Founded 1925
Incorporated
by Royal Charter 1961
To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering

Volume 51 No. 2 February 1981

The Radio and Electronic Engineer

The Journal of the Institution of Electronic and Radio Engineers

Legislating for Frequency Conservation

THE personal radio service, popularly known by its American name of 'Citizens Band' (CB), is currently a widely publicized technical controversy. Official British circles now refer to it as 'Open Channel' to distinguish the purpose of CB from the long established, and carefully regulated, amateur radio and private mobile radio. Nearly three years ago a Report by a Working Party of the National Electronics Council gave the following definition:

> 'A short range radio communication service available to private users (but not excluding the small business user) at an acceptably low cost and with the minimum of formality. The quality and reliability of the service and the probability of achieving the desired contact need not be as high as is required by emergency services or for security or major business communications. Ideally it would be introduced in such a way and with such characteristics that with little or no policing it did not cause unacceptable interference to any existing radio service or to other electronic equipment.'*

The Home Office response could not accept some points in the Report such as technical proposals for minimizing interference and observations on administration and social aspects. But the NEC definition was deemed a useful starting point for the Green Paper discussion document, 'Open Channel', published in September 1980. Since then illegal installations in the 27 MHz band have mushroomed and some estimates now put these in the order of 100,000. Such extensive use of this band has borne out the misgivings of both the NEC Report and the Green Paper.

Great argument has understandably followed the Home Office recommendation of 928 MHz as a future allocation. It is stressed throughout the Green Paper that the goal must be to avoid interference, whether harmonic or adjacent channel, and every available frequency channel was examined, and rejected, up to 928 MHz. Objections to 928 MHz come both from potential users and from manufacturers who doubt whether equipment can be made at the present time to a realistic price. The limited range—between 5 and 8 km according to the Home Office—is regarded as unacceptably low by the users. The Institution's Communications Committee, recognizing these and other technical difficulties, has sent to the government a professional radio engineering view urging re-examination of the criteria by which 225 and 450 MHz were considered unsuitable. An important consideration must be the compromise between acceptable frequency, required range and practical power.

The controversy continues in the national and technical press and some heated points have been made: 'Can British industry capture the potential market?' 'Could, or should, existing users be moved to accommodate CB?' 'Is biological damage a risk at 928 MHz?' 'What about environmental impact of fixed station arrays for 27 MHz?' Indeed, the cynic might remark that the famous electronics battles of the 50s and 60s—'a.m. versus f.m.', '405 versus 625 lines' and 'NTSC versus PAL'—were far, far tamer!

The rights, privileges and duties of the individual must apply to interference to existing services—domestic television viewers and mobile radio users pay licence fees, while aircraft and similar services must be safeguarded. On their part protagonists for CB put forward its potential for helping in emergencies in the same way as the radio amateur does.

A further statement expected from the Home Office may have appeared by the time this Journal is in print. Two factors however are incontrovertible: legislation must be sensible and enforceable; the radio frequency spectrum is finite and its use must be conserved like any other resource.

F.W.S.

* 'Citizens Band Radio', *Natl Electronics Rev.*, **14**, p. 46, May 1978; Home Office comments, p. 70, July 1978.

A New Service to Members

The Institution has concluded arrangements with MSL Engineering, part of MSL Group International Limited, under which MSL will extend its specialist recruitment activities in the electronics field. The new arrangement aims to improve the career development opportunities for IERE members at no cost to the Institution or to members who avail themselves of the service.

MSL Engineering was established in 1978 with the backing of the Institutions of Mechanical and of Production Engineers to assist companies seeking qualified engineers for managerial and technical posts. Even before MSL Engineering was established, about one-third of MSL's executive recruitment activity had been concerned with engineering industry.

MSL Engineering offers companies a comprehensive recruitment service—a major feature of which is the MSL Engineer File. To employers needing external professional recruitment assistance, it offers the prospect of incurring lower total recruitment costs.

The File will now be opened to members of the Institution. MSL will publicize the electronics component of the File by all means compatible with their membership of the Management Consultants Association.

The objective of the File is to develop a high quality alternative recruitment source in terms of the standards of the engineers whose details are recorded and also in terms of the accuracy of the information available on each candidate. To achieve this objective each application for File membership is vetted by a consultant against certain criteria such as age, qualifications and career progress. MSL maintains the right to determine who should be admitted to the File, and updates its candidate information at least every six months.

At the time of going to press, the Engineer File has a membership of 2000 qualified engineers. All hold at least an HNC qualification and about three-quarters hold a higher qualification. The majority of candidates are under 40 and almost all are at present in full-time employment. Whilst the File is concerned primarily with UK employers, increasingly an international dimension is being added to its activities, particularly in North America, the Middle East and Africa. Engineers currently resident abroad are eligible for File membership providing that they will be available for interview in the UK in the near future.

The principal benefits of the new scheme for members of the Institution are that their career details will be reviewed against every search assignment MSL Engineering obtains and that the ensuing relationship between File members and MSL will be conducted on a wholly professional basis.

File members will receive comprehensive information about jobs for which they are being considered. The provision of information at the recruitment stage is an important contribution in assessing the compatibility of prospective employer and employee.

MSL Engineering offers individual companies a highly flexible approach that can be geared to their particular needs. MSL is currently handling a wide range of assignments for different types and sizes of companies. Whilst the one-off recruitment exercise is common, the service is proving especially attractive to companies with multiple or continuing requirements for qualified engineers.

Members of the Institution who wish to receive details of the MSL Engineering File are asked to complete the reply card inserted in this issue or, alternatively, contact:

Information Section,
MSL Engineering,
52 Grosvenor Gardens,
London SW1W 0AW,
Telephone 01-730 0255

Letter to the Editor

From: **B. Priestley**, B.Sc., C.Eng., M.I.E.R.E.

Oscillator Design

Attending the symposium on Telemetric Studies of Vertebrates has finally triggered me into writing this letter about an appalling aspect of electronics.

I refer to the cavalier attitude of many so called engineers to the design of oscillators. At the symposium we heard of overtone crystal oscillators in production which jumped to another overtone or drifted off channel. Again I recently met a commercial microprocessor board where five out of five boards had to be modified because the oscillator only started half the time.

If our universities and colleges are unable to fit accurate information on oscillator *design* into their courses will they please stop telling the fable that an oscillator is simply an amplifier gone wrong! Even if the sole criterion is oscillation this isn't adequate; if amplitude or long or short term frequency stability is important it is hopeless.

Possibly *The Radio and Electronic Engineer* should help clarify the situation with an issue dedicated to a survey of the field? This would seem to be useful both to expose the dimensions of the problem and as a preliminary to any systematic classification.

B. PRIESTLEY

43 Raymond Road
Slough, Berkshire SL3 8LN
23rd November 1980.

UDC 621.391.812.7.029.64
Indexing terms: Electromagnetic waves, reflection, Microwave transmission

Dynamic and static reflection of microwaves from buildings

B. K. LEE, B.Sc.*

SUMMARY

This paper describes measurements of frequency spectrum and amplitude of 5 GHz signals reflected from large, generally flat, building faces. Two techniques were used, namely dynamic (i.e. fast moving transmitter source) and static, with ground-based equipment located near an aircraft runway. Although simple theoretical considerations had suggested that the reflection characteristics should be approximately specular, the results showed that there were both spectrum broadening and large rapid fluctuations of reflection amplitude over the specular zone, with significant scattering outside this region.

* *British Aerospace Dynamics Group, Bristol Division, P.O. Box No 5, Filton, Bristol BS12 7QW.*

1 Introduction

The investigation was designed to provide experimental evidence of the reflection characteristics of large buildings located near an aircraft runway, in the context of a microwave aircraft landing system operating at 5 GHz. A largely theoretical appraisal 1, 2 had concluded that at 5 GHz reflection multipath from large buildings could be approximately specular, based on considerations of the surface material of the building. In order to test this conclusion in a 'real life' situation an initial search was made to find buildings with flat surfaces which, as indicated in Reference 1 would be expected to produce specular reflection. The buildings chosen were located at RAF Fairford and are described in Section 3.

The basic measurement configuration is illustrated in Fig. 1; ground reflected rays (not shown) are also involved. The equipment employed is described in Section 2.1. The dynamic measurements (Section 2.2) were designed to simulate operation of the microwave landing system, with the aim of evaluating the spectrum of the multipath signal. With specular reflection a line spectrum would be obtained. However, initial investigation showed significant spectrum corruption, and additional static (i.e. reflection amplitude) measurements were made, as Section 2.3.

2 Experimental Technique

2.1 Equipment

The transmitter was installed in a vehicle which was moved during the measurements, and comprised a 5 GHz crystal controlled c.w. source with an adjustable output power, usually set to 1 W. The transmitter antenna was a waveguide horn with the following characteristics:

3 dB beamwidth:	$\pm 25°$ azimuth
	$\pm 10°$ elevation
Polarization:	vertical

This horn was mounted at the rear of the vehicle 2 m above local ground level and pointed rearwards so that an unobstructed field of view was obtained.

The receiver was installed in a vehicle which was static during the measurements, with the antenna located on an elevated platform so that its height above ground level could be varied between 6 and 12 metres. The antenna for the dynamic measurement was identical to the transmitter antenna, but for static reflection measurements a more directional antenna was used with the following characteristics in azimuth:

3 dB beamwidth:	$\pm 7°$
6 dB beamwidth:	$\pm 11°$
20 dB beamwidth:	$\pm 25°$
Polarization:	vertical

The signal from this antenna was fed to a 5 GHz receiver, the output being recorded on an instrumentation tape recorder together with a reference signal which was used to check the recording system.

0033–7722/81/020053+05 £1.50/0

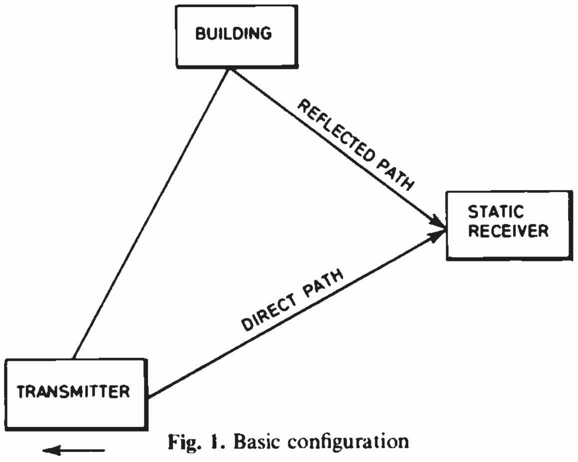


Fig. 1. Basic configuration

2.2 Dynamic Measurements

For each measurement, the transmitter horn was pointed to illuminate both the building face and the receiver antenna within the main lobe. The receiver antenna was positioned at the chosen height (up to 12 m) to receive both a direct signal from the transmitter and a reflected signal from the building, also within the main lobe. For each measurement run the transmitter was moved at a constant speed parallel to the building face considered, this direction being dictated by the geometry of the available tarmac surfaces.

2.3 Static Measurements

For these measurements both antennas were aligned towards the building face considered, the isolation against the direct signal being in excess of 30 dB for all geometries investigated. Whilst the transmitter was moved slowly (walking speed) to illuminate the building across the specular region, the receiver output was recorded. It was established that the motion of the vehicle did not modify results compared with a static transmitter. The configurations used were the same as for the dynamic measurements with the receiver antenna up to 12 m high.

The direct free-space signal strength was measured with the transmitter vehicle positioned at the point for reflection from the middle of the building face. With the transmitter vehicle stationary and the antennas directly aligned the receiver antenna height was varied to show a ground reflection maximum and minimum, and hence the free space signal level. Measurements of the direct link were also made with a fixed receiver antenna height and the transmitter mobile to confirm that there were no anomalous propagation effects; the ground reflection coefficient was found to be between 0·8 and almost unity, dependent on the measurement configuration.

3 Measurement Details

3.1 Building Faces

The buildings considered were aircraft hangars and each had some surface irregularity. The results given within Section 4 involved two different surfaces. One was a basically corrugated metal surface 10 m high and 100 m long. The corrugations were separated by 7·0 cm with a peak-to-trough distance of 1·5 cm, while the coarse surface irregularities had maximum peak-to-peak variations of a few centimetres over most of the length of the building, with the exception of the drainpipes, one small door and the very ends of the building. The second surface largely comprised flat hangar doors, which were in eight sections located in parallel closely spaced sliding tracks. Each section was faced with sheet metal on a backing framework, of size 9·3 m high and 6 m wide, the total width being 48 m. These doors were surrounded at the top and sides by corrugated metalwork.

3.2 Measurement Configurations

In all, 26 different transmitter/receiver/building configurations were investigated, with direct link lengths varying from a few hundred metres to about 1 km. The receiver antenna heights were 6, 8·5 and 12 m. For the dynamic measurements the vehicle speed was usually set to 97 km/h (60 miles/h), although speeds from 12 to 120 km/h were also employed to check linearity with speed. The two measurement configurations for which results are given in Section 4 were as follows, for mid-point reflection.

(a) *Corrugated Surface*

Transmitter to receiver:	560 m
Transmitter to building:	300 m
Receiver to building:	460 m
Angle to building face:	43°
Receiver antenna height:	8·5 m
Ground conditions:	flat and mainly grassed

(b) *Hangar Doors*

Transmitter to receiver:	160 m
Transmitter to building:	80 m
Receiver to building:	120 m
Angle to building face:	46°
Receiver antenna height:	6 m
Ground conditions:	flat tarmac

4 Results

The results presented in Sections 4.1 and 4.2 comprise a representative sample of the data obtained, which covered reflection angles of incidence from 50° to 15°. Reproducibility of results was confirmed by repeat measurements for several configurations at different times. All the experiments were carried out during dry weather.

4.1 Dynamic Measurements

The tape recordings of the dynamic results (i.e. the differential Doppler signal, see Appendix) were processed using a spectrum analyser. In all cases differential Doppler signals were present within (and, at lower amplitude, outside) the specular zone. The centre frequencies of these signals were consistent with theoretical calculations (see Appendix) and therefore validated the technique. The result for reflection from

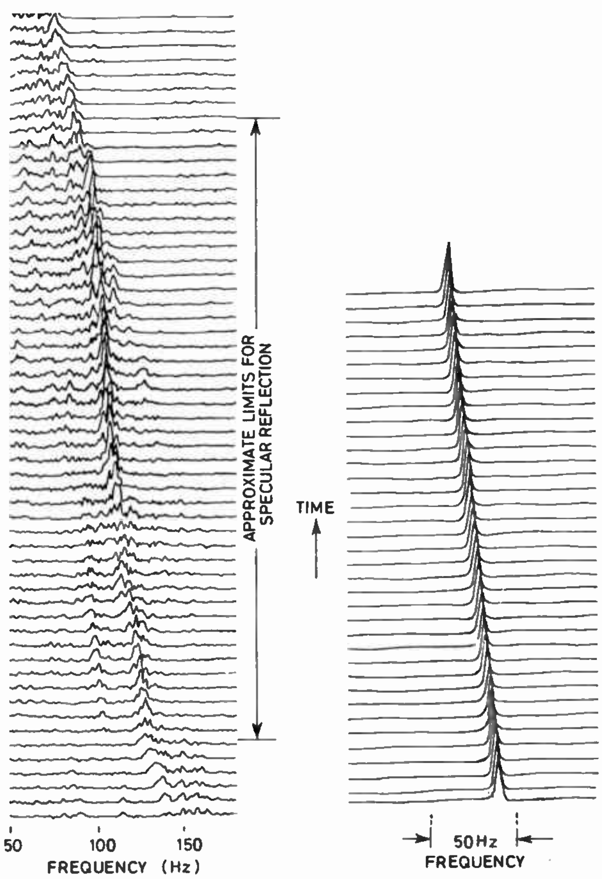


Fig. 2. Signal spectra

the corrugated surface is reproduced in Fig. 2(a) for a 1 Hz spectrum analyser bandwidth with a linear voltage display. The spectrum is noisy over a fairly wide band and low-level signals are present both before and after the specular region. The spectrum of a swept sinewave oscillator is shown in Fig. 2(b) for the same 1 Hz analysis filter bandwidth and illustrates that little spectrum broadening occurs due to the variation of frequency with time.

On average, the main line in Fig. 2(a) has a width at half the peak spectrum density of 2 Hz at 97 km/h with peak spreading of about 10 Hz. Experiments with different vehicle speeds showed that the spectrum width and differential Doppler frequency scaled linearly with speed for a given measurement configuration. The results for the corrugated surface are shown in Fig. 3. Overall, the spectrum widths showed no strong correlation with the receiver height used, or indeed the measurement range, for the situations considered.

