \qquad [5]$$

An analogue driver circuit suitable for voice and instrumentation signals is shown in Fig. 4b. This circuit is based on an operational amplifier. There are two aspects to this circuit: the signal path and the d.c. offset bias. The latter is needed in order to place the output voltage at a point where the LED lights at about one-half of its maximum brilliance when the input voltage, U_{in}, is zero. That way, negative signals will reduce the LED brightness, but will not turn it off (see Fig. 4c). In other words, biasing avoids clipping of the negative peaks. If the expected signals are monopolar, U_1 is set to barely turn on the LED when the input signal is zero.

The signal U_{in} sees an inverting follower with a gain of R_F/R_{in}, so the total output voltage (accounting for the d.c. bias) is:

$$U_o = \left(\frac{-U_{in} R_F}{R_{in}}\right) + U_1 \left(\frac{R_F}{R_{in}} + 1\right) \qquad [6]$$

Because the network R_2/P_1 is a resistor voltage divider, the value of U_1 will vary from 0 volts to a maximum of:

$$U_1 = \frac{(U+) P_1}{R_2 + P_1} \qquad [7]$$

Therefore, we may conclude that $U_{o(max)}$ is:

$$U_{o(max)} = \left(\frac{-U_{in} R_F}{R_{in}}\right) + \left(\frac{(U+) P_1}{R_2 + P_1}\right)\left(\frac{R_F}{R_{in}} + 1\right) \qquad [8]$$

Three different receiver preamplifier circuits are shown in Fig. 5: analogue versions are shown in Figs. 5a and 5b, while a digital version is shown in Fig. 5c. The analogue versions of the receiver preamplifiers are based on operational amplifiers. Both analogue receiver preamplifiers use a photodiode as the sensor. These p-n or p-i-n junction diodes produce an output current, I_o, that is proportional to the illumination on the diode junction.

The version shown in Fig. 5a is based on the inverting follower circuit. The diode is connected so that its non-inverting input is grounded, i.e., at zero volts potential, and the diode current is applied to the inverting input. The feedback current (I_F) exactly balances the diode current, so the output voltage will be:

$$U_o = -I_o R_F \qquad [9]$$

The non-inverting follower version shown in Fig. 5b uses the diode current to produce a voltage drop (U_1) across a load resistance, R_L. The output voltage for this circuit is:

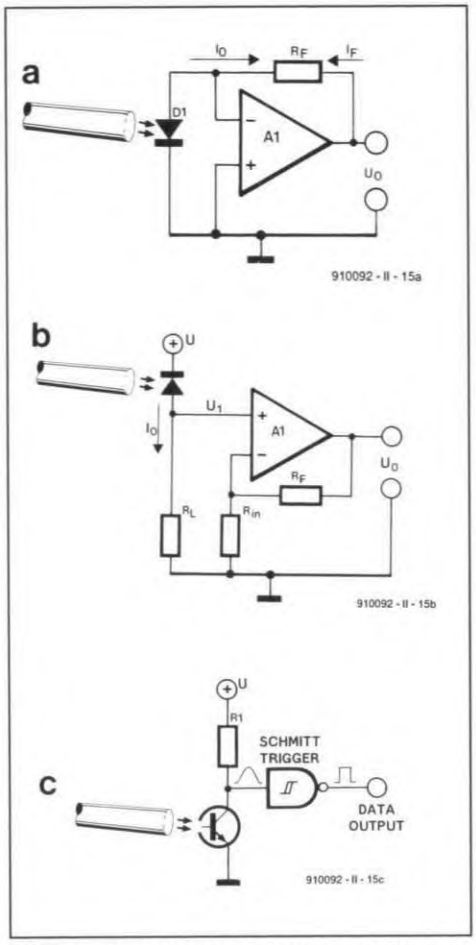


Fig. 5. (a) Inverting follower receiver; (b) non-inverting follower receiver; (c) digital receiver based on Schmitt trigger.

$$U_o = I_o R_L \left(\frac{R_F}{R_{in}} + 1\right) \qquad [10]$$

Both analogue circuits will respond to digital circuits, but they are not at their best for that type of signal. Digital signals will have to be reconstructed because of sloppiness caused by dispersion. A better circuit is that of Fig. 5c. In this circuit, the sensor is a phototransistor connected in the common emitter configuration. When light falls on the base region, the transistor conducts, causing its collector to be at a potential only a few tenths of a volt above ground potential. Conversely, when no light falls on the base, the collector of the transistor is at a potential close to $U+$, the power supply potential.

The clean-up action occurs in the following stage: a Schmitt trigger. The output of such a device will snap high when the input voltage exceeds a certain minimum threshold, and remains high until the input voltage drops below another threshold (these thresholds are not equal). Thus, the output of the Schmitt trigger is a clean digital signal, while the sensed signal is a lot more sloppy.

Conclusion

Fibre optics can be used by experimenters and small users: even Radio Shack (Tandy) has fiber optics kits available. Other suppliers can be found by perusing the ads of this magazine. Edmund Scientific (USA) has a particularly nice selection. ■

AM BROADCAST RECEIVER

In these days of stereo FM, medium-wave AM reception is often viewed with contempt, and the quality of a sophisticated hi-fi stereo tuner is solely judged by the specifications of the FM section, with the AM section taken for granted. No wonder that the AM reception of these hi-fi tuners is no better than that of any cheap portable radio.

by R. Shankar

THIS article describes a top-of-the-range AM receiver with the following features:

- High quality audio output with a –6 dB bandwidth of 12 Hz to 4.5 kHz, and a total harmonic distortion of only 0.5%.
- High input signal range of nearly 60 dB with virtually no change in the audio output level.
- A phase-locked loop (PLL) which keeps the intermediate frequency (IF) at exactly 455 kHz even when the receiver is slightly mistuned to a station. Since a small difference between the tuning of the antenna and that of the oscillator sections is inevitable in any receiver, this relaxation can be considered as a kind of fine tuning. Also, oscillator drift has no effect on the IF.
- An extremely steep IF cut-off characteristic thanks to a communications-grade ceramic filter.

Principle of operation

Basically a superheterodyne type, the proposed receiver is considerably more sophisticated than a good many hi-fi stereo tuners. A few salient differences are listed below.

AFC

As mentioned before, the local oscillator frequency is adjusted by a PLL so that the IF is held at exactly 455 kHz. This AFC (automatic frequency control) function may be turned off by a switch, in which case the local oscillator runs free, functioning just as in most other receivers. A certain level of carrier has to be present for the AFC to function. Otherwise, it is automatically switched off to en-

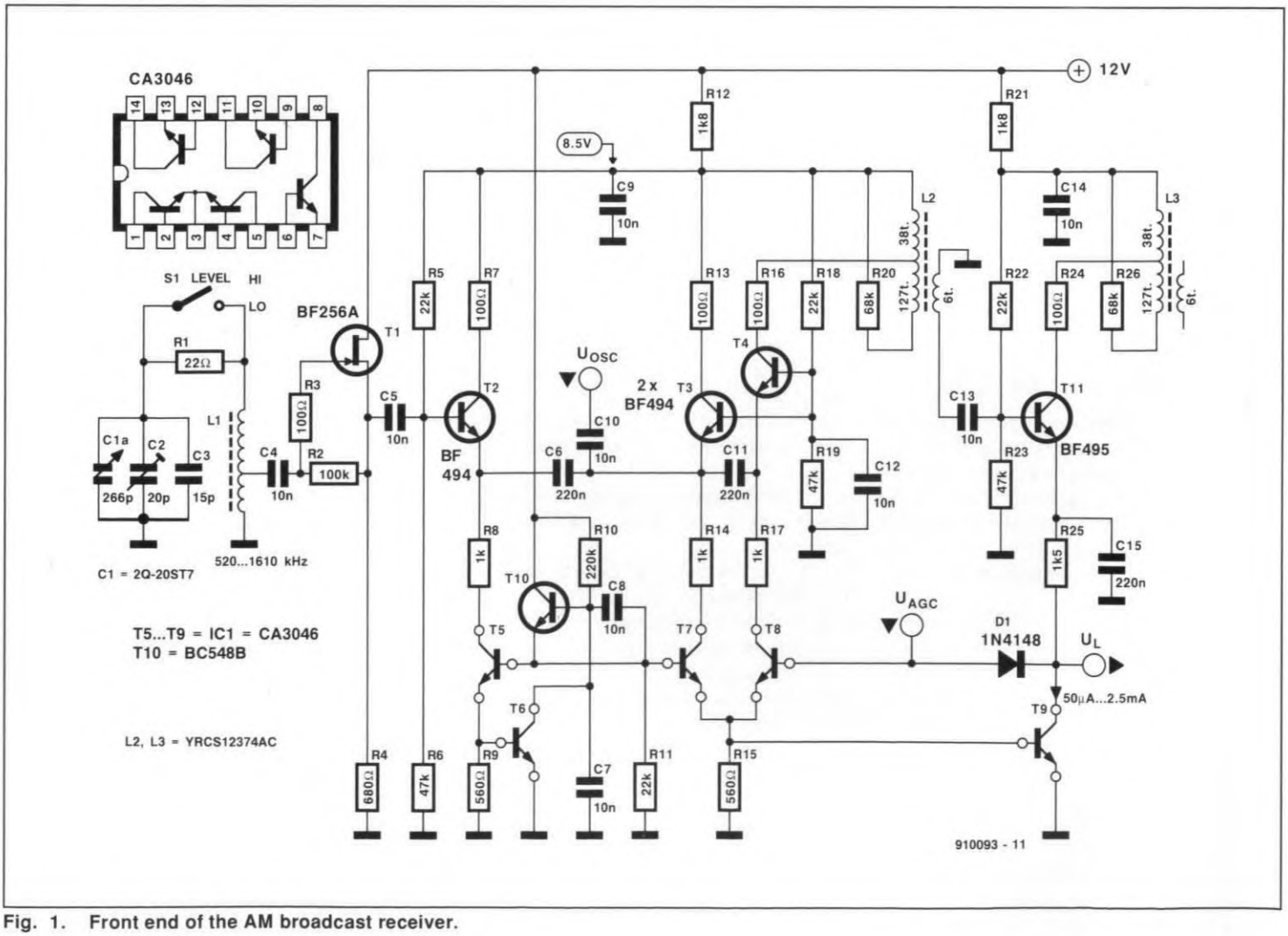


Fig. 1. Front end of the AM broadcast receiver.