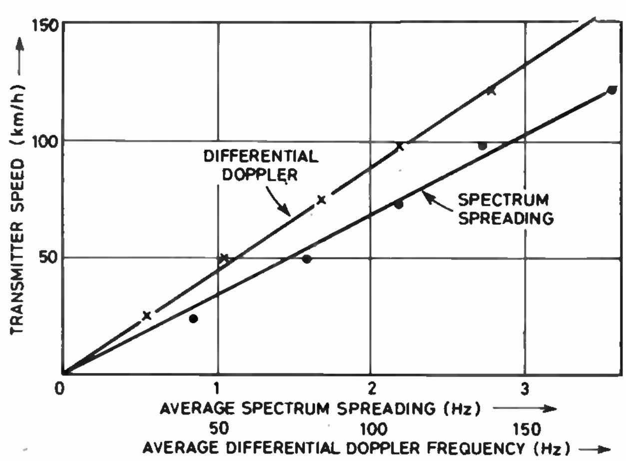


Fig. 3. Speed scaling (corrugated surface)

The amplitude of the main 'line' in Fig. 2(a) is very variable although this is smoothed from the true amplitude in these results by the 1 Hz analysis filter bandwidth.

4.2 Reflection Amplitude Measurements

The variation of the reflected signal amplitude for both the corrugated surface and the hangar doors are shown in Figs. 4 and 5 respectively. The high gain antenna was used at the receiver to discriminate (> 30 dB) between the direct and reflected signals; for reflection from the hangar doors the receiver gain varied over the region of interest and the 0, −3 and −6 dB positions are shown in Fig. 5.

5 Discussion

5.1 Nature of Results

The dynamic measurements showed non-specular characteristics with spectrum spreading, variable spectrum amplitude and background 'noise'. The differential Doppler frequencies of this background were consistent with scattering from the building surface; qualitatively, when reflection starts to occur from the surface (high differential Doppler), scattering produces only higher frequencies, and when reflection is about to cease the scattering produces lower frequencies only (see Appendix).

The reflection measurements showed that the signal amplitude varied rapidly (of the order of 1 m movement) from deep nulls to high level spikes. The signals outside the specular region at about −10 to −20 dB down on the direct free space signal were in excess of any direct transmitter to receiver leakage, confirming scattering: the levels observed were generally higher than those reported in Ref. 3, for tall buildings at X band. In this context it should be noted that the results in Fig. 4 and 5 were restricted to near the specular zone only.

5.2 Theoretical Considerations

The buildings considered were large enough to intersect a number of Fresnel zones. For the sample cases the first Fresnel zone radii were 3·3 m for the corrugated surface and 1·7 m for the hangar doors, whilst the geometrical reflection points were 4·6 and 3·6 m respectively above ground level. If the reflecting surfaces were perfectly flat the reflection coefficient at 5 GHz would be expected to be almost constant over all the building face with the exception of the ends where the finite wavelength would cause a ripple in the amplitude of the reflected signal.

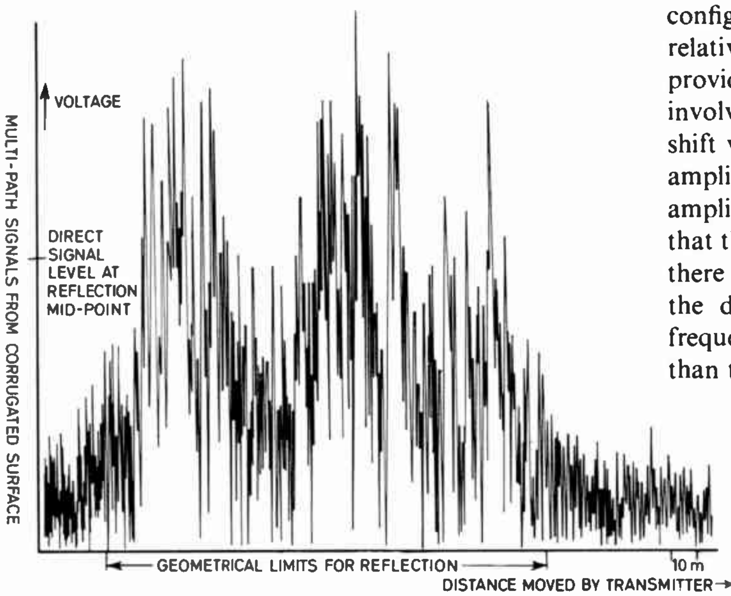


Fig. 4. Multi-path signals from corrugated surface

The effect of ground reflections modifies the situation, the two dominant rays being the simple building-reflected ray and the transmitter–ground-reflected–building-reflected–receiver ray. The relative difference in path lengths between these two rays for antenna heights h_1 and h_2 and a horizontal transmitter–building–receiver distance d is approximately given by:

$$\text{Path difference} = \frac{2h_1h_2}{d}, \quad \text{for } h_1, h_2 \ll d.$$

Assuming a $-\pi$ phase shift on ground reflection,

$$\text{Phase difference} = \frac{2h_1h_2}{\lambda d} \times 2\pi - \pi \quad \text{radians.}$$

In general, the relative phase shift between the two rays would be reasonably constant for long measurement ranges, where the variation in d would be small over the specular zone. This result is true for the configuration used for the corrugated surface, the relative phase shift varying from $0{\cdot}4\pi$ to $0{\cdot}6\pi$, hence providing some signal enhancement. The hangar doors involved shorter measurement ranges, the relative phase shift varying from $2{\cdot}5\pi$ to $3{\cdot}6\pi$; this includes a null, of amplitude dependent on the ground reflection amplitude. However, the results in Fig. 5 demonstrate that this is not the dominant mechanism, particularly as there is no single clearly identifiable null. With respect to the dynamic measurements, the differential Doppler frequency due to ground multi-path would be much less than that for building reflection.

The effect of surface irregularities on the reflection coefficient is discussed in Ref. 4. A useful parameter for roughness is given by:

$$\Phi = \frac{4\pi\sigma}{\lambda} \sin\theta$$

where σ is the standard deviation of surface roughness

λ is the signal wavelength

θ is the angle made by the incident ray to the surface.

In general if Φ exceeds unity, significant scattered energy would be expected at the expense of the coherent component. For the cases described Φ would be about unity over the first few Fresnel zones for the corrugated surface and approaching zero for the hangar doors where reflection occurred in the middle of a door panel. On this basis a significant coherent component would therefore be expected: nevertheless, even the results for the hangar doors demonstrate very rapid signal fluctuations.

6 Conclusions

6.1 Dynamic Measurements

The reflection of microwave signals at 5 GHz from surfaces such as hangar walls and doors produced a frequency spectrum whose width depended on the movement of the source. This width was found to scale linearly with the transmitter speed for a given measurement configuration, the average spread being typically 2 Hz at 97 km/h. Associated with this spreading of the main signal was a background 'noise' of differential Doppler frequencies which occurred both inside and outside the geometrical reflection zone. In addition large variations in the amplitude of the main signal were apparent.

6.2 Static Measurements

The received signal level rapidly fluctuated over the specular zone, the amplitude varying from a level comparable with the direct free space signal level to nearly zero with distance variations of the order of 1

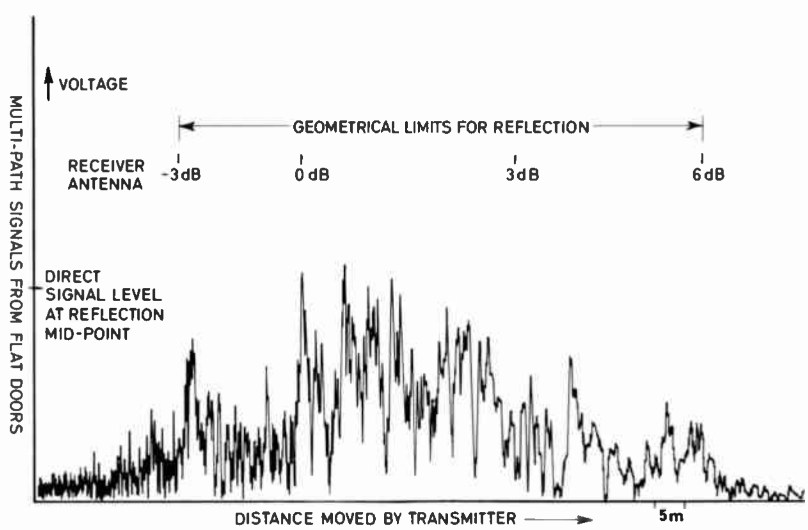


Fig. 5. Multi-path signals from flat doors

metre. The scattering level close to the geometrical reflection region was about -10 to -20 dB with respect to the direct signal level.

6.3 Application of Results to Other Situations

The results presented relate to real life measurements with large buildings, rather than extrapolations based on considerations of surface material. The effect of ground reflection was found to be secondary compared with the surface characteristics, with the roughness parameter Φ in the first Fresnel zone being between 0 and 1 at 5 GHz. For these values of Φ the rapidity and magnitude of amplitude fluctuations had not been expected from simple theory.[2] Therefore, it is expected that in practice large generally flat buildings (Φ about 0 to 1) would produce noise like reflection characteristics, as Figs. 2, 4 and 5, over the specular zone with scattering at about -10 to -20 dB close to this region. With such man-made structures it is apparent that the roughness over the first Fresnel zone does not adequately define the reflection characteristics. If surfaces other than buildings are considered the extrapolation of these results may not be appropriate, although the qualitative limitation of simple theory may still apply.

7 Acknowledgments

This project was undertaken for the National Air Traffic Services (a joint Ministry of Defence—Civil Aviation Authority service). The author wishes to acknowledge the co-operation afforded by the Royal Air Force in the use of RAF Fairford, and the assistance provided by his colleagues and, in addition, Mr F. Blackwell of RSRE, Malvern, in the spectrum analysis of the tape recorded signals.

8 References

1 Burchsted, R., Capon, J., Evans, J., Orr, R., Shnidman, D. and Sussman, S., 'Computer simulation of ICAO MLS candidates in representative airport environments', 'The Future of Aircraft All-Weather Operations', 1976, IEE Conference Publication 147, pp. 149–52.
2 Capon, J. 'Description of Computer Programs to Determine Obstacle Multipath Effects for MLS Simulation', MIT Lincoln Laboratory 44 WP-5020, 15th May 1976.
3 Bramley, E. N. and Cherry, S. M., 'Investigation of microwave scattering by tall buildings' *Proc. Instn Elect. Engrs*, **120**, no. 8, pp. 833–42 August 1973.
4 Beckman, P. and Spizzichino, A., 'The Scattering of Electromagnetic Waves from Rough Surfaces' (Pergamon, Oxford, 1963).

9 Appendix: Differential Doppler Frequency

The differential Doppler shift may be established by reference to Fig. 6. The Doppler shifted frequencies of the direct (f_d) and reflected signals (f_r) are given by:

$$f_d = f_0 - \frac{V}{\lambda}\cos\alpha \qquad f_r = f_0 - \frac{V}{\lambda}\cos\theta$$

where f_0 is the transmitter frequency, V is the speed of the transmitter vehicle and λ is the signal wavelength.

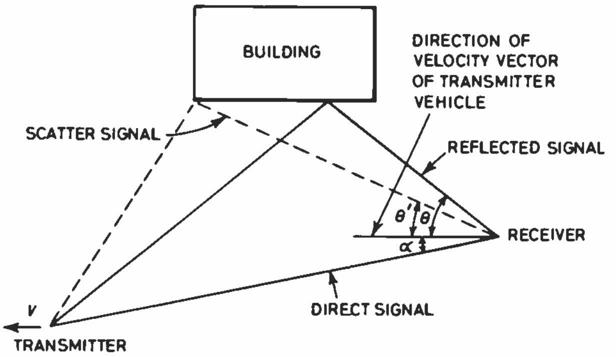


Fig. 6. Differential Doppler calculation

The differential Doppler frequency f_D is given by

$$f_D = f_r - f_d = \frac{V}{\lambda}(\cos\alpha - \cos\theta)$$

The differential Doppler frequency for a scattered signal is given by replacing θ by θ'.

For the example given in Section 4, at the beginning of the building reflection the differential Doppler frequencies for a speed of 97 km/h are:

f_D (reflected) = 129 Hz f_D (scattered) = 258 Hz (max)

Similarly, at the end of the building reflection the corresponding frequencies are:

f_D (reflected) = 89 Hz f_D (scattered) = 52 Hz (min)

The results in Fig. 2 are consistent with these values at the end of the building, although the maximum frequency indicated is limited to below that predicted: this is possibly due to a lower scattering coefficient and antenna gain at higher scattering angles (θ').

Manuscript first received by the Institution on 10th September 1979 and in revised form on 8th May 1980
(*Paper No.* 1970/*AMMS*106)

Standard Frequency and Time Service

Communication from the National Physical Laboratory

Relative Phase Readings in Microseconds NPL—Station
(Readings at 1500 UTC)

SEPTEMBER 1980	MSF 60 kHz	GBR 16 kHz	Droitwich 200 kHz
1	2·4	21·5	23·0
2	2·6	20.8	23·1
3	2·4	21·0	23·2
4	2·6	21·4	23·3
5	2·6	21·4	23·4
6	2·6	21·2	23·5
7	2·6	21·0	23·6
8	2·7	21·3	23·8
9	2·5	21·1	23·9
10	2·6	21·1	24·1
11	2·6	*	24·2
12	2·7	21·1	24·4
13	2·6	21·0	24·6
14	2·5	21·0	24·7
15	2·6	21·6	24·8
16	2·4	21·0	24·9
17	2·4	21·0	25·0
18	2·4	20·8	25·0
19	2·4	20·9	25·0
20	2·4	21·0	25·0
21	2·2	21·1	*
22	2·3	21·3	*
23	2·3	21·0	25·1
24	2·3	21·2	25·1
25	2·3	21·3	25·2
26	2·3	21·3	25·3
27	2·2	21·4	25·5
28	2·1	*	25·6
29	2·3	22·1	25·7
30	2·3	21·6	25·7

OCTOBER 1980	MSF 60 kHz	GBR 16 kHz	Droitwich 200 kHz
1	2·2	11·6	25·7
2	2·4	11·7	25·7
3	2·4	11·6	25·7
4	2·3	11·8	25·7
5	2·3	11·8	25·7
6	2·3	12·3	25.8
7	2·1	11·3	25·9
8	2·0	12·8	26·1
9	2·0	13·6	26·2
10	2·0	13·5	26·3
11	2·1	*	26·3
12	1·9	*	26·4
13	2·1	14·6	26·6
14	1·9	13·9	26·6
15	1·9	13·7	26·6
16	1·8	13·3	26·6
17	1·7	13·0	26·7
18	1·6	12·1	26·7
19	1·6	12·6	26·8
20	1·3	12·5	26·9
21	1·5	13·3	26·9
22	1·5	13·1	26·9
23	1·5	12.6	27·0
24	1·5	12·9	27·0
25	1·3	13·1	27·0
26	1·2	12·6	26·9
27	1·4	13·3	26·9
28	1·5	13·0	26·9
29	1·5	13·4	26·8
30	1·3	13·6	26·9
31	1·3	13·6	27·0

Notes: (a) Relative to UTC scale (UTC_{NPL}–Station) = +10 at 1500 UT, 1st January 1977.
(b) The convention followed is that a decrease in phase reading represents an increase in frequency.
(c) 1 μs represents a frequency change of 1 part in 10^{11} per day.

MSF 60 kHz TRANSMISSION. A phase change of approximately 2·5 microseconds occurring at about 2230 UT on 21 September 1980 has been removed from the table of accumulated phase.

GBR 16 kHz TRANSMISSION. A systematic re-examination of the published phase values over the period October 1979 to October 1980 has revealed a number of discordant results. These have occurred during the winter months, when there is an appreciable ionospheric component in the total received signal. The consequent phase fluctuations give rise to the possibility of the phase-tracking receiver transferring to an adjacent cycle, when the transmission is restored after the interruption for maintenance on Tuesday of each week. This results in an apparent change of precisely 2·5 microseconds in the phase-tracking record; four such changes in the same sense have been identified, in the period October 1979 to March 1980, leading to an accumulative change of 10·0 microseconds.

An appropriate correction has been introduced for the first and subsequent readings for the month of October 1980 and will rectify the error in relating GBR phase to UTC (NPL) on a long term basis. A complete list of amended results has been prepared and will be supplied on request to Division of Electrical Science, National Physical Laboratory, Teddington TW11 0LW, England.

UDC 62
Indexing Terms: Systems science, Operations research, Production control, Terotechnology.

A framework for systems engineering design

Professor P. K. M'PHERSON,
S.M., M.A., C.Eng., F.I.Mech.E., F.I.Mar.E., F.Inst.M.C.*

SUMMARY

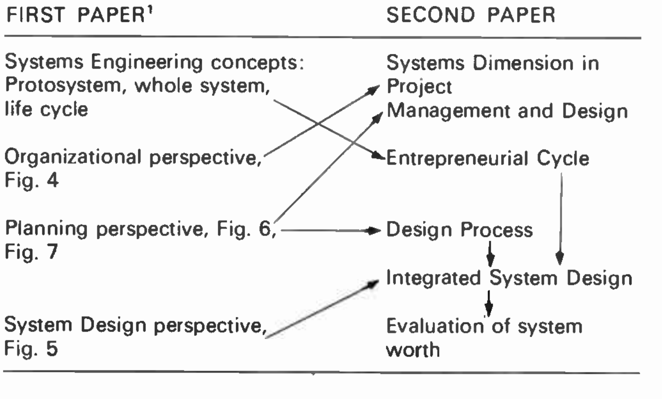
This paper follows on from the previous paper in the series and attends to the methodology of design as conceived within the whole-system whole-life dimensions of systems engineering. It assumes the concepts and general structure of systems engineering as outlined in the previous paper and goes on to elaborate the form and content of systems engineering from a design perspective.

The paper begins by considering the systems engineering organization within the hierarchical structure for project design and management that has evolved from the one-time individual project engineer. Then the important relationships between entrepreneur/contractor, customer/owner, user are explored to illustrate why design in the system engineer's hands is dominated by overall cost-effectiveness optimized with respect to the future whole-life operations. The systems engineering design process is very similar to multiple-objective decision analysis as both designer and decision maker are faced by a similar problem: to find a way of organizing action and deploying resources so that a desirable future state is realized efficiently. The decision analytical process is structured to provide a framework for the systems design methodology itself. This follows using a top-down (hierarchical) approach in which overall system worth is determined with respect to clearly articulated objectives and evaluated with respect to a comprehensive multiattribute value criteria. Cost-effectiveness is disaggregated into its components of availability, dependability, capability and discounted whole-life cost. These in turn are related to the sub-design problems of operational performance, reliability, maintenance, logistic support, risk and impact analysis.

* *Department of Systems Science, The City University, Northampton Square, London EC1V 0HB.*

TABLE OF CONTENTS

The following table bridges the two papers:



Note. A list of symbols is given in Appendix 5.

0033–7722/020059+35 $1.50/0

THE SYSTEMS DIMENSION IN PROJECT MANAGEMENT AND DESIGN

Technological Progress Overtakes the Individual Project Engineer

Time was when a competent project engineer could be expected to encompass all the technical and economic aspects of even large projects, as well as being his own project manager. The history of technology is full of the names of heroic engineers such as Telford, Stephenson and Brunel. These giants of the first industrial age were certainly matched by the bridge, railway and dam builders, the electrical engineers who brought power and light to towns and regions, the late 19th century chemical engineers who founded whole new industries, and the telephone and radio engineers who encircled the globe with their telegraph and wireless networks. All were pioneers and innovators, designers and managers caught up in the 19th and early 20th century excitement of technological progress and prowess. They had no computers, their science was often short, but their courage was long as they reached for their practical solutions. They often made mistakes, but life, energy, materials and money were cheap, and few thought that there might be limits to growth and that the planet was anything but a resource to be exploited. Where are the heroes now? What has happened?

There were no 'systems' as we understand them. Each branch of technological endeavour built as economic advantage dictated and capital allowed. The performance of their assemblages (rather than systems) was tied to whatever capability, efficiency and reliability was available from the components they used. Of course they were cost conscious and aware of the defects. But they operated within the state of knowledge, failures had to be accepted—and then studied so that improvements could be made. To avoid failures they used conservative designs with large safety factors. There was little of the complexity introduced by the electronic interlinking of components and systems. What could be done was what could be calculated by hand.

One can understand that the coal-fired power stations, guns, railway engines, balloons, telegraphs of 1900, even 1939, were designed—perhaps 'put together' is nearer the mark—by small self-contained teams under the leadership of a chief designer and project engineer who knew and could check all the detail. But the nuclear power stations, missile systems, high-speed trains, jumbo jets, satellite communications systems of 1980? These are so complex and expensive that trial and error development is at a premium, mistakes cannot be afforded (they are no longer acceptable socially anyway); in fact they have to be designed 'right' first time. And 'right' means very safe, reliable, clean, efficient, with continuous operation at high ratings and pared down safety margins, and still be highly available and cost-effective throughout a long operational life.

The Trend Towards Systems-Integrated Project Management and Design

In the capital goods sectors, particularly, the trend is towards turn-key contracts for which the main contractor has responsibility for the entire project—that is design to customer's requirements, development, delivery and deployment of the main operational parts together with all the related and ancillary supporting parts. And this can even include the provision of educational and training facilities for the future operators and maintainers. The customer is not just buying hardware, he wants a fully integrated and operational system together with all the knowhow for efficient life-time utilization. The trend has developed because, if 'big systems' are to work, it is best if they are brought into being as a total concept rather than being an assemblage of independently produced (off the shelf) items with the customer attempting to put the parts together. The trend has been most visible in the high technology military, nuclear and aerospace sectors. But it is also increasingly apparent in technological activity at large, particularly in the technology transfer market whereby major capital works and whole new industries are inserted into developing countries that have as yet relatively weak industrial infrastructures. At the top end of the scale participating firms combine as consortiums that then have to solve the problems of coordinated and integrated project management. Efficient performance at this level gives an advantage to larger-scale project teams with advanced systems know-how. But the trend is not limited to the giant projects. The complexity of modern technology, particularly where the operational units are integrated via sophisticated electronic instrumentation, control and computer systems, requires a systems dimension in design and project management from the start. More often than not, the customer has little understanding of all the technical detail and is more of a user/driver who cannot 'lift the bonnet' to do a running repair if something goes wrong. So customers are getting wary of ill-assorted technical packages: they want fully integrated systems right down to unit level, e.g. machine tool centres, integrated condition monitoring and alarm systems, integrated design of all the navigational aids on a ship's bridge. As a final spur to the systems trend, the capital cost and the repair and maintenance costs (including down-time penalties) of systems-integrated high technology are so high that customers increasingly require advance demonstration (at a pre-contract stage) that the proposed system will not only meet their performance specification but will continue to be cost-effective over its whole operational life in the particular context of the customer's intended operational environment. Consequently even small scale projects require the full systems treatment in their design and development.

Emergence of a Hierarchical Structure for Project Management and Design

The individual hero engineers of the past have become heroic teams of highly skilled specialist designers of the performance, reliability, safety, logistics, control and information aspects of the system, all applying their 'art' at the most advanced level if the system is to be competitive. In turn this has meant the development of design hierarchies for exactly the same reasons that hierarchical management structures develop in any complex organization:[2]

- the system is too complex for any one designer/decision maker (DM) to comprehend as a whole, hence separation into manageable systems;
- sub-system designers/decision makers (SDMs) have limited (but expert) information and attend primarily to local goals;
- to achieve the task in the same time the SDMs should work in parallel;
- goal coordination of the SDMs is necessary to ensure that the overall system goal is satisfied, hence superior control of the SMDs;
- SDMs can only track their interactions through constant intercommunication, hence the need for an efficient information system integrating the consequences of SDM activity at a whole-system level.

The result is a design hierarchy as shown in Fig. 1.

Design itself is but one of the major phases in the development of a project to its operational state, the management hierarchy endeavouring to control and coordinate a matrix of interdependent phases, functions and design tasks.

The Systems Engineering Layer in Project Management and Design

The system-wide concept of design and project management has introduced a new layer of design and management between the overall project manager and the specialist sub-system design teams and detail designers. This new layer is the province of the systems engineer whose task is to provide the system-wide integration and coordination between and within the phases of system realization, acquisition and operation. Referring back to Fig. 1 the systems engineering functions in the design phase are represented by the functional blocks with thick borders. The structure shows clearly that the systems engineer in design provides the integrating and coordinating linkage between the design manager and the subsystem design teams. He has not usurped the final authority of the chief designer (now better regarded as the design manager), nor has he substituted for the specialist design functions. But the systems engineer has taken over a fair part of the old-fashioned project engineer's responsibility for coordination and integration between the specialist design functions that are now the responsibility of the various subsystem design teams. Similar management structures can be drawn for the other main phases of the systems engineering process: development, production, trials and so on. In all these phases the systems engineers interpose between the phase manager and the specialist

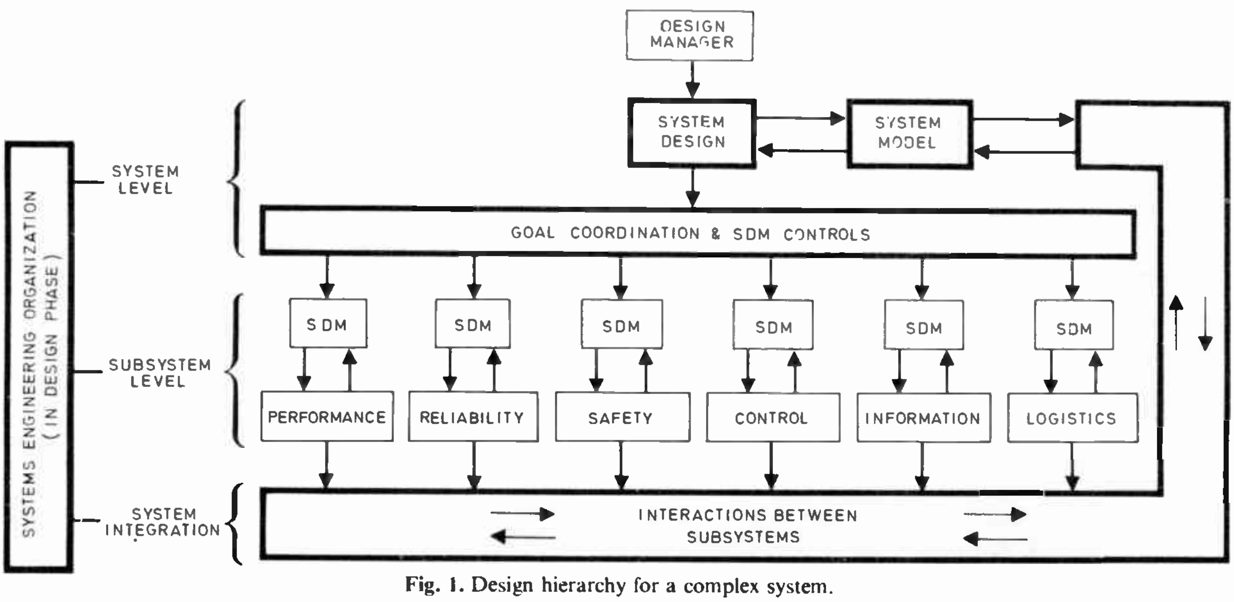


Fig. 1. Design hierarchy for a complex system.

other phases are Development, Production, Testing and Insertion. Each phase will have a management hierarchy of its own, and the several phases will need system-wide integration and overall coordination. Thus the original project engineer's task has now become a complex

teams. In turn the many phases require coordination and integration themselves, producing another level in the overall project management hierarchy. The general organization of systems engineering within project management is illustrated in Fig. 2. The normal hierarchical chain of command from project manager through phase managers down to the subsystem team leaders is shown. But a hierarchical structure is defective because lateral interlinkage occurs only through the superior elements in the chain. The systems engineering organization provides the project and phase managers with the necessary lateral and vertical linkage to ensure coordination and integration of the project within itself and through time. (References 3–7 list five books that cover many of the problems of management, organization and design in systems engineering.)

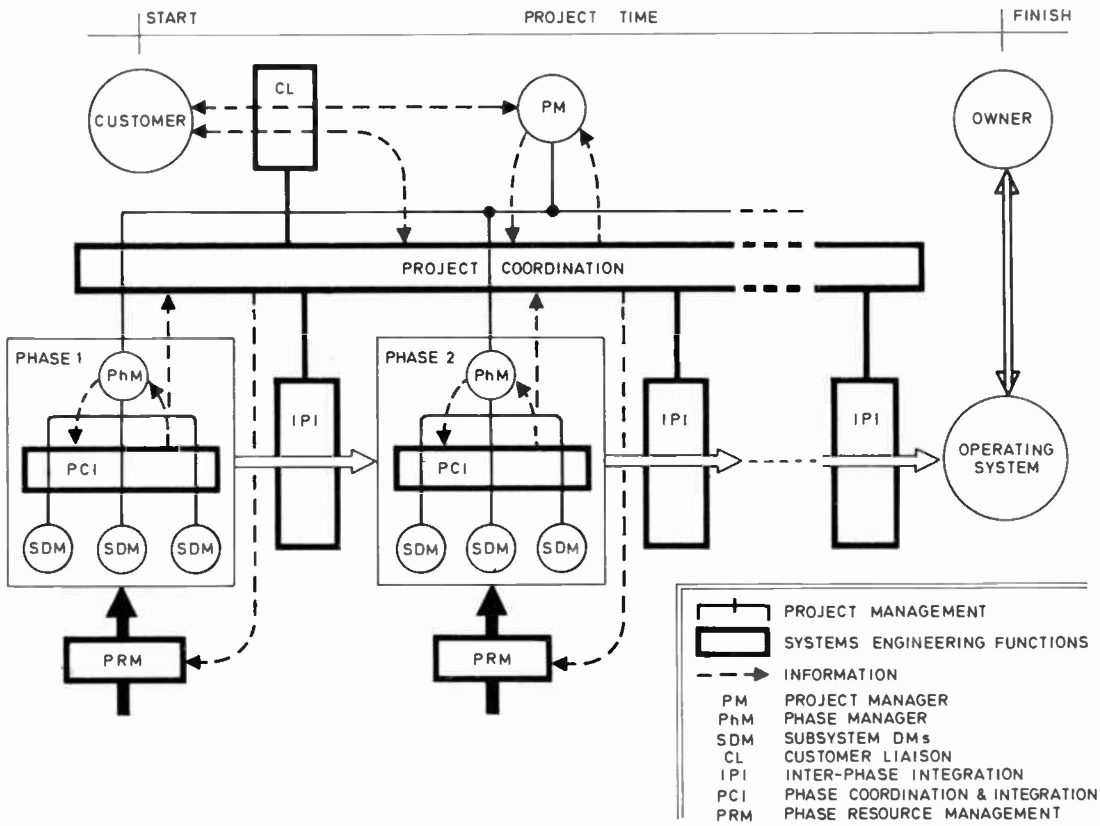


Fig. 2. Systems engineering organization in project management.

THE ENTREPRENEURIAL CYCLE

Any activity requires the conversion of resources. Technological activity is mainly concerned with the supply of useful products or services resulting from the operation of a technological system in which men and machines 'collaborate' to convert materials, energy, information and human effort into the desired output. The technological activity is itself divided into two distinct phases: that in which the potential technological activity is first being engineered, and that in which it is subsequently operated until it is phased out. The systems life-cycle in other words. But one other resource has to be found before any technological activity can take place at all: a cash flow to pay for the resources consumed in engineering and operating the system prior to the point at which the activities themselves convert into direct revenue (or the revenue equivalent of services rendered) and begin to repay the original cash flow source. It is sometimes difficult to get this simple fact across to people reared in the hire-purchase economics of 'have now, pay later', for in engineering the exact reverse is true: 'pay now, have later'. Somebody right at the start of the engineering process has to persuade a banker (or financial backer ... the same thing) to part with sufficient capital to launch the project in the expectation that the technological potential is sufficiently promising for a profitable return on the investment in the future. In large-scale capital projects that profitable future may be decades away ... and in some important classes of project there is never any direct profit—only costs, as in motorways, hospitals and military services.†

Cost-Effectiveness as an Economic Appraisal Index

All this is the stuff of corporate and investment strategy, capital project appraisal or just plain engineering economics.[8] It is also prime stuff for the systems engineer as whole-system whole-life cost-effectiveness is an indicator of a proposed system's merit in terms of the effectiveness likely to be achieved with respect to its predicted overall life cost.

† Of course tolls can be levied on the users of motorways, and hospitals can charge fees from their patients, but nobody has yet found a way for persuading the private citizen to buy voluntarily a share in a tank or warship ... until it is 50 years old when it becomes an antique.