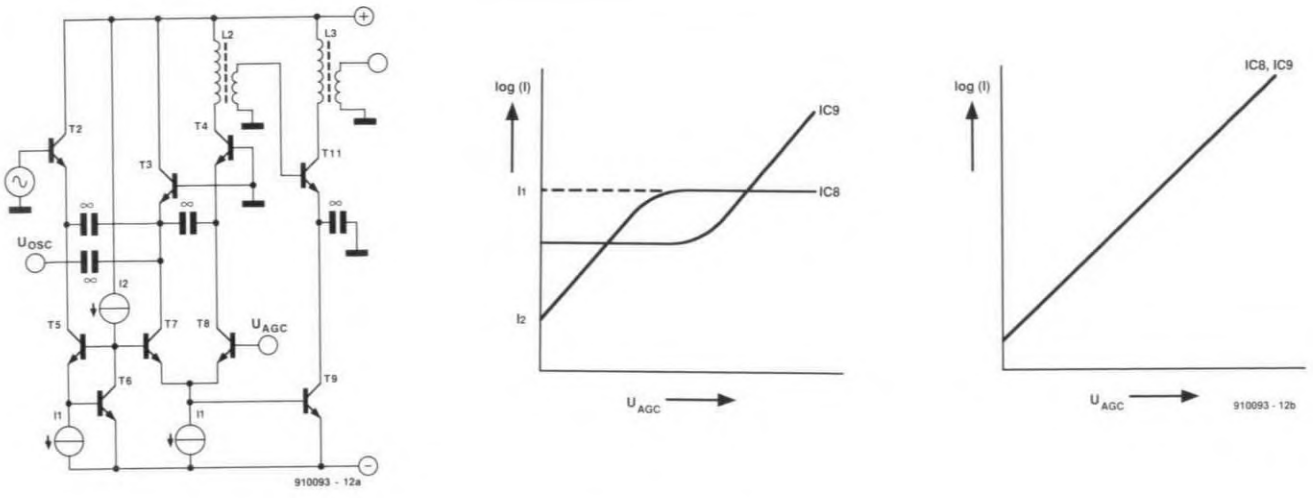


Fig. 2. Basic schematic and response of the of the automatic gain control (AGC).

able tuning to a different station.

AGC

Most AM receivers use a single transistor for the AGC (automatic gain control) function. By contrast, this receiver uses an elaborate two-stage AGC circuit to obtain a level control range of nearly 60 dB without overloading or excessive noise.

Active detector

Whereas most AM receivers use a single point-contact diode for the detector, the present design is based on an active detector stage, which has a distortion of less than 0.5% rather than the more usual 2% introduced by a diode.

Design background

The total circuit, excluding the power supply, uses 6 standard ICs and 15 transistors. Two ceramic filters are used. At first, it was decided to have only two IF stages, but in the end four were found necessary. The reason is as follows: in order to keep the audio output as flat as possible up to 4.5 kHz, the Q factor of each IF stage has to be reduced to a maximum of 25 by loading it. Inevitably, reducing the Q reduces the gain. Further gain reduction is caused by the insertion loss of the ceramic filter. So, in order to obtain a high input sensitivity (27 µV for 20 dB signal-to-noise ratio), four IF stages were found necessary.

For 4.5 kHz sidebands, the relative gain roll-off of the IF stages alone (including the ceramic filter) amounts to about 4 dB. The roll-off of the antenna tuned circuit is an-

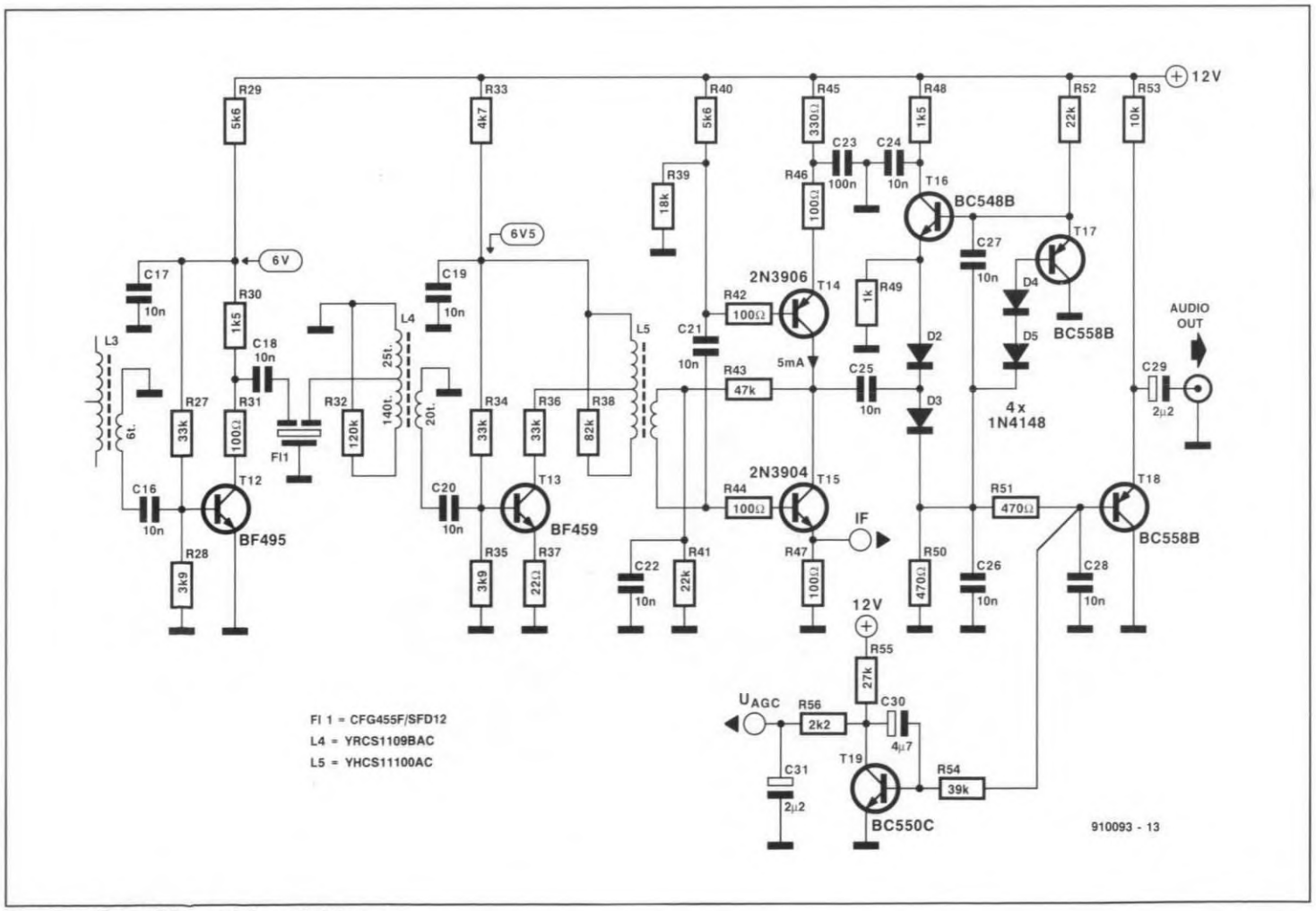


Fig. 3. IF amplifier and demodulator.

other 2 dB. This means that the total audio attenuation at 4.5 kHz is 6 dB, which is considered good for an AM receiver.

An important point must be made about the IFTs (intermediate frequency transformers) used in this circuit. These are basically 468-kHz types with built-in 180-pF capacitors. Though these IFTs can be tuned to 455 kHz, the stability suffers. For this reason, every IFT is fitted with an external 10-pF capacitor across the primary winding. Note that these 10-pF capacitors are not shown in the circuit diagram. The IFTs have not been used in the order suggested by the manufacturer, Toko. The reason that these particular 468-kHz types were selected is that there were no good 455-kHz types available from the author's supplier, Cirkit. Since the signal levels and impedances in the IF amplifier have been painstakingly ascertained, no attempts should be made to use types of IFTs other than the ones indicated, or the purpose of this project is defeated.