EFFECTIVENESS: the probability that an operational system will satisfy stated performance or mission requirements under stated operational environments during a stated period (up to whole-life).

COST: the discounted whole-life ownership cost of a system including capital cost, operational and support costs, maintenance and unreliability costs, renewal costs, hazard and environmental impact costs, phase-out costs.

Thus cost-effectiveness (CE from now on) is the systems engineer's measure for capital appraisal. It allows him to compare and justify alternative system designs in terms relevant to the future owner/operator's interest. The future owner/operator should not be interested only in the acquisition cost (i.e. the purchase price or capital cost) which is what concerns the contractor. To the owner the total cost of the system is its purchase price plus the total ownership costs likely to arise from future operations. Then the owner can make a proper assessment of the system's overall worth by balancing the likely future revenue (profit or benefit) against the total cost. A cost-effective system is not necessarily a minimum cost system. (CE will be examined in more detail later.)

The Entrepreneurial Process of Project Origination

CE is a technologically oriented indicator of the accountant's criterion of the return on investment (ROI). It is worth remembering that the assessment of CE or ROI is made during the early preliminary design phase *prior* to obtaining sanction to proceed with development. Thus the systems engineer is an important instrument in the entrepreneurial business of getting a project launched. One might describe the process as an entrepreneurial cycle (Fig. 3) in which there are four actors:

- the Originator who conceives a promising technological option and undertakes to develop and deliver it;
- the future Owner/Operator who will take delivery of the operational system and proceed to exploit it;
- the future Users/Consumers of the products/services

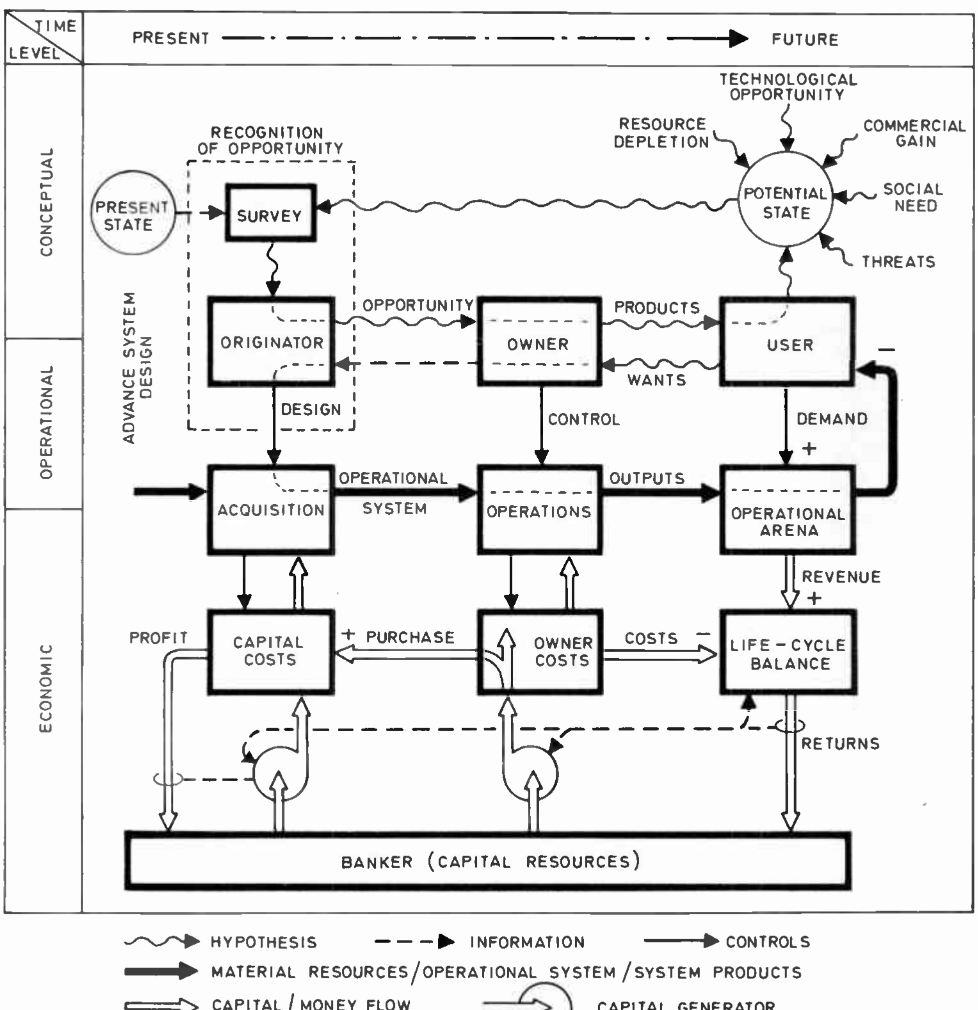


Fig. 3. The entrepreneurial cycle.

provided by the operating system;

- the Banker who provides the risk capital as the primary resource that converts into the operating system.†

The characters of originator, owner and user are reasonably clear-cut, but their forms can differ widely as illustrated in Table 1. For example, the originator, owner and user may be different parts of the same organization, they may be a completely separate organization, or they may represent mixed transactions between government and industry or between governments.

The objectives of originator/contractor and customer/owner are realized at the operational level. Here there is another cycle in which requirements are realized, via the design and acquisition of an operational system, in the form of outputs to the operational arena. The outputs satisfy the initial demand, but usually generate new wants. The two cycles repeat. But the operational cycle needs an economic incentive (usually) before it can be started. The fourth actor enters—the banker—who needs to be persuaded that the technological opportunity being discussed at the conceptual level is sufficiently promising for there to be a good chance of profit for the contractor, owner and banker (the banker means here the various sources of capital funding). Notice that it is the promise of the *eventual* returns to the owner that start the capital 'pumps' in the economic level of Fig. 3. The capital flow to the originator will start if the returns from future operations (perhaps years ahead) are expected to leave a satisfactory margin over *ownership* costs. The ownership costs include the purchase price—which is really a transfer of capital back to the contractor to cover the initial capital costs and contractor's profit. (The banker and government take their cuts as interest on the loans and taxes on profits.)

Table 1

Originator	Owner	User
Inventor + Developer	ABC Ltd.	ABC Ltd.
XYZ (Research) Ltd.	XYZ (Holdings) Ltd.	XYZ (Operations) Ltd.
ABC Ltd.	XYZ (Operations) Ltd.	Consumer market
ABC Ltd.	Public Authority	Public Service
XYZ (Research) Ltd.	Government	Military Service
Government Research Lab.	ABC Ltd.	XYZ (Operations) Ltd.
XYZ (Research) Ltd.	Government of K via UN Aid Programme	Developing industry in K via XYZ (Operations) Ltd.

In the beginning there has to be somebody with an idea for an undertaking that could solve a problem. The somebody could be the originator or the future owner who feeds the idea to the originator (now really the contractor). The problem arises from the realization of a mismatch between the present 'state of the world' and a more desirable future state that could occur if the undertaking is brought to fruition. The future state may be discerned from either a defensive or offensive standpoint:

- Defensive undertaking: e.g. a technological shift consequent on resource depletion, a socio-economic or socio-technological need amenable to a technological solution, improvement of an existing muddled or degraded technology.
- Offensive undertaking: e.g. exploitation of a technical opportunity (invention, innovation), raising or increasing a market, making the enemy's military technology obsolescent.

So the first part of the entrepreneurial cycle is at the conceptual level (Fig. 3) in which the mismatch between present and future states is recognized as an opportunity which could bring profit to the originator and owner. Notice that the left-hand part of Fig. 3 is in the present, while the right-hand part is in the future. In present time the future owner is the originator's *customer*.

† More often than not the Government is the 'Banker' for profitless but socially desirable ventures, e.g. those motorways and weapons. The originating funds are drawn from tax revenue, and the 'return on investment' is the social benefit received by taxpayers and society at large . . . sometimes.

The Systems Engineer as Customer's Friend

The journey round the entrepreneurial cycle may have provided a simple lesson in engineering economics, but the main point of the exercise has been to demonstrate why the systems engineer is in many respects the customer's friend, even though he may be employed by the contractor. The systems engineer is part of the originator's organization and is in the present with respect to Fig. 3. While he is assisting with the Systems Survey and Advanced System Design in the Conceptual Phase he is matching the customer's needs to the likely future operational requirements, and he looks far into the future to estimate the whole-system whole-life cost-effectiveness of the candidate designs from the owner's point of view. With this form of assessment the customer will know that when he takes over from the contractor he is getting for his money a system that has been optimized to his overall operational needs and not just to a trial performance specification.

THE SYSTEM DESIGN PROCESS

Systems Assurance

The overall aims of the systems engineering organization are twofold:

(i) to establish overall system performance requirements and to ensure that the system candidate that goes forward for acquisition and operation satisfies whole-system whole-life cost-effectiveness criteria with respect to the defined requirements.

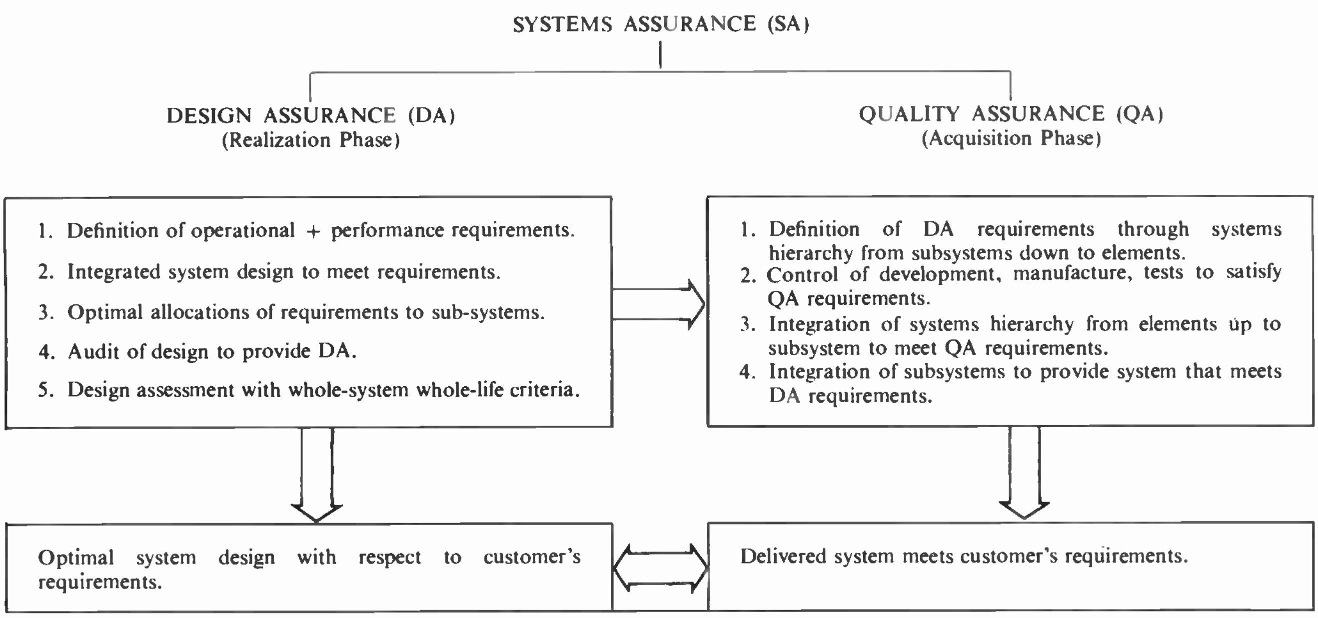
(ii) to ensure that all aspects of detail design, development and production during the acquisition phase are so planned, monitored and integrated that the assembled and operating system does in fact achieve the specified performance within the required cost-effectiveness boundary.

These functions of the systems engineer are becoming known as Systems Assurance. It embraces the above two aims which are named respectively as (i) Design Assurance and (ii) Quality Assurance.[6,21] Table 2 summarizes the systems assurance concept.

One of the main thrusts of systems assurance is towards integrated systems design (ISD). Here 'design' is not a matter of only achieving a stated performance specification, instead 'design' has imbedded in it all the considerations necessary to ensure that the system not only meets its functional requirements but also continues in a satisfactory working state, i.e. it performs according to its design specification and is not degraded by any malfunctions due to wear, breakdowns, inefficient operation, low quality supplies, etc. The systems engineer has to take these degrading factors into account as it costs money to provide a support system that will ensure that the functional system remains in its 'as-good-as-new' working state. In other words ISD requires an integrated model incorporating both the functional processes and the support system. Without the support system the functional processes will continue to operate until the first failure that takes the operational system into a failed state. Then it does not matter how well the functional processes were designed: the failed system is useless and represents only so much idle capital. Moreover, maintenance and logistic costs accumulate over a system's life, and it is by no means unusual for the total to amount to *more* than the initial capital cost of the plant. For example, a modern wide-bodied jet airliner (with a nominal 14 year life) accumulates technological ownership costs greater than its purchase price (at constant prices) after about 12 years.[22] To the owner of that aeroplane its true price is more than double the actual purchase price. And just as much may be said for a control system or an instrument: however accurately it may meet its design specification, its real price to the owner could be considerably higher than the purchase price if it is unreliable and needs continuous maintenance and tuning. Nor is that the end of the story: an instrument wandering from its calibrated setting can detune the host plant and cause wasteful production; and if the plant has to be shut-down while the instrument is changed—then the consequent loss of revenue should really be charged to that instrument or to its ineffective maintenance.

Table 2

Systems assurance



General Structure

The essential structure of a Design Process is that of a cyclic decision process that iterates towards the selection of a concept whose predicted behaviour (when and if the concept becomes operational) best meets a stated design criterion. A structure for the process is shown in Fig. 4. It belongs to the family of applied research methodologies such as Applied Systems Analysis, Decision Analysis, Operational Research and Systems Engineering.[3–7,9–20] Each of these bases its methodology on the following general sequence of stages:

- ∗ Problem definition (Needs, Objectives, Requirements, Constraints)
- ∗ Synthesis (Candidate Generation)
- ∗ Modelling (Design)
- ∗ Analysis (Predict and Analyse Performance)
- ∗ Evaluation and Optimization
- ∗ Selection

The reader is encouraged to spend a little time studying Fig. 4, following the paths through in the order of the numbered boxes whose labels should be self-evident. Note the 'Customer' who provides the input to the problem description [1]†, whose preferences should be reflected in preference analysis [23], and who also makes the final decision [30] as to which candidate system will be selected to go forward for development.

The structure is compartmented by Design Stage and Level. The six design stages (at top) follow the usual decision or design sequence. The fourth stage 'Analysis' implies an analysis of the predicted system's performance with respect to its

• satisfaction of design objectives	Performance Analysis‡
• effect on systemic objectives§	Impact Analysis
• hazard level	Risk Analysis

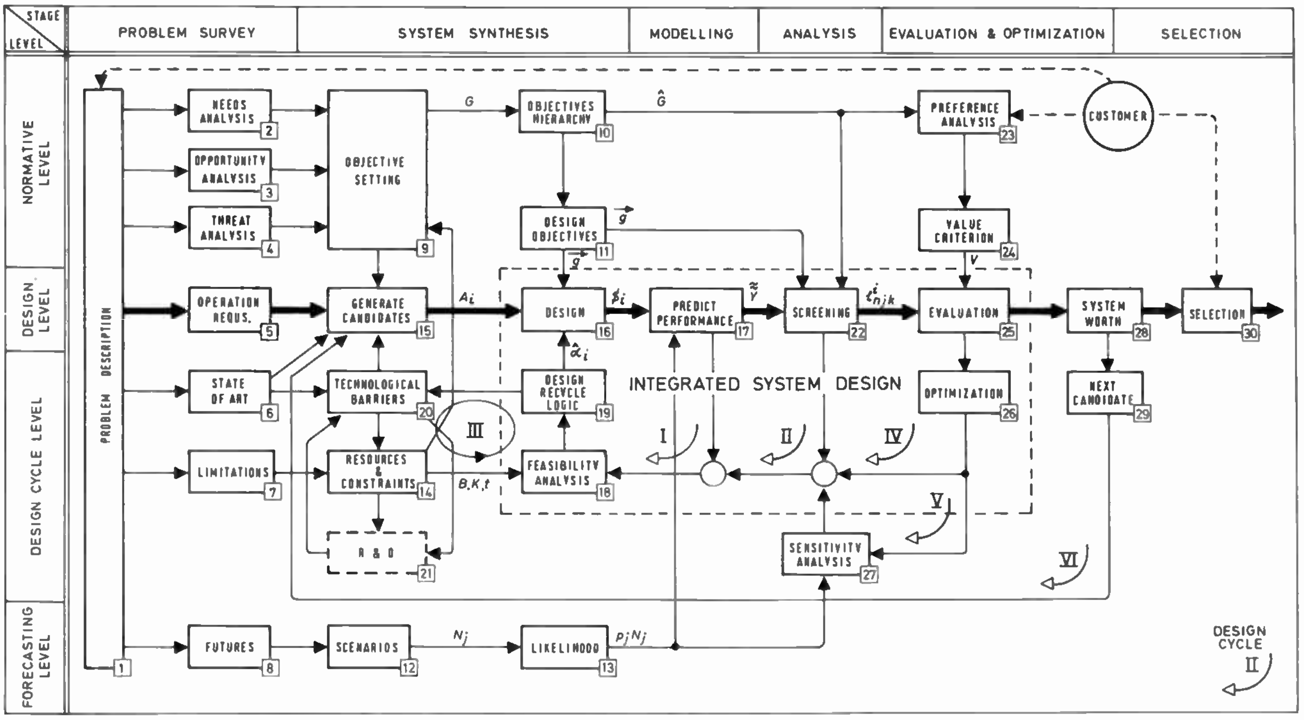


Fig. 4. The system design framework.

The Design Level contains the sequence of functional operations that combine the detail of the forwards design process starting from the definition of the operational requirements [5] through to the selection [30] of the preferred candidate at the end. In the hands of a systems engineer the design sequence is quite close in form to decision analysis as:

(i) detailed equipment design is not being undertaken;

(ii) the system performance has to be projected forward to uncertain future environments to cover the whole-life design requirement (the Forecasting Level);

(iii) the design objectives are systems wide and must

† Numbers in square brackets refer to box numbers in Fig. 4.

‡ In Decision Analysis this would be referred to as Consequence Analysis.

§ Systemic objectives refer to the satisfaction or avoidance of specified social and environmental factors.

meet the customer's preferences—unlike the equipment designer who usually designs to a given performance specification. Thus the elaboration of the Normative Level.

Hence the System Design Framework is modelled on the analytical structure developed for the multiple objective decision problem under uncertainty or risk.[20, 23, 24]

Design Sequence

The sequence of operations in the forwards path of system design follows the similar sequence in Decision Analysis fairly closely: both are searching for the best candidate action or design to fulfil stated objectives in the future. The sequence is summarized below in analytical form.

- Let the set of generated candidates be

$$A = \{A_1, A_2, \ldots, A_i, \ldots, A_I\}. \tag{1}$$

- The designer will synthesize a configuration $\$_i(\hat{\alpha}_i)$ with the vector of design alterables $\hat{\alpha}_i$ for the candidate A_i.
- The predicted performance Y_i of the candidate will depend on
  (i) the configuration and setting of alterables;
  (ii) the environmental state N_j in which the system is presumed to be operating.
- Let the set of possible environmental states be

$$N = \{N_1, N_2, \ldots, N_j, \ldots, N_J\} \tag{2}$$

where the state N_j is a distinct future incorporating variables to which the system is sensitive but is not in the designer's control, e.g. one of several possible operational states, climate, quality of resources, prices, threats ....

- The state N_j is assumed to occur with the probability p_j at the designated future time, i.e.

$$p_j(t_f) \Rightarrow (N(t_f) = N_j | N(t_0) = N_0); \ \sum_j p_j = 1 \tag{3}$$

where (N_0, t_0) is the initial environmental state. Denote the set of probable future environments by $\{p_j, N_j\}$.

- Let the set of performance functions be

$$Y_i = \{Y^i_1, Y^i_2, \ldots, Y^i_n, \ldots, Y^i_N\} \tag{4}$$

for the candidate A_i.

- Let the level of the nth performance function, given that the jth environment exists, be y^i_{nj}. This is a design outcome.
- Then the expected value of the nth performance function for the ith candidate with configuration $\$_i$ and alterables at $\hat{\alpha}_i$ evaluated over the range of possible environments is given by:

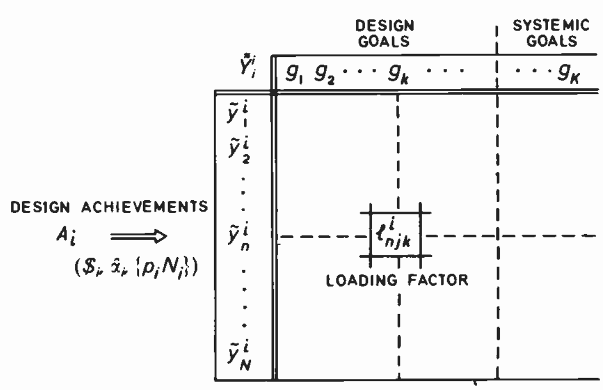
$$\tilde{y}^i_n \Rightarrow (y^i_n | (A_i, \$_i, \hat{\alpha}_i), \{p_j, N_j\}) = \sum_j p_j y^i_{nj}. \tag{5}$$

- Let the set of expected performance levels be

$$\tilde{\tilde{Y}}_i = \{\tilde{y}^i_1, \tilde{y}^i_2, \ldots, \tilde{y}^i_n, \ldots, \tilde{y}^i_N\}. \tag{6}$$

For example, in a servo design, the performance functions would be the expected response time, stability factor, accuracy, reliability, energy consumption, cost ... for a specified range of possible operational environments.

- The expected values of the performance functions (called the design achievements) now have to be screened [22] against the design goals and systemic objectives $\vec{g}$. The screening is effected via a loading matrix (Fig. 5) in which each design achievement is assessed in turn against each design goal and systemic goal. The assessment is subjective and is expressed as a loading factor l^i_{njk} entered in the cell recording the loading (or impact) of the achievement $\tilde{y}^i_n$ on the goal g_k. The loading factors range between 0 and 1 for very low/high impacts. They may be scores or drawn in the form of a loading function that rests on the same axiomatic basis and construction as the utility function in Decision Analysis (e.g. Refs. 4, 14, 18, 19). The completed loading matrix is signified as L^i_{NJK}.



The loading factor may be in the form of:
(i) a score, e.g. 1 for very high impact, 0 for very low impact;
(ii) a loading function, e.g.:

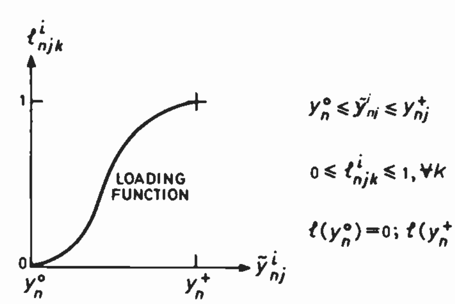


Fig. 5. Loading matrix and function (see also Appendix 3).

Screening is effected as follows:

- Since all goals g_k are expressed in desirable form (high capability, high pollution avoidance), each g_k column should have at least one high loading factor l^i_{njk}.
- A column with many high l^i_{njk} is no more effective than one with a single high factor, but it signifies a better A_i with respect to that goal g_k in that many of the achievements contribute towards g_k.
- Achievements with no high loading factors in their row are probably mismatched or redundant.
- Achievements with many high factors in their row are important, perhaps too much so as goal satisfaction should not depend on too few achievements.
- Candidates with one or more g_k columns empty or with only a few low factors should be discarded or designed to increase the loading of g_k.

The diagonal loading matrix has $l^i_{njj} = 1$, $l^i_{njk} = 0$, and $J = K$. This signifies a unique 1 : 1 relationship between the achievements and the goals. It is a special case that occurs quite often in engineering design where the design objective is expressed only in terms of desirable levels of design achievement. But it is by no means general as the design achievements are instrumental values only with respect to the general objectives hierarchy.

- The load matrix has now to be evaluated [25] with respect to the objectives hierarchy [10] and customer's preferences [23] to obtain an overall figure of merit, or worth [28] for the candidate design A_i. This requires the design of a multi-attribute value criterion [24] that matches the multiple objectives and customer's preferences.[25] It is not a straightforward matter and will be left till later.
- The Design Problem facing the designer can now be expressed quite simply:
  - (i) Find that configuration $\mathcal{S}_i$ and alterables setting $\hat{\alpha}_i$ that maximizes the worth of each candidate design A_i when evaluated over the appropriate range of future environments, using a value criterion V that is matched to the objectives G and the customer's preferences.
  - (ii) Each candidate design must be feasible and, when optimized, must not exceed any of the stated constraints.
  - (iii) The best candidate design with respect to V is that which has maximum worth W relative to the objectives set G:

    The design

$$A^* \Rightarrow \max_{\forall i} W^*(A_i) \tag{7}$$

where

$$W^*(A_i) = \max_{\mathcal{S}_i, \hat{\alpha}_i} V(L^i_{njk}|G, \{p_j, N_j\}) \tag{8}$$

and

$$B^i_m \leqslant B^+_m, \ \forall_m; \qquad K^i_p \leqslant K^+_p, \ \forall p. \tag{9}$$

The process of finding the maximum worths is an iterative constrained optimization search [26]. B_m and K_p are resource and cost constraints.

The Design Cycles

The concept of the design cycle or spiral is well established due to the essential feedback and iterative nature of design.[9,26] The System Design Framework has several important cycles imbedded in its structure:

I Feasibility screening [16, 17, 18, 19, 16] eliminates configurations and alternatives that have little chance of producing the desired performance or that contravene a constraint.

II Impact screening of the candidates [16, 17, 22, 18, 19, 16] is required to eliminate those that have only a weak impact on the objectives or, conversely, to generate candidates that have strong desirable impacts and weak undesirable impacts.

III It may be that a strong constraint has become a barrier, e.g. a technological barrier [20] which the state-of-art [6] cannot overcome. This leads to the R & D loop which may either remove the constraint [18, 19, 20, 21, 20, 14, 18] or suggests a new approach following a breakthrough [16, 17, 18, 19, 20, 21, 20, 14, 18, 16]. Of course the R & D loop can be indulged in only if research resources and time are available. If not, the candidate has to be discarded as unfeasible, and another one studied ... unless the additional costs and delays are accepted.

IV This is the familiar optimizing loop [16, 17, 22, 25, 26, 18, 19, 16] in which the alterables α_i of a particular candidate A_i are tuned to achieve the best performance or trade-offs between the objectives.

V Sensitivity analysis [27] is desirable after optimization to ensure that the design configuration is not too sensitive either to parameter variations within the system or to environmental parameters. If sensitivity is noticed then another cycle is required to design it out.

VI When a particular candidate A_i has been explored and optimized the next one A_j is subjected to the design process until the set of optimized and ranked alternatives is complete (enough) and submitted to the customer for selection.

The design cycles are really a misnomer as the designer never returns to 'square one'. For one thing time is always ticking away so that the cycle process would have a waveform when plotted against time. And the designer always knows more about the problem at the end of each cycle. A better way of representing the design cycles is in the form of a spiral.

Figure 6 takes the six cycles of Fig. 4 and spins them out as a spiral sectioned into the six design stages. In this representation the design advances in time along the spiral moving from the centre outwards. Each stage is visited at least five times as the design elaborates *even if there is no recycling*. The cycles are shown dotted and slip from an outer to an inner spiral for a repeat of the contained design process, but this time with modifications, improvements or just clearer insight. Each time the designer visits a stage he has the advantage of the knowledge acquired during his previous visit to that stage. In some sectors the radial arrows pointing outwards indicate where the knowledge reinforcement is particularly important. The reader is invited to walk carefully along the spiral, relating it to his own design experience.

The structure of design is an important matter to the systems engineer. No longer is design in the personal control of a chief designer and a small team undertaking design in a 'natural' manner. It is a large scale activity requiring hierarchical management control to ensure coherence across all subsystem aspects (Fig. 1 again). Consequently the system engineer becomes a keen student of the designer's art so that he can formulate a comprehensive structure for the design process with tasks, information inputs and outputs, interactions, etc., all itemized and properly sequenced. On this framework successful design management can be organized. (The next paper in this series looks at Design more closely.)

INTEGRATED SYSTEM DESIGN (ISD)

There is a dotted rectangle in the centre of Fig. 4 marked ISD. This boundary marks off the outer more systems-analytical parts of the design framework from the inner more detailed parts. The outer parts may be identified as:

Applied Systems Analysis for Systems Design

Problem survey
Objectives setting and criterion design
Research
Scenario generation and likelihood estimation
Overall assessment of system worth and selection

The inner part defines the process where the systems designer is working hard to provide the design framework within which the subsystem and detail design teams should operate. This may be described as:

Integrated Systems Design

Performance design
Reliability design
Support design
Integration through mutual trade-offs
Optimization through cost-effectiveness analysis

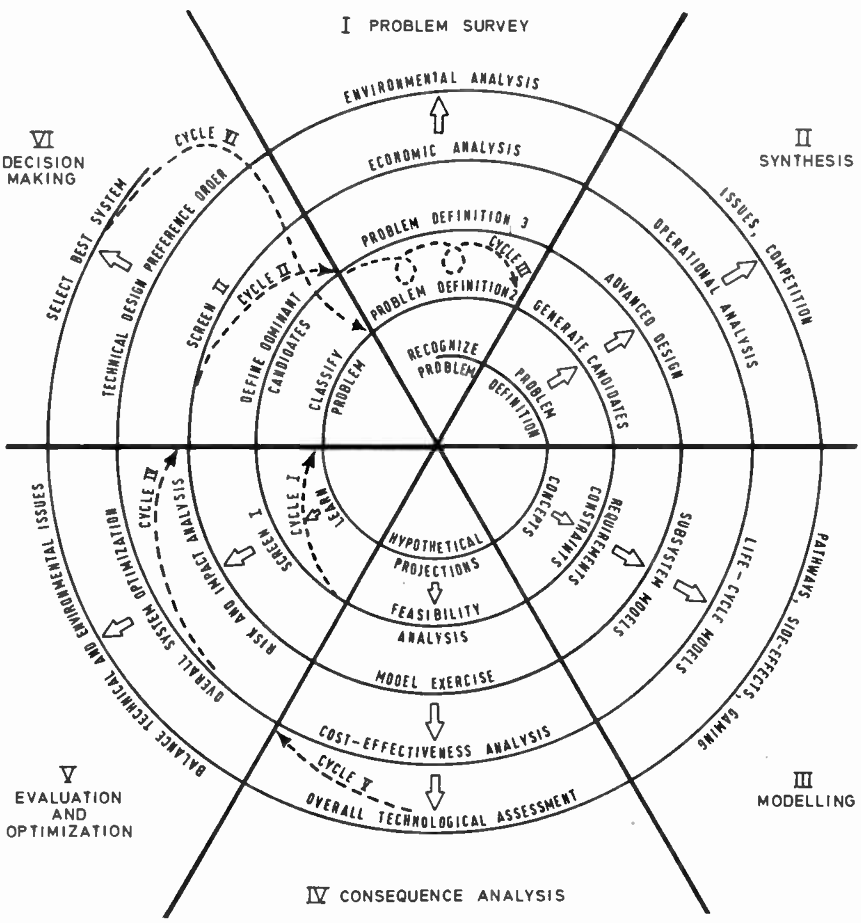


Fig. 6. The system design spiral.

ISD is the detailed implementation of the hierarchical and lateral form of systems design displayed earlier as Fig. 1.

The System Engineer's Model for an Eventful Operational Life

Just as the control engineer designs a control system using a dynamic model of the controlled process as a test-bed, so the systems engineer has to predict the probable dynamic behaviour of the whole-system during future operational cycles if he is to bring whole-life considerations into the analysis. A dynamic operational model of the whole-system has to incorporate the interdependent dynamic behaviour of both the functional processes and the support system over projected future operational cycles in a sufficient variety of environmental scenarios to cover the range of possible futures. The dynamic functional behaviour of any system—large or small—can be represented by the familiar state equation:[27]

$$\frac{d}{dt}\hat{X} = f(\hat{X}, \hat{M}, \hat{N}) \qquad (10)$$

where $\hat{X}$ is the vector of functional state variables, $\hat{M}$ the vector of control inputs and $\hat{N}$ the vector of environmental states to which the system is sensitive.

But equation (10) is valid *only if* the functional or operational system is assumed to be in a satisfactory working condition. The implications of this assumption are clear from Fig. 7(a) which shows the four essential subsystems necessary to ensure that an operational system *continues* to do what is required of it. The usual control and information feedback loop is shown. But the operating parts need:

(a) the supply of logistic resources $\hat{B}$ (fuel, power supplies, spares, provisions, personnel);

(b) the attention of support cycles (maintenance and repair) to counteract both the inevitable degradation of performance due to wear and tear, and the possibility of complete breakdown and failure.