As regards distortion in amplifiers based on bipolar transistors, odd-order harmonic distortion is negligible for base-emitter voltage swings of less than 5 mV. At 3 mV, for example, the third-order harmonic distortion is only 0.06%. The second-order harmonic distortion, however, is nearly 3%. Fortunately, even-order products do not matter in the IF stages as they fall outside the pass-band. The upshot is that the simple single-ended common-emitter stage is perfectly acceptable. Not so in the mixer stage, however, where even-order intermodulation products are troublesome when the receiver is tuned to about 910 kHz (twice the IF), because they coincide with the image frequency. As the second-order products are considerable even for small input swings, a differential amplifier should be used for the mixer.

Contrary to popular belief, bipolar transistors are better suited to AGC circuits than FETs or MOSFETs. There are two reasons for this. First, the linearity of a FET or MOSFET deteriorates rapidly when its transconductance is made smaller than a tenth of its maximum value. Interestingly, precisely that condition would arise when a strong signal is received. Second, the exponential characteristic of a bipolar transistor results in an AGC loop gain which is independent of the carrier strength. For a FET, it can vary by a factor of 100, giving rise to wildly varying loop response times and lower cut-off frequencies.

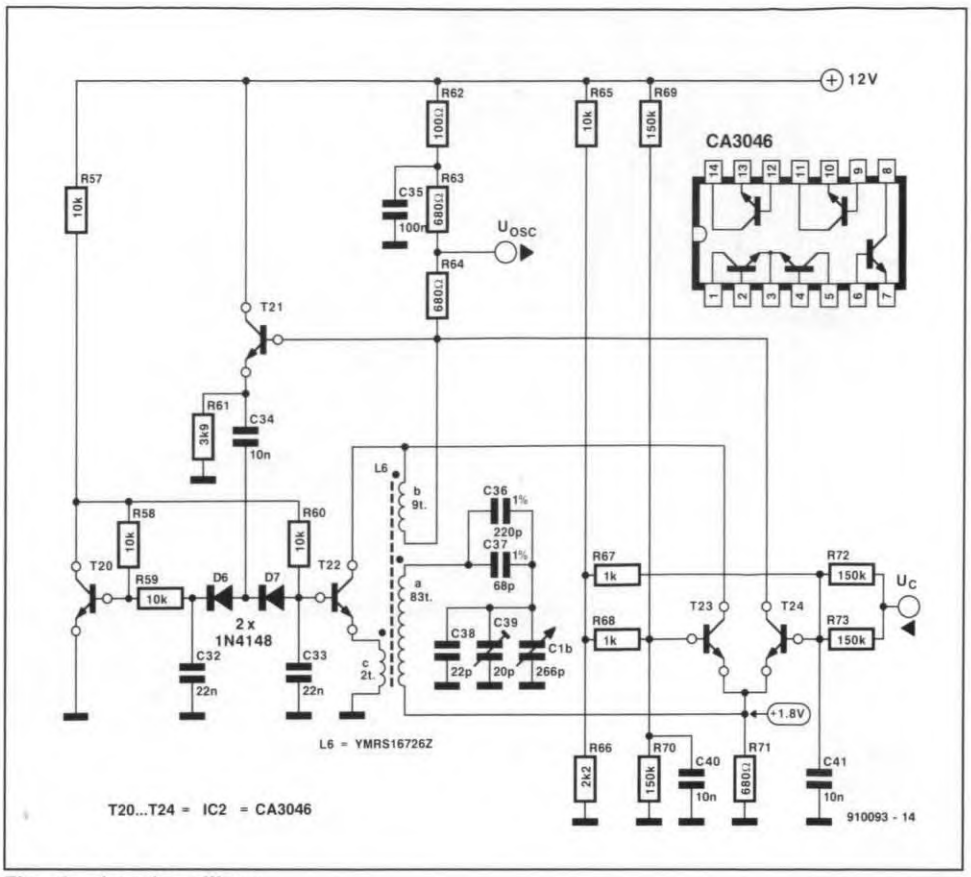


Fig. 4. Local oscillator.

The circuit

Input stage

Figure 1 shows the input stage of the receiver, which comprises the mixer, the first IF amplifier, and the AGC. The antenna tuned circuit comprises L1, C1a, C2 and C3, with the ferrite rod, L1, tapped at a fourth from earth. When the signal strength is high, a resistor can be switched in series with the tuned circuit for two purposes. First, it increases the bandwidth so that the high-frequency response is improved. Second, it decreases the signal level, and prevents distortion.

Transistor T1 is a buffer that prevents the mixer stage from loading the tuned circuit, especially at the higher end of the band. T1 operates at $I_d=I_{dss}$, which results in maximum gain and linearity.

The mixer stage is formed by T2, T3 and T4, with the local oscillator signal injected into their emitters. The mixer stage also functions as a part of the AGC. At low input levels, T4 conducts fully so that the maximum possible gain of the mixer stage is achieved. The operation of the AGC now depends mainly on the current through T11. When the input signal is fairly strong, T3 steals some of

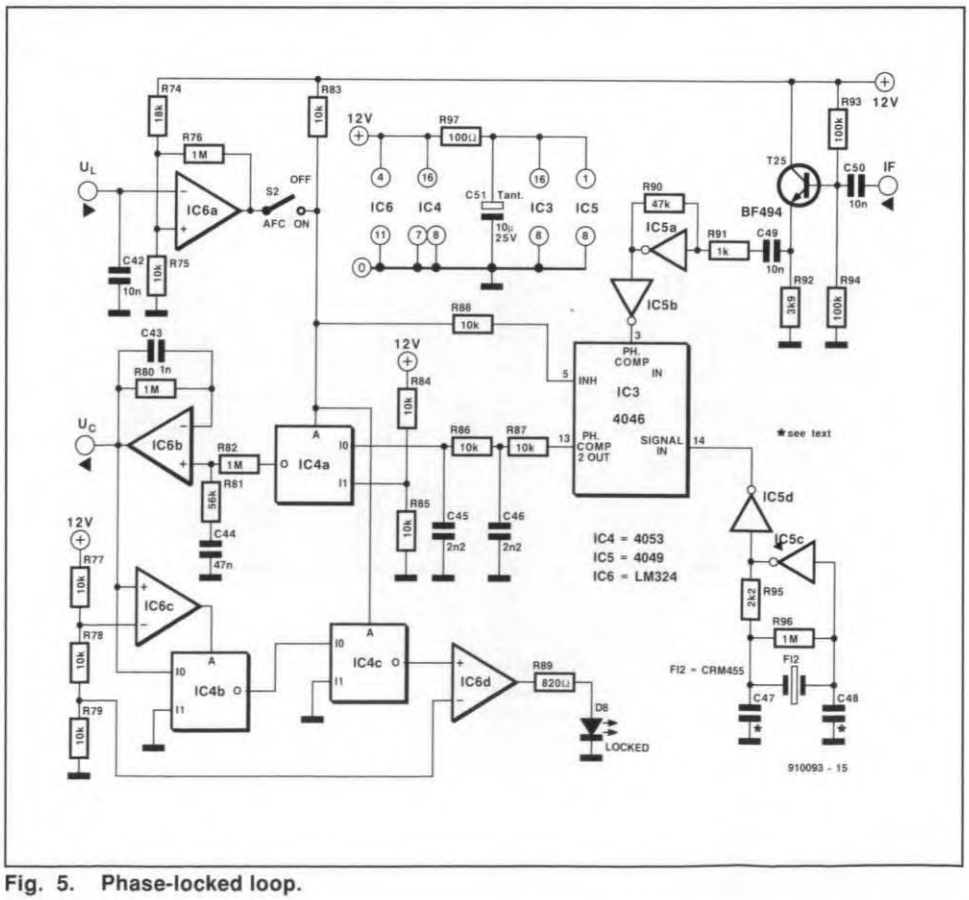


Fig. 5. Phase-locked loop.

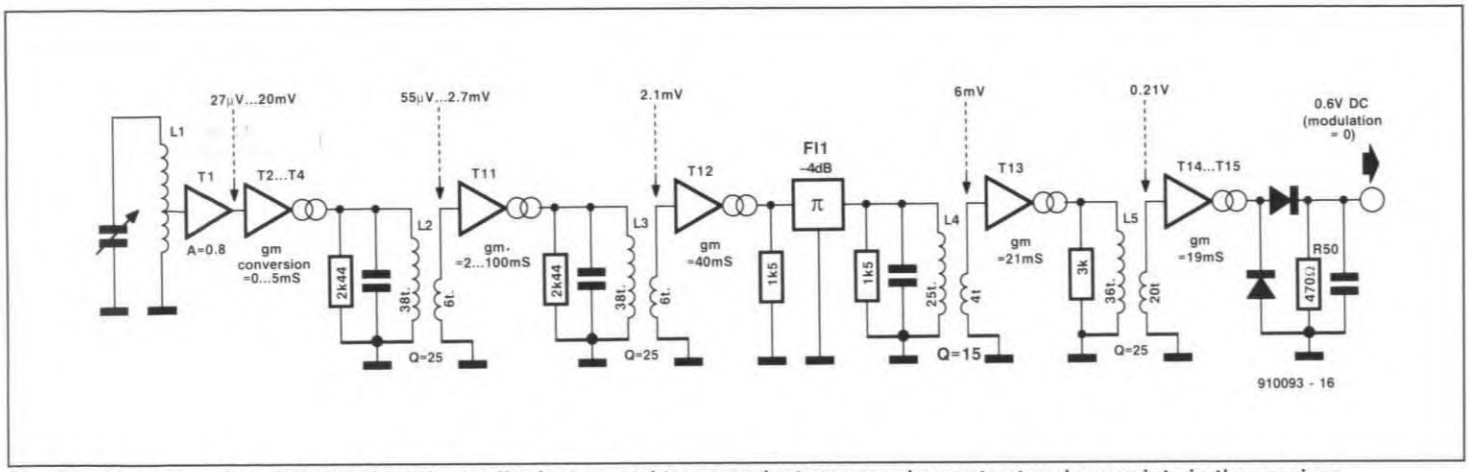


Fig. 6. Overview of crucial signal levels, quality factors and transconductance requirements at various points in the receiver.

the current through T4, so that the signal at the base is limited to about 3 mV. At still higher input levels, T4 starts to conduct less while T11 has a flat gain. Figure 2 shows a simplified model of the AGC stage. Here, current sources have been used instead of transistors. Currents I_1 and I_2 are 1 mA and 50 µA respectively. It can be shown that the total gain varies purely exponentially with the control voltage.

The two-stage AGC ensures that the first IF stage can not be overloaded by strong signals. This means that the signal handling capability of the receiver is increased by a factor equal to the mixer stage gain.

Transistor T10 is included because the base currents of T5 and T7 are comparable to the collector current of T6. Compensation is achieved with C8 and C7 which also ensure an AC path from the bases of T5 and T7 to ground. To improve stability, resistors R8, R14 and R17 provide RF isolation for IC1. The 100-Ω resistors likewise prevent spurious VHF oscillations of the transistors.

IF amplifier and AM detector

The circuit diagram of the second, third and fourth IF stage, and the active detector, is shown in Fig. 3. The input and output of the ceramic filter are matched to 1.5 kΩ.

The active AM detector is formed by T14-T17. T14 and T15 form a voltage-to-current converter with the collectors of the transistors at a relatively high impedance. A push-pull rectifier based on D2-D3 is used. The rectified RF current that flows through D3 develops a voltage across R50, which is smoothed by C26. The circuit around T16, T17, D4 and D5 provides a quiescent current of about 1 µA through D2 and D3 to keep the voltage swing at their junction to a minimum. The detected audio signal across R50 is filtered by R51 and C28, and buffered by T18 before it is taken to the output of the receiver (to reduce traces of IF to an absolute minimum, one more ladder section consisting of a 470-Ω resistor and a 10-nF capacitor can be added before T18).

The AGC feedback is also taken from this point. The instantaneous carrier level is integrated by T19, and fed to the base of T8. Excluding T19, the gain of the AGC from the base of T8 to the base of T18 is V_{be}/V_t, or about 23. The lower cut-off frequency of the audio signal is $23/(2\pi R_{54}C_{30})$, or 20 Hz. As mentioned earlier, this cut-off frequency is virtually independent of the carrier strength.

Local oscillator

Figure 4 shows the local oscillator based on a single CA3046 transistor array. The oscillator is basically a Hartley type, although a portion of the current through L6a is also fed to L6b by T23. This causes the effective inductance of L6 to change by a small amount. The ratio of the currents through T23 and T24 varies between 3:1 and 1:3, corresponding to V_c values of 4 V and 8 V respectively. The total frequency variation achieved is about 2.5%. At the lowest oscillator frequency, 975 kHz, the variation is about 25 kHz — more than enough to correct tracking deviations between the antenna and oscillator sections. When the AFC is disabled, V_c is held at half the supply voltage to minimize frequency drift. In this case, T23 and T24 conduct equally.

T22 is the positive feedback transistor whose d.c. setting determines the strength of the oscillations. T20 and T21 are used to control the amplitude to about 1 mA(0.6/R64) by reducing the d.c. voltage at the base of T22 when D6 and D7 start conducting.

For optimum tracking of the antenna and oscillator sections, the parallel combination of C36 and C37 should be equal to 0.978 times the total antenna tuning capacitance at the lowest frequency. This works out at 285 pF. Capacitors C36 and C37 should be 1% types. When C2 and C39, the trimmers associated with the tuning capacitor, are properly adjusted, a worst-case deviation of smaller than 5.5 kHz is possible.

PLL section

The zero crossings of the IF output signal are shaped by N1 and N2 and fed to one phase comparator input of the 4046 PLL. The other input is supplied with the reference frequency of 455 kHz from a ceramic resonator oscillator based on N3. The output of the phase comparator is fed to a ladder filter to remove traces of the product IF/4. When the AFC is switched on, the 'A' inputs of IC4a and IC4c are made logic 0, provided a sufficient carrier strength exists (for this, the current through T11 has to be smaller than 1 mA). The output of the PLL is fed to the loop filter formed by R81, R82 and C44. The AGC time constant is about 0.05 s while the damping factor is about unity.

If V_c is between 4 V and 8 V when the AFC is switched on, D8 lights to indicate that the PLL is locked. If the AFC is off, or if the carrier is too weak, or if V_c exceeds its range, D8 is off. When the AFC is not on, IC4a selects the resistive divider, and V_c becomes 6 V. At the same time, IC3 is disabled.

Signal levels

Figure 6 shows the signal levels at various points in the RF-IF chain. For the IFTs with 190 pF total capacitance, the parallel loading corresponding to a Q factor of 25 is 46 kΩ. Since the IFTs have an inherent Q of 90, additional loading is necessary to reduce the Q factor to this value. With the transistor types used, the base-emitter swings of all stages, except the mixer, have been kept in the 2 to 3 mV range as this is the best compromise between noise and distortion. Note, however, that even the mixer can handle up to 13 mV of RF, and still produce less than 1% distortion. At higher signal levels, R1 may be switched on. ■

THE DIGITAL COMPACT CASSETTE

WILL IT SUCCEED?

910083

based on an original article by P. van Willenswaard

Just when consumers were getting reconciled with the idea that digital audio tape (DAT) is virtually dead and started to buy more CDs, the two DAT protagonists, Sony and Philips, decided to confuse the market afresh by each announcing another new system: the Mini Disc (MD) and the Digital Compact Cassette (DCC) respectively. We briefly explain the reasons behind these new systems and have a closer look at the digital compact cassette technology.

WITH the demise of the digital audio tape (DAT) system—killed by the music industry which believed (probably rightly) that a system that enables the making of perfect copies from copies would cost them thousands of millions of dollars—audio equipment manufacturers began to reformulate their development programmes with the aim of getting a larger slice of tomorrow's market.

They found that cassettes are bought primarily for portable or mobile equipment; CDs and LPs almost exclusively for use at home. Moreover, three times as many cassettes (recorded and blank) as CDs and long-playing records (LPs) are sold world-wide. Because of their higher prices, CDs and LPs bring in a much larger revenue, of course. See Fig 1. This schism is further illustrated by considering the apportioning of tape decks over the various types of audio equipment. About 12 per cent are used in hi-fi separates or integrated systems; the remainder, that is, almost 90 per cent, are used in car radio/cassette players, walkman equipment and other portable sets—see Fig. 2.

Annual sales of compact cassettes amount to a staggering 2,600 million

The realization that the bulk of the demand for audio equipment is for portable/mobile units form the basis of Sony's and Philips' development programs.

Sony, still smarting from the collapse of the DAT market on which they had pinned much hope for the future, decided to get away from cassettes for the time being and have another look at their tremendously successful CD system (CDs today account for about half the £13,000 million – $22,000 million – world market for recorded music).

Knowing that CDs cannot compete with the cassette in the popular mobile/portable markets, Sony recently unveiled Mini Disc (MD). The MD is considerably smaller than the Compact Disc, but it has the tremendous advantage over its big brother that it is recordable and erasable. Moreover, the system is immune to shocks and jolts, which means that it can be used on the move, that is, in portable and mobile (car) units, with impunity.

Philips has gone back to its earlier invention, the compact cassette, but now in digital format: the Digital Compact Cassette (DCC) system. Since a large part of the 2,600 million cassettes sold every year are used in fairly inexpensive equipment, Philips reasoned that, to stand a chance of cornering a sizeable part of that market, they had to offer something that works with relative inexpensive tape decks and tape. Even better if the new system would also accept standard compact cassettes.