To deal with the dynamics and costs of the support system an operational state model has to be brought alongside the functional state model for $\hat{X}(t)$. The operational state $Z(t)$ is one of the many possible states that the operational system can occupy depending on its state of readiness and degree of malfunction.

Thus:

$$z(t) \in Z \qquad (11)$$

where

$$Z = \{\{Z_S\}, \{Z_U\}\} = \{z_1, z_2, \ldots, z_r, \ldots, z_R\} \qquad (12)$$

with z_1 as the 'as-good-as-new' state, higher indexed states signifying increasing degradation of performance; $\{Z_S\}$, $\{Z_U\}$ are respectively the sets of satisfactory states and unsatisfactory states for the operational system.

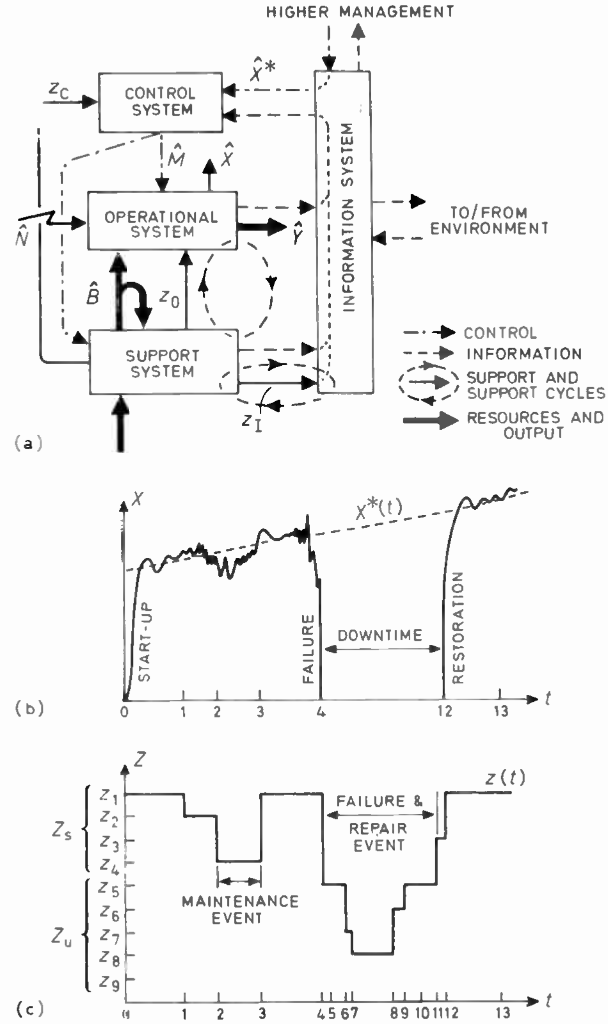


Fig. 7. Functional and operational states of a system.

Accordingly the proper state equation for the operational system is:

$$\frac{d}{dt}\hat{X} = f(\hat{X}, \hat{M}, \hat{N}) \quad \text{iff } z(t) \in \{Z_S\}. \qquad (13)$$

The relationship between the two equations (10) and (13) is demonstrated in Fig. 7(b) and (c) in which a rather dramatic slice of a system's life is shown. Figure 7(b) traces the functional state vector $\hat{X}(t)$ in time: it is supposed to follow the desired state $X^*(t)$. The system starts up satisfactorily at time t_0, there is some degradation of performance during the interval t_2–t_3, but the system returns to normal. However, at t_4 there is a catastrophic failure and the system is non-operational for quite a period while it is repaired and restored to the operational state at t_{12}. The equivalent operational states $z(t)$ are shown in the profile of Fig. 7(c) and described in Table 3.

As far as management was concerned the system was required to be operational and 'on-load' for the whole of the interval t_0–t_{13}. But the failure event made it non-operational between t_4 and t_{11}. At t_{11} it was returned to

Table 3

Time	Events	Operational state		
t_0	Start-up + operation	Z_1	Z_S	Operational
t_1–t_2	Degradation of functional state	Z_2		
t_2–t_3	In-service maintenance	Z_4		
t_3–t_4	Operation	Z_1		
t_4	Breakdown	$Z_1 \rightarrow Z_5$	Z_U	Failed
t_4–t_5	Waiting for attention	Z_5		
t_5–t_6	Removal to repair facility	Z_5		
t_6–t_7	Diagnosis (fault recognized)	Z_7		Failed and under repair
t_7–t_8	Under repair	Z_8		
t_8–t_9	Testing (found OK)	Z_6		
t_9–t_{10}	Waiting for attention	Z_5		Repaired and waiting
t_{10}–t_{11}	Returned to operational site	Z_5		
t_{11}–t_{12}	On standby	Z_3	Z_S	Standby
t_{12}	Start-up + operation again	Z_1		Operational

site in an operational state but not brought back on load until t_{12}. Thus the operational system was available for use (being in a satisfactory operational state) between t_0–t_5 and t_{11}–t_{13}.

The *Availability* of the system during this period is:

$$A_\infty = \frac{\text{actual time available for operation}}{\text{total time required for operation}} = \frac{T-DT}{T} \quad (14)$$

where

$$T = \text{total time} = t_{13} - t_0$$
$$DT = \text{down time} = t_{11} - t_5.$$

(Note the source of an argument. As far as the support system is concerned the operational system was 'down' from t_5 until restored to site in an operational state at t_{11}. But to the operators the system was 'down' until t_{11}.)

Clearly a high availability is desirable, implying a low failure rate and efficient (fast) repair. The economic implications of the events of Fig. 7 are shown in Fig. 8(a) for an operational system designed for steady operation with planned maintenance. The actual costs, however, are increased by the unreliability loss occasioned by the failure event. The cost components are shown in Fig. 8(b).

The planned costs include operational and planned maintenance, capital charges and overheads. Note that there is a 'non-operational gain' in that operational costs stop during downtime. But these are more than offset by the downtime penalty (loss of revenue, fines, etc.). The unreliability loss is equal to the total cost of the repair event plus downtime penalty minus downtime gain. It represents a loss that might be avoided through better system design and management.

Model of Systems Effectiveness

Clearly an effective system implies a system with high availability. In turn, high availability implies:

- operational units with high reliability to reduce the need for maintenance and the risk of failure;
- an efficient support system that reduces downtime for maintenance and repair by providing:
  - a fast and efficient maintenance and repair service;
  - a logistic service that ensures the supply of operational resources and spare parts.

An effective system must, of course, also be able to meet its operational requirements. The concept of Systems Effectiveness has been developed to provide a figure of merit that combines the three essential attributes of a

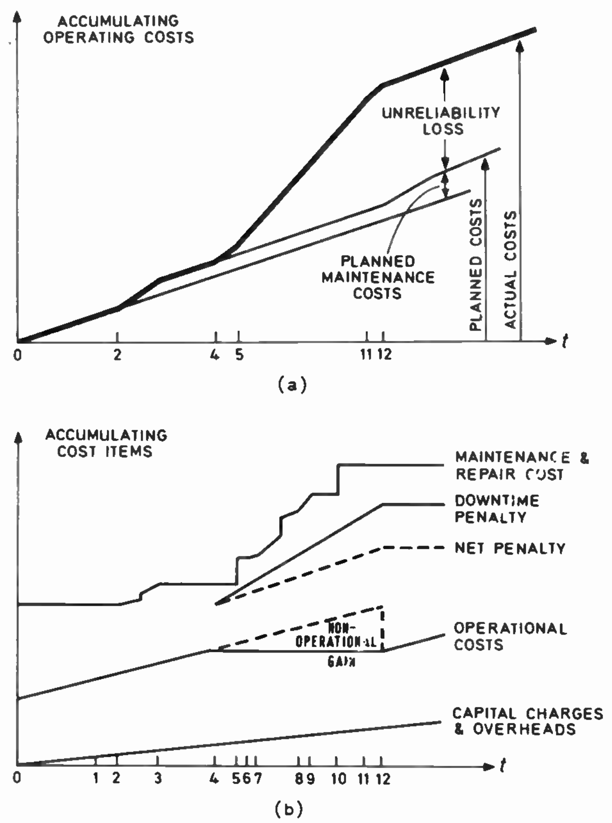


Fig. 8. Breakdown of operating costs:
(a) Accumulating operating costs.
(b) Accumulating costs items.
Note. (a) and (b) are not to the same vertical scale.

'good' system:

- High *Availability* so that the system is ready for operational use when required.
- High *Dependability* so that the system, once in operation, can meet the operational stresses with little risk of failure, i.e. it is reliable.
- High *Capability* so that when the system is in operation it can meet the operational requirements under the prevailing environment conditions.

The concept comes from the systems technology developed to provide the American armed forces with weapon systems that are as strategically and tactically effective as possible for the millions of dollars that they cost.[28–30] It has remained a high technology and military related concept (weapons, aerospace, nuclear power), but is none the less highly relevant as a figure of merit for 'good' design in any technological application. It filters through to the industrial sector and is frequently reinvented in the systems engineering variants of Industrial Engineering, Systems Assurance, Terotechnology and Design Audit.

The previous Section will have shown that the model of operational states is dealing with probabilities only: the operational profile $z(t)$ in Fig. 7 is discontinuous with a stochastic pattern of transitions depending on the occurrence of uncertain failure events and then on repair events of uncertain duration. Thus it is usual to define Effectiveness in terms of probabilities:

For a system with n operational states:

- Availability is the vector of probabilities a_r that the system will be in state z_r at t_0

$$\mathbf{A} = [a_1, a_2, \ldots, a_r, \ldots, a_R], \quad \sum_r a_r = 1. \quad (15)$$

  Availability answers the question 'will the system be working'?

- Dependability is the $R \times R$ matrix of state transition probabilities d_{ij} such that d_{ij} represents the probability that the system will be in state z_j given that it was in state z_i at t_0.

$$[D] = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1R} \\ d_{21} & d_{22} & \ldots & d_{2R} \\ d_{R1} & d_{2R} & \ldots & d_{RR} \end{bmatrix}, \quad \sum_j d_{ij} = 1 \quad (16)$$

$$d_{ij}(t_0, t) \Rightarrow p(Z(t) = z_j | Z(t_0) = z_i). \quad (17)$$

  Dependability answers the question 'will the system go on working'?

- Capability is the column vector whose elements c_r represent the system's ability to meet its objectives when in state c_r under stated environmental conditions.

$$\mathbf{C} = [c_1, c_2, \ldots, c_r, \ldots, c_n]^T, \quad 0 \leqslant c_r \leqslant 1. \quad (18)$$

  Capability answers the question 'if the system is working, how well is it working?'.

Clearly Effectiveness is going to be a number between 0 and 1. A system could not be more effective than one that is always available, completely dependable and capable. Such a system will have an effectiveness of 1. If any one of the attributes is zero, effectiveness must also be zero: e.g. a completely dependable and capable system is useless if it is never available. Thus the combination of the three attributes is multiplicative:

- Effectiveness is the probability that an operational system will satisfy its objectives under stated operating conditions throughout a given period.

$$E = \mathbf{A}[D]\mathbf{C} \quad (19)$$

and since

$$0 \leqslant \begin{Bmatrix} A \\ [D] \\ C \end{Bmatrix} \leqslant 1, \quad \text{then } 0 \leqslant E \leqslant 1. \quad (20)$$

The computation of effectiveness as defined above requires a certain amount of mathematical analysis on the state-transition model that computes the probabilities of the system being in any of the z_r operational states. The mathematical basis of state transition models and the effectiveness calculation are outlined in Appendix 1 where it is shown that:

- The elements c_r of the capability vector are calculated (estimated) from the system's functional outputs degraded as necessary for the various sub-operational states. Thus capability is primarily a function of the Performance Design (PD) of the operating system, but Reliability Design (RD) also influences it as it is sensitive to performance degradation due to unreliability.
- The elements d_{ij} of the dependability matrix are related analytically to the state transition probabilities which are quantified in terms of the failure and repair rates designed into the operating and support systems. Thus Dependability—which is a measure of the system's ability to resist performance degradations during service—is a function of Reliability Design (RD) and Maintenance Design (MD).
- The elements a_r of the availability vector reflect past operational history of the system including the need for major repairs and overhauls; in fact $\mathbf{A}(t_0)$ is the solution to the state probability equation from the start of the operating cycle to t_0. Inherent availability $\mathbf{A}(\infty)$ indicates the proportion of time during the whole-life that the system is likely to be in an operational state. Availability is a function of RD, MD and overall Support Design SD.

The effectiveness equation can be elaborated further in accordance with the operational scenario. For example, one may include additional states to represent the existence of undetected or erroneously diagnosed faults with which to explore the behaviour of a system

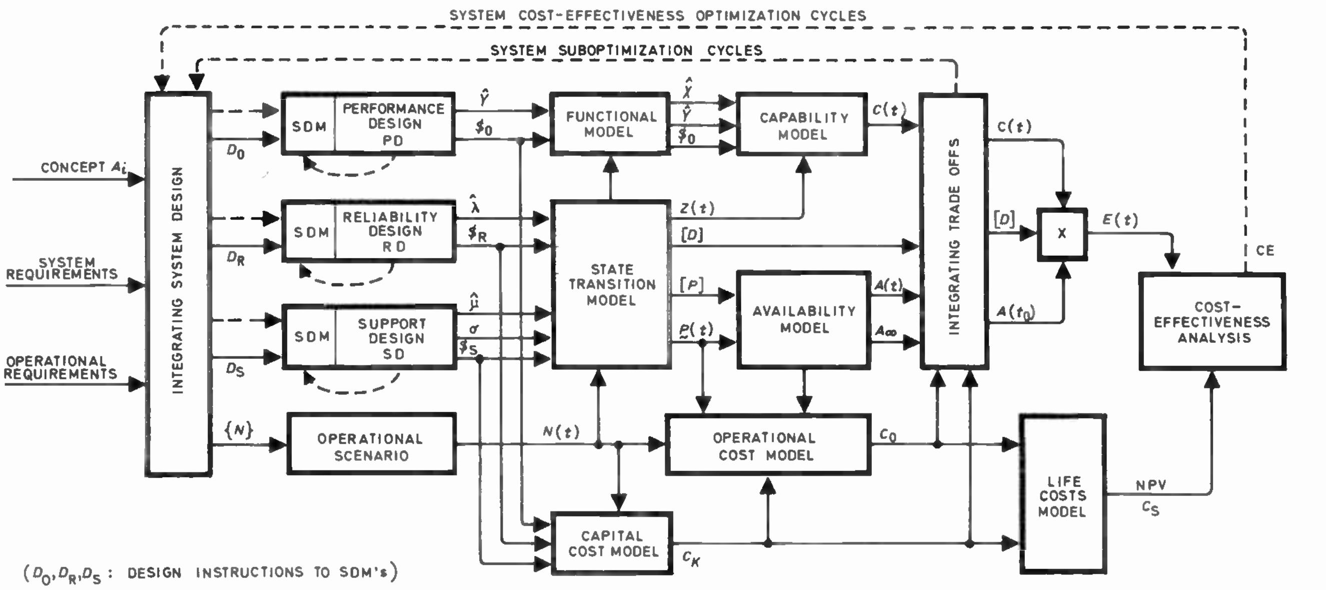


Fig. 9. Framework for integrated system design.

operating in a true state when the operators think that the system is occupying another (false) state. Such procedures are useful for establishing diagnostic routines or designing condition monitoring systems. A good diagnostic or monitoring system ensures a minimum divergence between true and false states.

The effectiveness indicator as defined above provides a comprehensive and sensitive figure of technical merit that integrates the various attributes of Performance, Reliability, Maintenance and Support Design as they influence the system's performance in prescribed operational environments. The price of the indicator is a certain amount of not too difficult mathematical modelling followed by sufficient computer runs using quite a powerful machine to calculate the life-time effectiveness. It is not as bad as it sounds as it need be used only as the final overall system optimization to refine the suboptimum designs. Nor does the early criticism of this approach hold any more:[34] it is quite valid provided it is limited to quantifiable engineering problems (as in this paper), and the advance of computing power and availability over the last 20 years means that the technique is within the reach of any competent project manager.