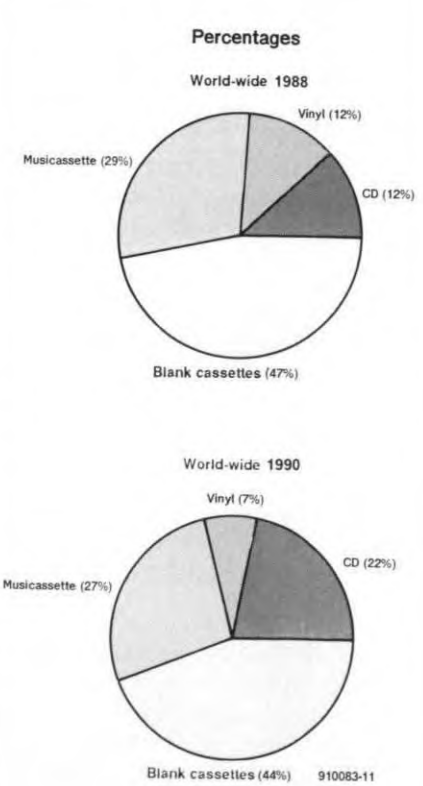
Both systems are promised (threatened?) to be on the market in late 1992. The success, or otherwise, of the new systems is in the hands of confused consumers and a cautious music industry which will insist that both systems are fitted with appropriate circuitry that will prevent copying from copies. It is still their, and many others', opinion that home taping is nothing but theft.

DCC requirements

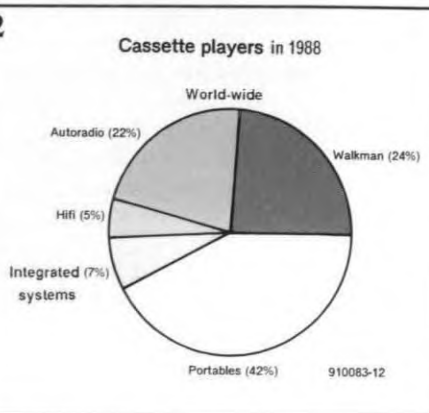
Recording 16-bit audio samples at a rate of 44 100 times a second in two channels requires 1.4 Mb/s; if space for encoding and error correction is included, that figure should be doubled. This kind of high-density recording is impossible on cheap tape. Ways have, therefore, to be found to make recording with fewer data possible.

At the same time, the choice of inexpensive, slightly modified compact cassette tape decks means a considerable reduction in development cost and time. The most frequently heard complaint about the current cassette system is the varying quality of sound reproduction: there is loss of high frequencies

1



2



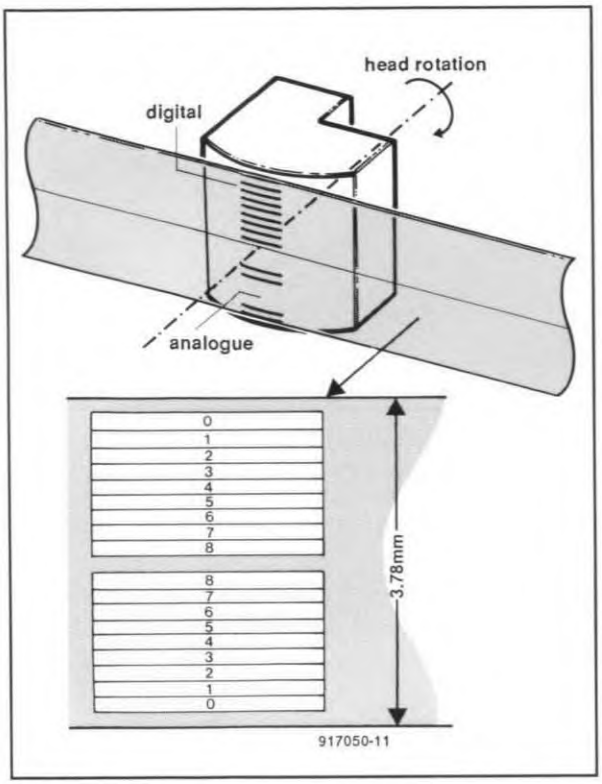


Fig. 3. Division of the tape into nine digital tracks; at the end of the tape, head and direction reverse and the lower half is written to or read from. The head also has two slits for the playing of analogue recordings.

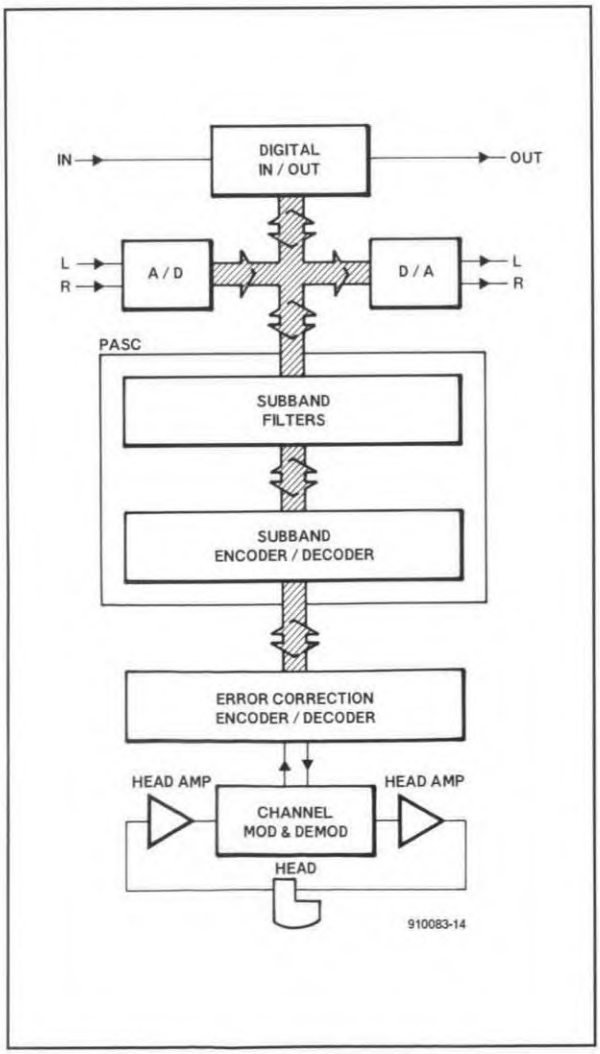


Fig. 4. Block schematic of a DCC player.

owing to misaligned or dirty heads; the sound quality is proportional to the quality of the tape transport (causing fading at high frequencies and wow and flutter) and to minor damage of the tape.

A digital recording, if adequate error correction is provided, is fundamentally immune to such imperfections. A digital compact cassette would, therefore, in many instances give an appreciable improvement in sound reproduction, which would make it immediately attractive to large numbers of potential users. This presupposes, of course, that production costs are low enough to make the sales price acceptable at the lower end of the market. This is quite feasible, because the tape decks of the DCC player are not very different from those of current cassette players, while the cost of the electronics would become low enough if manufacturing quantities were large. There are an estimated 200 million cassette players in the world; even replacement of 5–10 per cent of these would ensure success.

Reverting to the question of audio quality: how is good quality to be ensured if the audio signal is stored digitally in a much smaller space than 1.4 Mb/s? The tape width of a compact cassette is 3.78 mm. The vertical tolerance of transporting the tape across the head is about 0.05 mm. This means that the DCC remains within that tolerance if half the tape width is divided into nine tracks, each 185 µm wide, of which 70 µm is used for reading (play-back) and the remainder for writing (recording)—see Fig. 3. Eight of the nine tracks are allocated to audio; the ninth, to system information.

It has been found that at the standard tape speed of cassette players (4.76 cm/s), about 100 kb/s per track can be saved reliably. There is thus about 800 kb/s space for the audio signal; this has been standardized at 768 kb/s. Half of that, however, is used for the Reed-Solomon interleaving (in aid of effective error correction), synchronization bits, and a 10:8 modulation mode, which means that only 384 kb/s remain for the audio information proper. That is about a quarter of the earlier established 1.4 Mb/s; in other words, at a sampling frequency of 44.1 kHz, there would only be space for a four-bit amplitude description per channel. That is, of course, quite unacceptable. Philips has, therefore, developed precision adaptive sub-band coding—PASC.

Precision adaptive sub-band coding

PASC is not, as one might suppose, a highly developed form of data compression as used in computers or data links. Whereas pulse-code modulation (PCM) systems give a total and perfect (within the constraints of 16-bit linear encoding) digital representation of the sound that reaches a microphone, PASC uses a perceptive approach: it takes human hearing as reference. PASC analyses the signal and computes what may be audible and what not, and stores only what is in a very efficient notation. Subsequently, the resulting digital information is reorganized in such a manner that the available space is reallocated optimally.

Only what may be audible is stored and encoded

The way a signal is processed in DCC is shown schematically in Fig. 4. First, the a.f. range is divided into 32 sub-bands: the reason for this will become clear later. The division is linear, not logarithmic as in human hearing: each sub-band is, therefore, about 600 Hz wide. In digital signal processing (DSP), it has been possible for some time to divide the a.f. band into sub-ranges that can later be recombined without any errors or deterioration as far as frequency and phase are concerned. Philips is confident of being able to put this entire process on to one chip in the near future.

Threshold and masking of hearing

The next step occurs in every sub-band. The minimum intensity a sound requires in order to be heard depends on its frequency—see Fig. 5. This means that in each sub-band a threshold has to be determined; any information below that limit is ignored.

Another aspect is masking: it is well known that there are circumstances in which a strong signal will mask, that is, make inaudible, a weaker signal at a frequency not far removed from its own—see Fig. 6. When someone whispers something to you in a quiet street, you can understand it perfectly. If, however, just then a noisy lorry passes by, you have to be a lip-reader to understand.

Another form of masking occurs in time. If a weak sound is preceded a few milliseconds earlier by a strong sound, we cannot hear the weak sound. Even the opposite can happen: a weak sound emitted a few milliseconds before a strong one may, in certain circumstances, mask the strong sound. It all depends on the difference in intensity and the separation in time.

The PASC processor analyses the signal in each of the sub-bands and constantly adapts

the various thresholds to the actual signal. The thresholds are, therefore, not static, but follow the signal content. This results in higher efficiency, because only the part of the signal above the threshold, i.e., that which is audible, is stored and encoded.

Mantissa and index

The amplitude of the part of the signal that remains above the thresholds must be quantized. Whereas compact disc technology uses 16-bit linear quantization, PASC uses a floating-point technique. This works in a manner similar to the presentation of a large number on a pocket calculator: with a mantissa and an index (exponent). The mantissa of most pocket calculators is normally eight digits long and represents the value of the number. This is followed by the exponent, normally two digits (positive or negative), which weights the numbers.

In PASC, the mantissa is 15 bits long, but the processor may decide to work with a shorter one if not all the 15 bits appear necessary. This may happen, for instance, when the signal is just above the noise level or threshold of hearing. The mantissa thus determines the resolution of the system, while the index shifts the area over which the mantissa operates up and down the dynamic range.

The index is not linear over the entire dynamic range, but reflects the non-linear action of the ear. After all, the ear does not differentiate between sounds at 118 dB and 120 dB: both are too loud. At the other end of the range, however, 2 dB steps would be much too coarse, because the ear is highly sensitive to subtle differences when listening to weak sounds. Defining the amplitude in this flexible manner gives a further increase in efficiency compared with the linear encoding used in PCM.

If PASC fulfils its early promise, it will mean a revolution in digital audio

The maximum length of the mantissa of 15 bits is reflected in DCC by what Philips calls the (theoretical) THD+N figure of –92 dB, which is 1 bit, or 6 dB, above the –98 dB of CD technology. But that technology is not flexible, so that the THD+N figure is always referred to the same absolute level: a –20 dB signal will, therefore, have its THD+N figure only 78 dB lower. Because of the floating-point amplitude description, in DCC the THD+N distance of 92 dB will, however, be maintained until the noise threshold of the system is reached. Philips states a dynamic range of –108 dB, which is 10 dB better than that of CD.

All internal calculations are carried out with 24-bit wide words. The associated noise level is –146 dB, but Philips claims that the resolution made possible by the mantissa and index system will handsomely exceed that level. The real problems arise not in this part of the system, but in the analogue-to-digital (A-D) converter that needs to precede the PASC. The crux of the matter is that even the best state-of-the-art converters cannot provide more than 18–19 bits of reliable a.f. information.

Reallocation

One of the nicer properties of music is that the waveform varies constantly. This means that not all available space in the 32 sub-bands is in use all the time: that in a number of them will, in fact, be empty (above the threshold, that is). A certain space is reserved in the ultimate code that is written on to the tape for each of the sub-bands. A sub-band that is temporarily partially, or wholly, empty would not be very efficient. When the recording code is being composed, PASC moves information from full sub-bands to partially, or wholly, empty ones (and provides it with an address code to ensure that during play-back the signal is reconstructed correctly). The process of moving the information is called reallocation.

The gain in efficiency resulting from this method of signal processing is surprising: from 16 bits per channel for linear PCM to an *average* 4 bits per channel for PASC. *This means a gain of a factor 4 in data content.* PASC does not work with two channels, but with a stereo signal (the correct encoding of stereo signals proved to be quite difficult; stability gave particular problems). Just as there is no sharp division between left-hand and right-hand, there is none between one sample of the signal and the next: PASC works with small groups of samples. On average, the code for each stereo sample contains eight bits of space that are written on to the tape. These are not bits of waveform: the 8-bit code generated by PASC is much more intelligent than that. The 8-bit words should be considered as pieces of a continuously developing jigsaw puzzle, which are resolved by the decoder during playback.

The decoder need not be high-tech, because the intelligence is stored in the code itself, since this contains all necessary keys for encoding and reallocation. In other words, the keys have become part of the 8-bit codes.

Storing the keys in ROM near the decoder would seriously limit the system (perhaps even make it unworkable), because each sound is different from others and thus requires its own unique PASC code. The keys in the 8-bit code control the operation of the decoder: what is the reallocation information; which mantissa and index were used; and so on. The code is the brain: the decoder only carries out orders.

The decoder is, therefore, suitable not only for working with DCC codes, but with all codes within the same family, such as Digital Audio Broadcasting (DAB) codes (in the framework of the European Eureka project, Philips partakes in DAB experiments), or even simple 2-bit codes of limited audio quality. Moreover, the code may change from one to another instantly because the decoder does not have to switch over.

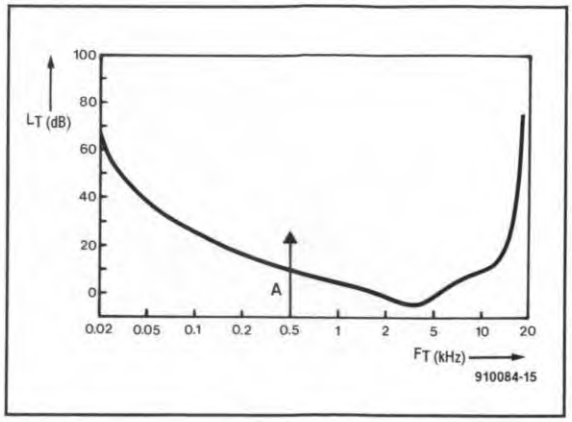


Fig. 5. Characteristic to indicate the threshold of hearing, both as regards loudness L (vertical scale) and frequency F (horizontal scale). Sounds below the curve can be omitted. Sound A is just perceptible.

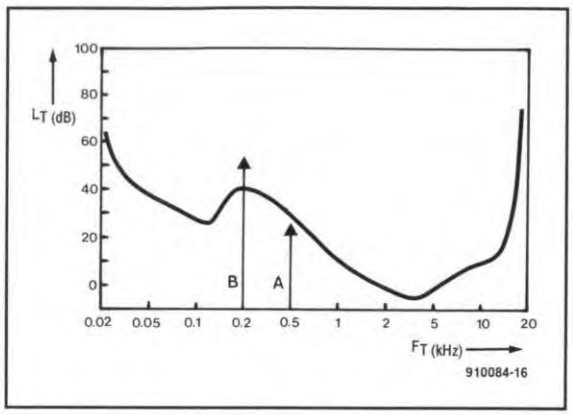


Fig. 6. Strong signal B raises the threshold of hearing locally, as it were, so that sound A is no longer audible: B masks A.

Breakthrough in digital audio

If PASC fulfils its promise, it will mean a revolution in digital audio. Storage space for the large amount of data generated by digital audio has always been a problem. Until now, storage with simple means, a longer playback time, a substantially higher sampling rate than the current 16 bits of amplitude, and so on, were not possible in practice. It may also well be, because a signal after being processed in PASC is technically less complex, that the analogue electronics in the play-back chain can become less complex.

As far as the sound quality of DCC is concerned, it is too early to make a definitive judgement. However, first impressions gained during a recent demonstration of the system to a number of international journalists were good. When amplifiers of reasonable (not high-end) quality were used, there was no perceptible difference between a CD and a DCC. ■

APPLICATION NOTES

The contents of this article are based on information obtained from manufacturers in the electrical and electronics industry and do not imply practical experience by *Elektor Electronics* or its consultants.

SWITCH-MODE VOLTAGE REGULATORS LM2575/LM2577 (National Semiconductor)

THE recently introduced LM2575 and LM2577 switch-mode voltage regulators in the Simple Switcher series from National Semiconductor have a wide voltage range, and are remarkably easy to get going with only a handful of external components. The devices in the LM2575 series are step-down regulators, and those in the LM2577 series step-up regulators. Offering an efficiency of over 80%, these ICs are a good alternative to the ubiquitous 1-A regulators in the 78xx series.

The new step-up and step-down converters are based on switch-mode voltage regulation, which gives them a higher efficiency (70-90%) than linear voltage regulators (30-60%). By virtue of the on-chip power stages, the external component count remains low — basically, only a choke, a power diode and a few decoupling capacitors are required. The adjustable regulators in addition require a voltage divider that determines the output voltage. All devices in the LM2575 and the LM2577 series are available in a 5-pin TO220 case (LM257xT), and a 4-pin TO3 case (LM257xK). The case of these devices is always connected to ground. Fixed voltage as well as variable voltage devices are available — see Table 1. The LM1575 and LM1577 should be used in heavy-duty applications where an extended temperature range (T_j up to 150°C) is required. The main technical data of the regulator family are summarized in Table 2.