Table 4

System Whole-Life (ownership) costs $C_S = C_K + C_O$

CAPITAL COSTS

$$C_K = C_{RL} + C_{AQ} + C_{KC}$$

Realization costs .... C_{RL}		as required for:
Survey, Preliminary design	Phase	R + D
Acquisition costs .... C_{AQ}		Operational System
Development, Manufacturing, Assembly, Test, Deployment, Training	Phase	Support System
Capital charges .... C_{KC}		Phase M + I†
		Phase Personnel
		Phase Resources + Support

OPERATING COSTS

$$C_O = C_{OS} + C_{SS} + C_{Eth}$$

Operating system costs .... C_{OS}

Equipment:	Resource use, (Support), Modifications, Renewal, Phase-out
Personnel:	Wages, Benefits, Training, (Support), Retirement, M + I
Operations:	Dues, Licences, Facilities, M + I
Marketing:	Research, Advertising, Sales, Personnel, M + I
Intelligence:	Equipment, Personnel, M + I

Support system costs .... C_{SS}

Personnel:	Victualling, Transport, Health, Recreation, Lodgings, M + I
Maintenance:	Personnel, Monitoring + Diagnosis, Scheduled Maintenance, Equipment, Facilities, Downtime Costs, M + I
Repair:	Personnel, Diagnosis, Unscheduled Maintenance, Equipment, Facilities, Downtime Costs, M + I
Logistics:	Personnel, Spares + Stores Inventories, Transport, Facilities, Procurement, M + I

Ethical costs................ C_{Eth}

Safety:	Monitoring + Alarms, Equipment, Personnel, Damage, Liability, Insurance, M + I
Environment:	Amenity, Public Relations, Pollution Abatement, M + I

† M + I = Management and Information.

Framework for Integrated System Design

The analytical structure of the effectiveness calculation defines an appropriate framework for Integrated System Design (ISD). This is shown in Fig. 9. The effectiveness model follows on the subsystem designs for PD, RD and SD. Capital and operational cost models are added to allow the final CE analysis. The cost models are not elaborated here—they are standard; Table 4 summarizes the main cost headings in overall whole-life ownership costs. These costs are, of course, discounted to their Net Present Value (NPV) for the CE analysis. Comparing Fig. 9 with Fig. 1 will show how well the system cost-effectiveness approach satisfies the requirement for a closely integrated system design hierarchy.

The Inter-Subsystem Trade-Offs

Another problem in overall cost-effectiveness design is the coupling that exists between the various design loops: local optimizations arrived at independently for performance, reliability, maintenance and logistics do not necessarily mean that the whole-system has optimal cost-effectiveness overall. The couplings present in the design process of Fig. 9 are indicated in the matrix of Table 5.

The interactions between RD, MD and LD are complex as shown in Fig. 10. This is an implication diagram tracing the effect of an increase in MTTR but keeping inherent availability constant at some desired value A^x_γ.

(Note: for fixed A_γ, MDT $= A/(1-A)$ MTTR.)

Variables that increase are shown shaded, the half-shaded ones could rise or fall depending on the balance of effects. The design variables controlled by MD are MDT and the preventive maintenance inspection interval T. As a result of RD's action in increasing MTTR, MD has to increase MDT if A^x_γ is to stay constant. Because of the rise in MTTR, MD can lengthen the inspection interval T so reducing inspection costs C_T. The risk of failure n_F is reduced if MTTR rises and increases if T lengthens. With luck this conflict might balance out. Moving down to LD, the rise in MTTR means that the demand rate for spares σ drops so that LD can reduce the spares inventory I, reducing the costs C_I. Another conflict arises as reducing σ will lower the probability of stock-out π with fixed inventory, while reducing the inventory will raise π. If the probability of stock-out π rises the mean down-time MDT may also rise due to the extra waiting time implied by the rise in π.[5,35,36]

Table 5

	Cost-effectiveness										
	Effectiveness			Life-cycle costs							
Design	Availability	Dependability	Capability	Capital	Operating	Risks	Maintenance	Logistics	Downtime	Modification	Phase-out
PD			×	×	×	×				×	×
RD	×	×		×		×	×		×	×	×
MD	×	×		×		×	×	×	×		
LD	×			×			×	×	×		

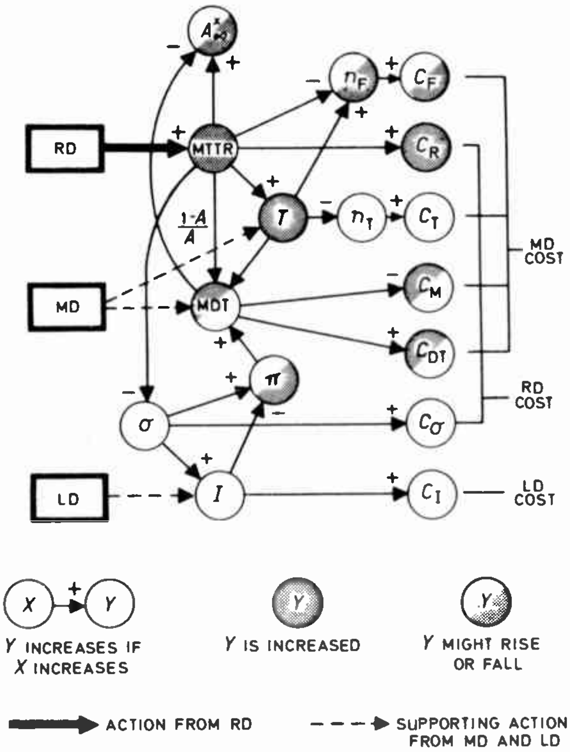


Fig. 10. Interactions between reliability, maintenance and logistic designs.

A^x_γ	Desired inherent availability	I	Spares inventory
MTTR	Mean time to repair	C_F	Failures and repair costs
MDT	Mean down time	C_R	Reliability cost
T	Inspection interval	C_T	Inspection cost
n_F	Number of failures	C_M	Maintenance cost
n_I	Number of inspections	C_{DT}	Downtime penalty
σ	Spares demand rate	C_σ	Spares cost
π	Probability of stock out	C_I	Inventory cost

The marginal cost following the increase in MTTR is

$$C(\delta\text{MTTR}) = +\delta C_R \pm \delta C_M \pm \delta C_F \pm \delta C_{DT} - \\ -\delta C_T - \delta C_\sigma - \delta C_I.$$

If the negative signs win, the increase in MTTR is cost-effective because the cost of achieving the availability A^x_γ has been reduced. Further increases in MTTR may then be tried until a minimum availability cost is found.

The above exercise has been a demonstration of a typical trade-off between the Reliability, Maintenance and Support Designs. It is obvious that the sub-system designs cannot be optimized independently, consequently the trade-offs have to be done at a superior level in the design hierarchy. Moreover, by integrating the trade-offs at system level the sub-system designs are brought into close coordination with each other. (The detailed methodology of Reliability and Support Design will be studied in a later paper in this series.)

Trade-offs can often be entertained between Performance and Availability. For example, it might pay to lower the standard of functional performance requirements and use the consequent capital cost reduction to increase reliability (say) in the interest of higher operational availability. Which means a more cost-effective system overall, but a non-minimum cost system with respect to the purchase price (capital costs). What does the customer really want? The cheapest system that he can buy or the cheapest system to own and operate? The answer will depend on the short or long sightedness of the customer.

EVALUATION OF SYSTEM WORTH

General Value Attributes for a System

The final stage of every design cycle is an assessment of the design achievement with respect to the specified objectives of the system under study. The assessment is the designer's attempt to answer the question 'how good is the system?', and goodness has many dimensions in a systems engineering context:

- whole-life cost-effectiveness consequent on the technical design;
- commercial attractiveness as predicted from market forecasts and economic analyses;
- social benefit dependent on the level and distribution of services or job-opportunities that the system may provide in a particular location;
- environmental impact resulting from the disturbance (or otherwise) that the insertion of the technological activity may have on the surrounding ecosystem;
- safety level taking into account the consequences of possible accidents and the risks to the health of the operators and the surrounding population.

The first two dimensions (cost-effectiveness and commercial attractiveness) are commonly encountered as part of most engineering project assessments. But the remainder (social benefit, environmental impact, safety level) loom more largely in 'big' technology and public utilities. For example, the safety and siting of nuclear power plant are studied meticulously as part of the overall power system concept. But such care is not always exercised: a glaring example is to be found in the transport system for bulk oil at sea, e.g. the combination of very large oil tankers and the ineffective policing of navigational, operational and safety procedures producing hazard levels that are increasingly unacceptable on commercial, social and environmental grounds.

The overall dimensions of system 'goodness' are the value attributes of the system that constitute what the customer, contractor, designer, operator, user and the watchdog authorities acting on behalf of society and nature would regard as desirable objectives. A system design may be examined for its contributions to those value attributes; the overall aggregate of the contributions will be called 'worth'. Table 6 shows a general structure for the value attributes of a system arranged in the form of an Objective Tree. The overall System Worth $W(A_i)$ for the ith system concept is conveniently compartmented into Principal Worths:

- Technical Worths — Technical Worth w_T; Commercial Worth w_C
- Ethical Worths — Social Worth w_S; Environmental Worth $w_{N \cdot}$,

The principal worths form the major branches of the tree which further subdivide into minor branches formed from the:

- Principal Objectives (which should be accessible to quantification)

and thence down to the

- Principal System Attributes (which should be compositions of the specified design goals of the system).

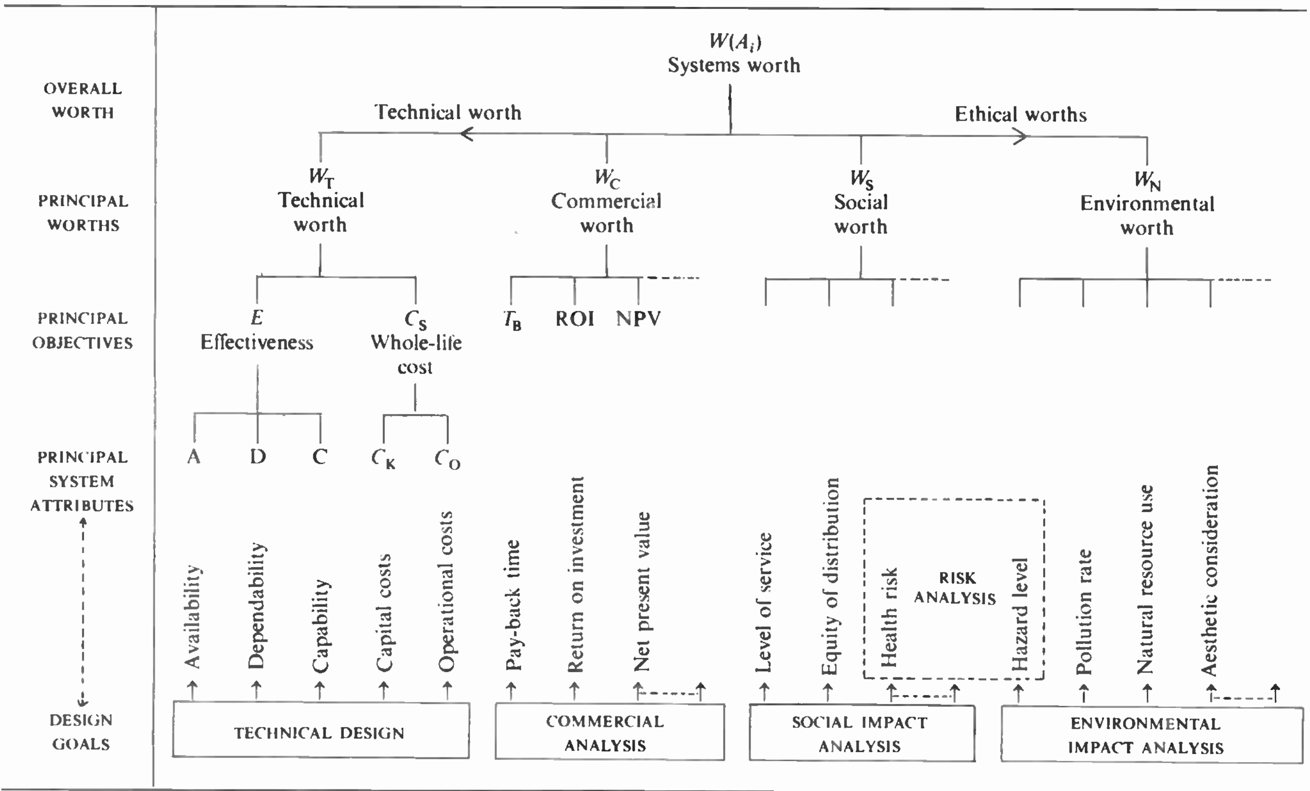
The tree structure in Table 6 provides the skeleton of a general value criterion with which to assess the overall worth of a system. Clearly there is no standard tree structure: different problems, systems, contexts, assessors will produce different trees. But the point is made that the design of a comprehensive value criterion for each overall system assessment is an essential part of Integrated System Design. Traditionally engineers and economists (i.e. the technicians) have attended to the technical worths and have left the difficult problems raised by social and environmental considerations to political judgement aided by systems analysis or cost-benefit analysis. But the systems engineer cannot ignore the ethical worth dimensions if he is to conduct a proper analysis of whole-system whole-life operations and advise his responsible decision-makers appropriately. (The decision-maker is the authority with the responsibility for committing the resources to acquire the system and, by implication, the consequences of that system's operation.) Thus an adequate design methodology for the construction of comprehensive value criteria is a necessary part of a systems engineer's tool kit. Other things being equal, the quality of the design will depend on the quality and comprehensiveness of the value criterion.

Designing a Value Criterion

Talk of objectives and value criteria has brought the discussion to the normative level of Fig. 4 from which

Table 6

General tree structure for technological system objectives



one may infer the bare bones of value criterion design. Referring back to the Design Sequence and its related equations (1)–(9), the final stage of the exercise is represented by the optimization process of equation (8) repeated below:

$$W^*(A_i) = \max_{\mathfrak{s}_i, \hat{\alpha}_i} V(L^i_{NJK}|G, \{p_j N_j\}). \quad (8)$$

The value criterion is $V(\ldots)$. Mathematically speaking it is a function whose formulation allows the designer to search for an optimal design configuration by extremising the value function $V(\ldots)$. There are three vital problems in determining a value function:

- Provide an acceptable and testable method for assessing the merit of the design achievements (levels of the performance functions Y^i_n) with respect to their relative contributions to the stated desirable goals $\vec{g}$ in G.

  (The assessment is performed by determining (in some way) the loading factors l^i_{njk} in each cell of the loading matrix L^i_{NJK} (Fig. 5). Where necessary the uncertainty of alternative future environments N_j has to be allowed for.)

- Provide an acceptable and testable criterion that combines and aggregates the value contributions of the design along each goal in $\vec{g}$ in such a way that it both satisfies the requirements of a mathematical extremal function and reflects properly the meaning of all the terms in the objectives hierarchy $\hat{G}$.

- Ensure that the proper reflection of meaning of the terms $\hat{G}$ fully articulates the preferences of the decision maker within the structure of system objectives $\hat{G}$.

It will be realized that the above implies a value function that is something more than a simple additive account and comparison of profits, benefits and costs. A comprehensive analysis of system worth requires a criterion that is able to handle both objective and subjective value attributes. It must then combine them in a way that may have to subordinate mathematical convenience in order to produce a function that encompasses the full value context of the problem as perceived by the human observers of the value problem. As such, value criterion design is the problem of defining an appropriate multi-attribute value function for the multiple criterion decision (or design) problem.[23, 24, 25, 41, 42] A methodology for determining an appropriate value criterion for a problem in Integrated System Design is outlined below—with the aid of Fig. 4.

Objectives Setting

A prerequisite for value criterion design is a clear statement of:

- all the objectives relevant to the problem.
- an arrangement of the objectives in the form of a tree indicating relationships and dependencies between the objectives.

Table 7

Objectives for a power generation system

General objective: Provide an efficient electrical power generation system

ELEMENTS OF THE OBJECTIVES SET	CATEGORY TECHNICAL	SOCIAL	ENVIRONMENTAL
Generate power—	×	×	
• to the specified requirements and load factor	×		
• economically (costs, prices, profit)	×	×	
• reliably	×		
• controllably	×		
• safely	×	×	
• provide safe, healthy, interesting work enviroment for operators	×	×	×
• match the operation to:			
1. the labour and skills available	×	×	
2. the industrial and logistic infrastructure available	×	×	
Distribute the power generated—	×	×	
• to the specified requirements economically, reliably, etc.	×		
• fairly between the client regions and categories of customer		×	
Avoiding—			
• undue sensitivity to:			
natural calamities	×		
industrial action		×	
enemy action and terrorists		×	
• wastes and toxic discharges	×	×	×
• health risks to surrounding society		×	×
• environmental pollution			×
• aesthetic damage to the neighbourhood		×	×
• consumption of scarce resources etc.	×		×

The objectives G [9]† result from a detailed survey [1] of the customer's desires and requirements in terms of:

- Needs analysis [2]: satisfaction of basic needs for services, resource procurement, survival, avoidance of undesirabilities, etc.
- Opportunity analysis [3]: identification of technological/market/political opportunities that may exist in the future and which could be in the customer's power to capture.
- Threat analysis [4]: identification of threats from the competition (business, military, technological ...), and of technological barriers that could prevent progress, and of hazards to be avoided.