Since the switch-mode regulators are based on a fixed oscillator frequency with a variable duty factor, they do not require the usual minimum load current to operate correctly. The ICs feature an internal thermal overload protection which is actuated at T_j=125 °C, and an output current limiting circuit. A further advantage of these new ICs is that they are inexpensive and uncritical of the choke type used.

Table 1. Overview of devices in LM2575/LM2577 Simple Switcher family

Type	Output voltage	TO220 case	TO3 case
Step-down	5 V	LM2575T-5.0	LM2575K-5.0
	12 V	LM2575T-12	LM2575K-12
	15 V	LM2575T-15	LM2575K-15
	adjustable	LM2575T-ADJ	LM2575K-ADJ
Step-up	12 V	LM2577T-12	LM2577K-12
	15 V	LM2577T-15	LM2577K-15
	adjustable	LM2577T-ADJ	LM2577K-ADJ

Step-down regulator LM2575

The operating principle of the so-called buck (step-down) regulator is illustrated in Fig. 1. Switches S1 and S2 close in alternate fashion at a certain rate. When S1 is closed, the self-inductance of the charge choke, *L*, causes a slowly rising current to flow. This current causes energy to be stored in the form of a magnetic field in the core of the choke. When the switches toggle (S1 is opened, and S2 is closed), the self-inductance first causes a current I_L to be fed through *L*. As a result, the magnetic energy stored in the choke, and with it I_L, decreases gradually. This process is repeated as S1 and S2 close and open alternately. Capacitor *C* smooths the output current, I_{OUT}, because it is charged as I_L rises, and discharged by the load as I_L drops again.

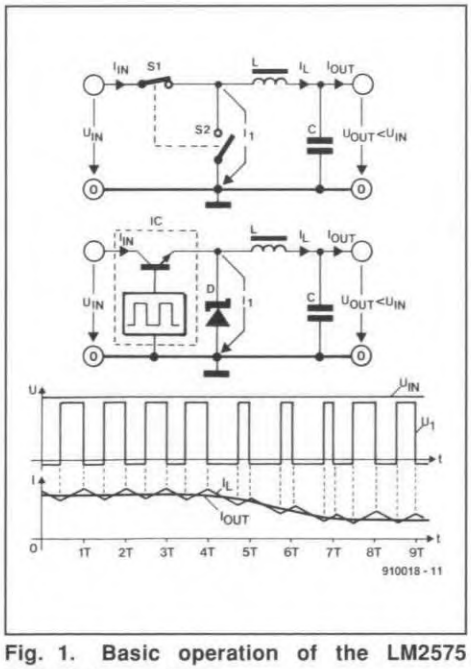


Fig. 1. Basic operation of the LM2575 step-down regulator.

By controlling the mark-space ratio of individual pulses, I_{OUT} can be set such that the desired output voltage U_{OUT} appears across the load. The drawing in Fig. 1 illustrates this process when the output current drops. Switch S1 is actually a power transistor controlled by a rectangular wave generator (Fig. 1b). The other switch, S2, is formed by a diode that enables I_L to continue flowing when the power transistor is switched off. To prevent losses as a result of the relatively high currents, the diode must have a fast switching characteristic. This requirement is met by the use of a Schottky power diode.

The internal circuit of the step-down regulator Type LM2575 is shown in Fig. 2. The power transistor is an n-p-n type that switches the input voltage, U_{IN}, to the output pin marked OUT. The transistor is driven by

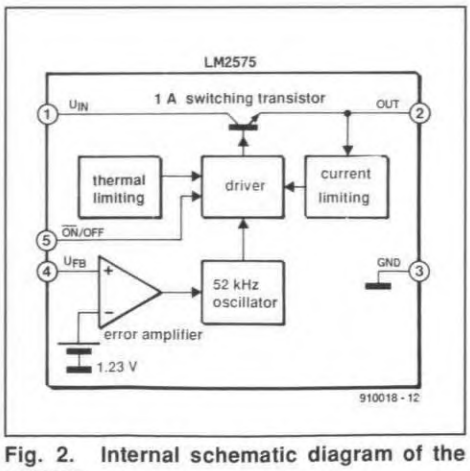


Fig. 2. Internal schematic diagram of the LM2575.

an oscillator with a fixed output frequency of 52 kHz. The pulsewidth of the oscillator output signal is controlled by an error amplifier, which compares the voltage supplied by a potential divider connected across the output with a fixed voltage supplied by an internal 1.23-V reference.

Apart from the protection circuits for the power transistor (junction temperature; output overload), there is a reset generator, which is also used to switch off the output voltage via the TTL-compatible ON/OFF input of the regulator.

A typical application circuit of the LM2575 is shown in Fig. 3. Here, the adjustable device Type LM2575T-ADJ is used. The output voltage is set with the aid of an external circuit. When a fixed-voltage device is used in this circuit, pin 4 must be connected to the output. A 3-A Schottky diode Type 1N5821 is used as the previously mentioned power diode. Components R2, D2 and C5 supply a voltage of +5 V for the ON/OFF control input, whose function may be tested with the aid of jumper J1. When the ON/OFF function is not required, the 5-V supply and the jumper may be omitted, and pin 5 connected permanently to ground. As indicated in Fig. 3, the circuit must have a single earthing point. This arrangement prevents voltage fluctuations as a result of current peaks upsetting the control loop, and should be taken into account when a PCB is designed for the supply. Spurious pulses are suppressed by electrolytic capacitors C2 and C3 at the input and output of the regulator. In some cases, the a.c. resistance of C3 may be too high, when the capacitor is best replaced by a number of capacitors in parallel with the same total capacitance (small electrolytic capacitors have a lower parasitic inductance than large ones). The value of C3 must always be greater than 330 µF. The datasheet advises against the use of tantalum capacitors, which cause instability of the control loop. When the output voltage is higher than 25 V, C3 must be replaced by a type with a working voltage of 40 V or more. Capacitors C1 and C4 form a virtual short-circuit for the fast pulse edges, and so secure the necessary RF suppression.

Table 2. LM2575/LM2577 main technical data

Parameter		LM2575	LM2577
Input voltage	U_{IN}	3.5 - 35 V	3.5 - 40 V
Output voltage	U_{OUT}	1.23 - 30 V (always < V_{IN})	max. 60 V (always > V_{IN})
Output current	I_{OUT}	1 A	2 A (depends on V_{OUT})
Current limit (typ.)	I_{CL}	2.2 A	4.3 A
Switch saturation voltage	U_{SAT}	0.9 V	0.5 V
Quiescent current	I_Q	10 mA	
Reference voltage	U_{FB}	1.23 V (±2%)	
Input current at VFB	I_{FB}	50 nA	100 nA
Oscillator frequency	f_{OSC}	52 kHz (±10%)	
Efficiency		80%	
Operating temperature		−400°C < T_j < +1250°C	
Protections		internal current limit thermal shutdown	
Special features		ON/OFF input undervoltage detector	Soft-start

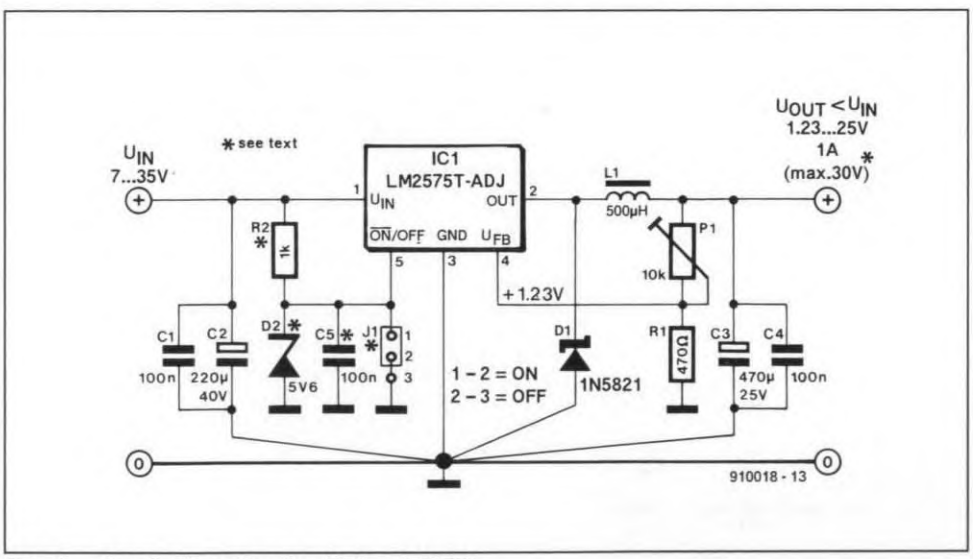


Fig. 3. Application circuit with the LM2575.

Table 3. Test data for application circuit in Fig. 3

	I_{OUT} (A)	P_{OUT} (W)	I_{IN} (A)	P_{IN} (W)	Efficiency	Ripple (mV)
12 V to 5 V	0.1	0.5	0.07	0.84	60%	1
	0.3	1.5	0.2	2.4	63%	1
	0.6	3.0	0.35	4.2	71%	1.2
20 V to 12 V	0.1	1.2	0.08	1.6	75%	2
	0.3	3.6	0.2	4.0	90%	2
	0.6	7.2	0.4	8.0	90%	2.2
	1.0	12.0	0.7	14.0	86%	2.5

The output voltage, U_{OUT}, of the regulator is simple to calculate. Potential divider P1-R1 reduces the desired output voltage to 1.23 V. The equation for the current, I, allows us to determine the values of P1 and R1:

$$I = 1.23\text{ V} / R1 = (U_{OUT} - 1.23\text{ V}) / P1$$

This allows the value of P1 to be determined when R1 is known:

$$P1 = R1\ (U_{OUT} - 1.23\text{ V}) / 1.23\text{ V}$$

$$P1 = R1\ (U_{OUT} / 1.23\text{ V} - 1)$$

The value of P1 should lie between 1 kΩ and 10 kΩ. Evidently, the operating principle of the regulator does not allow an output voltage that is higher than the input voltage. The results of a test on the circuit of Fig. 3 at two different settings are given in Table 3. Particular attention was paid to the behaviour of the LM2575 when the input voltage was increased slowly. The result: the set output voltage was maintained exactly (by contrast, some other regulators produce spurious output voltages higher than the set value when the input voltage drops below a certain minimum value). The measured values show an efficiency of between 60% and 90%, and an output voltage ripple of a few millivolts (measured in mV_{rms} with an a.c. millivoltmeter).

Step-up regulator LM2577

Before discussing the operation and application of the LM2577, it may be useful to recapitulate the principle of the step-up or boost voltage regulator. Figure 4 shows the basics. It is seen that the charge inductor, L, is connected ahead of switches S1 and S2 (compare Fig. 1) which are closed alternately. When S2

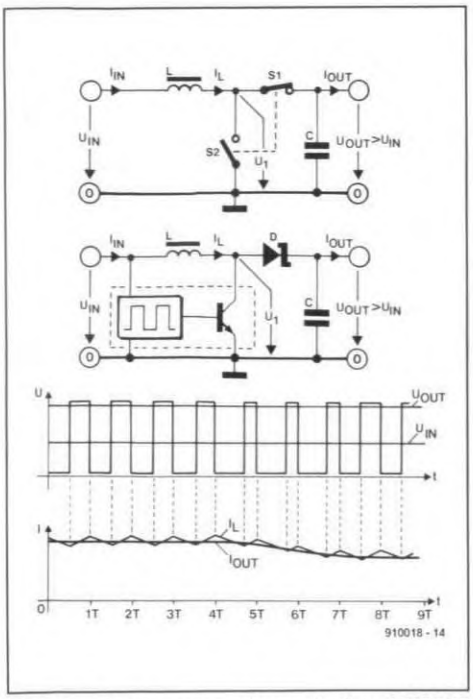


Fig. 4. Basic operation of the LM2577 step-up regulator.

is closed, the choke is connected to the full input voltage, U_{IN}. Because of the self-inductance, the current through L can only rise linearly, i.e., not pulse-like. As in the step-down regulator, magnetic energy is stored in the core of the choke. As soon as S2 is opened, and S1 is closed, the self-inductance the choke forces the current I_L to keep flowing in the same direction. The voltage developed by the self-inductance is polarized such that it is added to the input voltage. Hence, the output voltage is higher than the input voltage. Capacitor C smooths the output current, I_{OUT}. In the LM2577 (Fig. 4b), switch S2 is formed by a power transistor, and S1 by a fast switching diode. The diode is reverse biased when the transistor conducts. The voltage and current waveforms that occur in the circuit when the output current is reduced are shown in Fig. 4c.

Figure 5 gives the block schematic diagram of the LM2577. Like the LM2575, the LM2577 contains a 52-kHz pulse-width modulated oscillator driven by an error amplifier which obtains its control information from the voltage at the VFB input. In addition to thermal and current overload protection circuits, the LM2577 features a soft-start generator that prevents a high rush-in current through choke L. Another preventive function is that provided by the undervoltage limit circuit. When the input voltage drops below 2.9 V (typ.), the power stage is automatically switched off to prevent it conducting almost continuously. When the undervoltage detector is actuated, the output voltage drops to virtually the input voltage.

An application circuit of a step-up voltage converter based on the LM2577 is given in Fig. 6. Here, the adjustable device Type

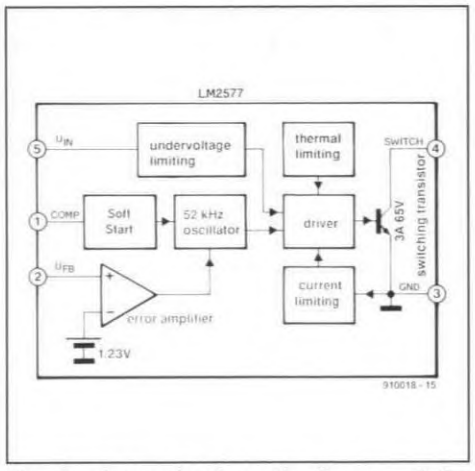


Fig. 5. Internal schematic diagram of the LM2577.

LM2577T-ADJ is used. The 1N5821 and the 500-µH choke are familiar from the step-down regulator (Fig. 3), as are the central ground point and the possible problems with a single electrolytic capacitor in position C3.

Since the same reference voltage of 1.23 V is used, the calculation of the output voltage is identical to that shown above for the LM2575. It should be noted, though, that the circuit in Fig. 6 can not supply output voltages lower than the input voltage. The maximum output current depends on the properties of choke L1, the ratio U_{OUT}/U_{IN}, and the maximum permissible current through the power transistor in the LM2577 (4.3 A typ.). The datasheet provides the following information on I_{OUT} and U_{OUT}:

$U_{OUT} < 60$ V
$U_{OUT} < 10\ U_{IN}$
$I_{OUT} < 2.1\ A\ U_{IN} / U_{OUT}$

Also, when the output voltage is higher than 30 V, the 1N5821 should be replaced by a type with a higher rated voltage.

The main results of tests carried out on the circuit of Fig. 6 are listed in Table 4. Clearly, the efficiency rises as the difference between the input and the output voltage becomes smaller. The ripple on the output voltage is higher than with the step-down regulator, but a mere 10 mV_{rms} or so will not be a problem for most applications. When experimenting with the LM2577, make sure that the voltage divider is always connected. Without a properly dimensioned and closed control loop, the IC may be destroyed easily. ■

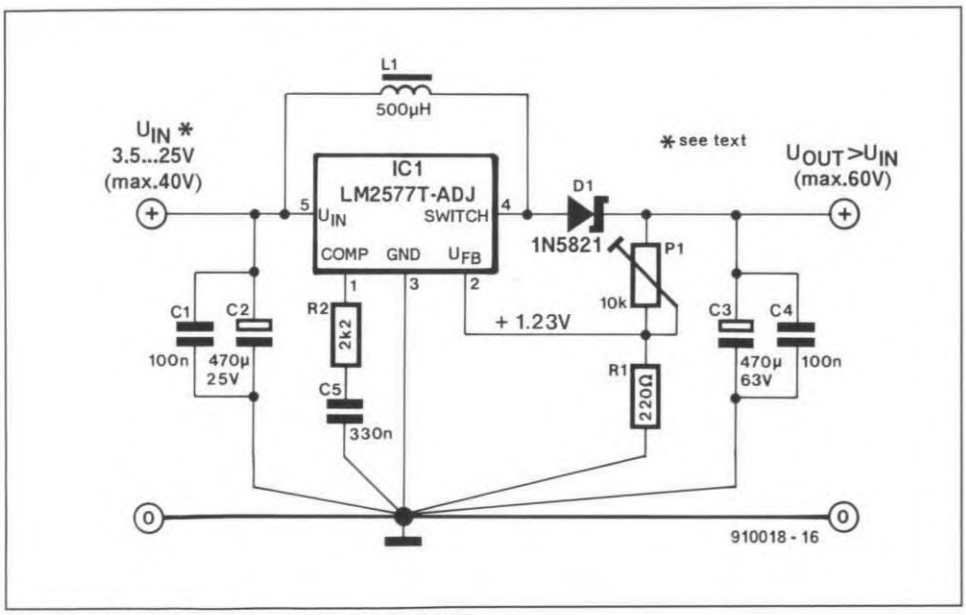


Fig. 6. Application circuit with the LM2577.

Table 4. Test data for application circuit in Fig. 6

	I_{OUT} (A)	P_{OUT} (W)	I_{IN} (A)	P_{IN} (W)	Efficiency	Ripple (mV)
5 V to 12 V	0.1	1.2	0.3	1.5	80%	4
	0.2	2.4	0.6	3.0	80%	8
	0.3	3.6	0.9	4.5	80%	15
12 V to 28 V	0.1	2.8	0.3	3.6	77%	6
	0.2	5.6	0.6	7.2	77%	10
	0.3	8.4	0.9	10.8	77%	15

Source:

Datasheets LM1575-ADJ / LM2575-ADJ — Simple Switcher Step-Down Regulator. National Semiconductor 1990.
Datasheets LM1577-ADJ / LM2577-ADJ — Simple Switcher Step-Up Regulator. National Semiconductor 1990.