Some objectives will be clearly defined right from the start, others will emerge as a result of deeper analysis of both the customer and his problem, and the system design and its consequent problems. In less structured problems the objectives will have to be pulled out of the air with a brain-storming exercise. The end result of objective setting is a set (a euphemism for a jumble) of objectives:

$$G = \{G_1, G_2, \ldots, G_k, \ldots, G_K\}. \qquad (23)$$

Tables 7 and 8 list some of the main objectives that might occur typically for a power generation system (Table 7), or a small portable instrument (Table 8).

† Refer to box numbers in Fig. 4.

Scanning these two Tables it is clear that:

- There are multiple objectives for any non-trivial design problem.
- The objectives fall quite conveniently into the three categories: technical, social and environmental.
- Some of the objectives are easily quantifiable (mainly the technical ones), others are more difficult to define and may be entirely subjective (e.g. aesthetic considerations).
- There is interaction between the categories and between the objectives, some reinforce each other, others conflict with each other.
- Individuals are likely to give different priorities to the objectives.
- No two people, let alone expert groups or ideological constituencies, are likely to produce the same objectives set, and they are unlikely to agree on the priority order.

Determining the Objectives Hierarchy

Structure now needs to be given to the objectives set G to show the relationships and dependencies between the objective elements. The result is an Objectives Hierarchy $\hat{G}$ [10] or tree in which the objectives become more general and abstract at higher levels. The bottom level should—if possible—match directly to a precise and monitorable set of operational goals $\vec{g}$ ... which could

Table 8

Objectives for a small portable instrument

General objective: Provide a marketable instrument for profit						
ELEMENTS OF THE OBJECTIVES SET	EFFECTIVENESS C	EFFECTIVENESS [D]	EFFECTIVENESS A	CATEGORY TECHNICAL	CATEGORY SOCIAL	CATEGORY ENVIRONMENTAL
Operational performance						
Range	×			×		
Sensitivity	×			×		
Error	×			×		
Drift	×			×		
Response time	×			×		
Stability	×			×		
Reliability		×		×		
Useful life time			×	×		
Support						
Power supplies			×	×		
Maintainability		×	×	×		
Servicing requirements			×	×		
Spares requirements			×	×		
Utility						
Robustness			×	×	×	×
Portability	×			×	×	
Readability	×			×	×	
Adjustability		×		×	×	
Useability	×				×	
Safety				×	×	×
Style and appearance					×	×
	COSTS C_K		COSTS C_O			
Economics						
Capital costs	×			×		
Running costs			×	×		
Price	×		×	×		
Marketability	×				×	
Profitability			×	×		

be the design objectives [11]. If the problem is fuzzy to the extent that no tree structure is self-evident then some analytical assistance may be obtained from graph theory.[38] A tree is a directed graph in which the line segments indicate a dependency between the connected elements. Analytical methods exist for constructing objectives trees via Interpretive Structural Modelling.[7,39,40] An illustration of the development of a simple objectives hierarchy is given in Appendix 2.

The Structure of Valuation

The structure of the objectives hierarchy $\hat{G}$ provides the routes by which the value contributions from the design achievements trickle up, combine and yield an overall worth for the system A_i (with a particular configuration $\mathcal{S}_i$ and setting of the alterables $\hat{x}_i$). It enables the designer, decision-maker and customer to agree on and define both the general value content (implied by the objectives set) by which the system will be assessed and the complete set of desirable system attributes ($\vec{g}$) that the designer must try to satisfy. As a result the initial vagueness of a system design problem with its aims imprecisely articulated over many technical and ethical dimensions is clarified by the precise content and structure of the hierarchy. However, the tree structure of the hierarchy is only a skeleton on which to build the value criterion [24] and the evaluation methodology [25] that together form the key processes of valuation.

The structure of valuation is illustrated in Fig. 11 for a simple case in which only two design achievements y_1, y_2 are to be valued with respect to an elementary hierarchy in which worth is disaggregated into only one branch with three attributes g_1, g_2, g_3.

The valuation process can be separated into two quite distinct parts:

- Evaluation in which the design achievements are assessed for their value contributions to the desirable system attributes, taking uncertain future states of nature into account.
- Conflation of the value contributions by means of a *Value Criterion* into a single overall worth to

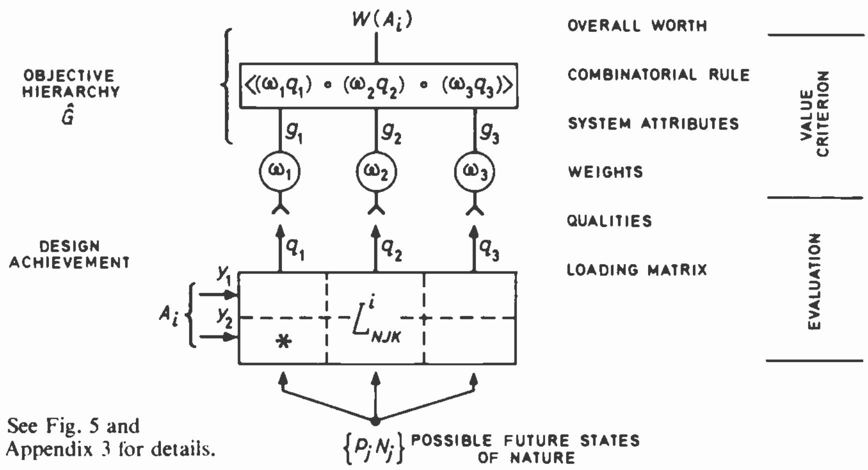


Fig. 11. The structure of valuation.

indicate the relative merit of the design A with respect to the objectives hierarchy $\hat{G}$.

There is no way that valuation can be done either objectively or independently of the decision-maker (DM)—unless all subjective attributes are neglected and the preferences of the DM with respect to the consequences on offer from the design are ignored.

The Evaluation Process

Attending first to the evaluation process, the problem here is to:

- Determine the sensitivity of the design achievements to the range of possible future states-of-nature.
- Compute expected values for the design achievement over the range of probable states-of-nature. The expected values are called *design outcomes*.
- Decide which design outcomes load which attributes.
- Assess the value contribution from each design outcome to each attribute that it loads over the range of possible outcome levels.

The assessment of the value contributions must be performed taking the DM's preferences into account. This requires a preference analysis of the DM [23] to ascertain:

- his attitudes to risky decision-making;
- his preferences with respect to the variable loading of the attributes over the range of the design outcomes;
- his assessment of the relative importance of each consequence to the attributes they load;
- his direct assessment of subjective impacts on social or environmental attributes that are not covered by analytical formulations driven by the design outcomes;
- the manner in which he senses that the value contributions combine across the range of different design outcomes to yield an overall value contribution to each attribute.

A comprehensive evaluation of N design achievements against K attributes in the context of J possible states of nature is not lightly undertaken! It involves the interleaving of many problems of subjective measurement, impact analysis, decision-making under risk and multi-attribute utility theory. In addition the DM himself (themselves) should become part of the analysis. An outline of a coherent evaluation methodology is given in Appendix 3. The methodology is culled from many sources, but it depends mainly on comparatively recent developments in multi-objective decision analysis—which is still somewhat unfamiliar territory for engineer designers and project managers.[7,23,24,25,41,42,46–52]

The end result of evaluation is a set of value contributions, one for each attribute g_k. The contributions are called 'qualities' and represent the overall goodness of the design with respect to the desired attributes. The term quality is used to signify a general utility or value function independent of how it may have been obtained during evaluation. Like utility it is a scalar between the values 0 and 1: 0 means no value contribution, 1 means complete satisfaction of the attribute g_k. And, again, like utility, it can represent by its shape the DM's preferences for attribute satisfaction via design variations. Thus the input to the Value Criterion is a set of quality contributions, one for each attribute g_k, conflated over the N performance outcomes:

$$q(A_i|g_k) = \langle (s_{1k}q_{1k}) \circ \ldots \circ (s_{nk}q_{nk}) \circ \ldots \circ (s_{Nk}q_{Nk}) \rangle \quad (24)$$

where q_{nk} is the quality contributed by the nth design outcome;
s_{nk} is a scaling factor representing the relative importance of the nth design outcome to the satisfaction of g_k

and $\langle (\cdot) \circ (\cdot) \circ \ldots \circ (\cdot) \rangle$ signifies an appropriate combinatorial rule between the bracket terms.

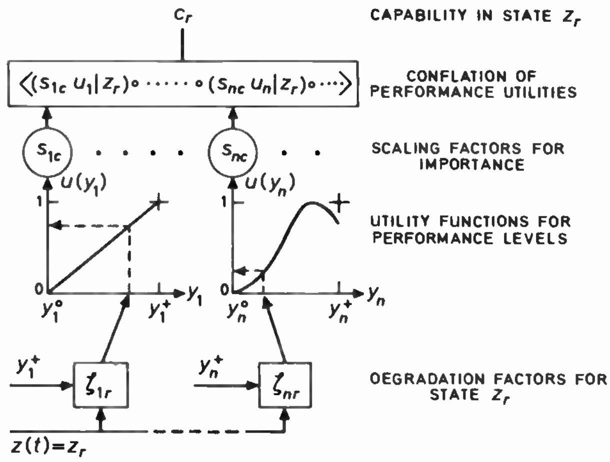


Fig. 12. Construction of a capability element.

The form of the combinatorial rule has to be determined in the context of the problem and the DM's perceptions—it cannot be assumed that qualities (or utilities, value, etc., for that matter) are always additive in the sense that such measures as money, length, force, etc., can be combined additively with themselves. Finding the appropriate combinatorial rule for multi-dimensional value judgements is the key problem in the design of a value criterion.

Effectiveness as a Value Criterion

The concept of system cost-effectiveness as the overall indicator of merit for Integrated System Design has been argued earlier (e.g. Fig. 9). Clearly effectiveness E on its own, with its three subordinate attributes of availability, A dependability, D and capability, C acts as a measure of technical worth. The analytical relationship between these variable parameters has been defined as

$$E = A[D]C \qquad (19)$$

where both A and $[D]$ are functions of the achievements of the Reliability and Support Designs, and C is a function of Performance Design (Fig. 9). It has also been shown that E, A, D, C will have values that lie between 0 and 1, so they are already scaled appropriately for worth and qualities. Relating equation (19) above to Fig. 11 one may infer that the Loading Matrix from System Design Achievement to Effectiveness will have analytical loading factors because the 'qualities' of availability and dependability are known functions of Reliability and Support Design parameters, while the loading factors onto capability may contain both analytical and subjective elements (Appendix 3). The formulation for effectiveness (equation (19)) also implies that availability, dependability and capability have equal importance to effectiveness, and the combinatorial rule for conflating the three attributes into effectiveness is multiplicative. Thus equation (19) provides a suitable combinatorial rule for conflating the qualities of availability, dependability and capability into effectiveness.

Effectiveness/Cost Ratio as a Criteria for Technical Worth

The branch of the hierarchy (Table 6) subtending from technical worth W_T contains the principal objectives effectiveness E and whole-life costs C_S. The multiplicative combinatorial rule for the three attributes A, D, C conflating to effectiveness has just been found. It is obvious that the combination of the two money attributes of C_S, namely capital and discounted operational costs C_K, C_O, requires a simple additive rule without weights: $C_S = C_K + C_O$. (Discounting of future C_O is effectively a weight reducing its importance relative to C_K.) Thus the combinatorial rules for the lower level of the technical worth branch can now be entered.

$$W_T$$

$$\langle E \quad ? \quad C\rangle$$

$$\langle A \times D \times C\rangle \qquad \langle C_K \quad C_O\rangle$$

An appropriate rule for the conflation of effectiveness and cost into technical worth remains to be found. The criterion in common use is the ratio of effectiveness to cost: the larger the ratio the 'better' the system or design.

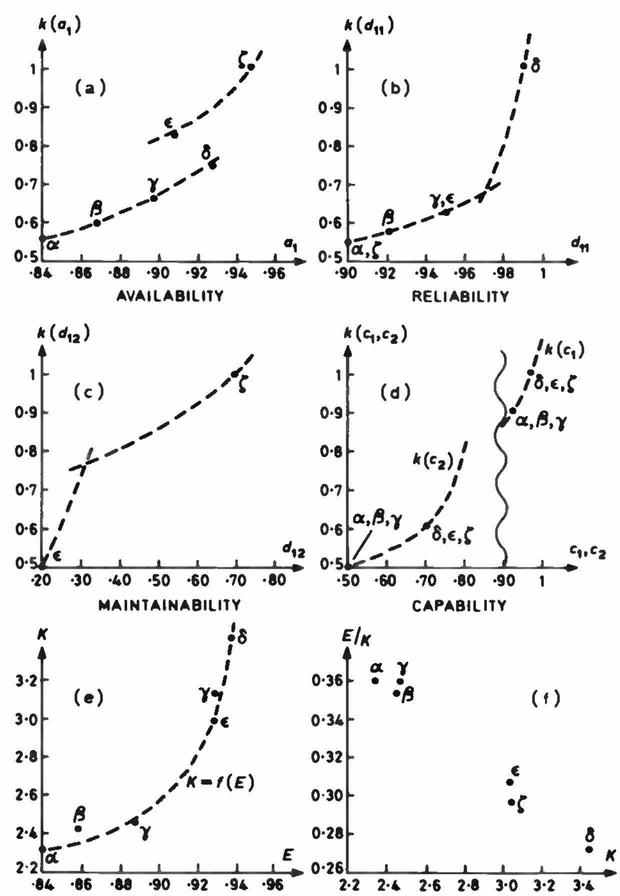


Fig. 13. Cost and effectiveness functions. (See Appendix 4.)

A standard but simple cost-effectiveness analysis is illustrated in Appendix 4. Six designs, labelled $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta$, are to be compared and placed in a preference order. The six designs represent different combinations of high and low cost solutions to the Performance, Reliability and Support Design problems. The attendant cost functions are shown in Fig. 13(a–d): they are assumed to be the results of design exercises incorporating differing concepts. The same operational and environmental contexts apply throughout. The main thrust of the exercise is to explore the trade-off between the series of designs $(\alpha, \beta, \gamma, \delta)$ which require no in-mission maintenance, and a more complex concept (ε, ζ) that involves self-maintenance during operation (without becoming non-operational). The analysis is summarized by the Table in Appendix 4 and Fig. 13(e, f). The locations of the designs in $(K–E)$ space are shown in Fig. 13(e): an approximate curve for cost as a function of effectiveness $K = f(E)$ is added—it has a typical form (which serves to justify the design scenario of the simple example). The effectiveness/cost ratios are plotted against cost in Fig. 13(f). The low-cost non-maintenance solutions (α, β, γ) form a cluster with an E/K ratio distinctly higher than the high-cost with-maintenance solutions (ε, ζ); the no-maintenance very-high-reliability solution δ is the worst. The preference order using the E/K criterion is

$$(\alpha, \beta) \succ \gamma \ggcurly \varepsilon \succ \zeta \ggcurly \delta. \tag{25}$$

In fact there would be little to choose between α, β, γ in view of the uncertainty inherent in much of the data used in cost-effectiveness analysis. But the indication is clear: adopt one of the low-cost no-maintenance solutions.

The E/K criterion is simple mathematically and produces preference orders that appear to be rational, but as a value criterion it is flawed because it does not take into account the logic of a rational DM's trade-offs between increments of effectiveness and expenditure.

Introduction of the Vector Regret Criterion

The criticism of E/K as a criterion is developed briefly using Fig. 14. Take as reference the two cost-effectiveness functions in Fig. 14(a): a linear one $K = aE$ and the more typical form $K = f(E)$ which implies that increasing increments of effectiveness that strive towards feasibility limits become more and more costly. The curves are plotted in $(K–E)$ space between feasible limits (E^0, E^+) and (K^0, K^+). The effectiveness/cost ratio K/E for the two sample curves is plotted against E in Fig. 14(b). The criterion E/K gives no answer to a linear curve $K = aE$: all points on the curve are equally preferable. The curvature of the $K = f(E)$ curve allows the E/K criterion to indicate an optimum towards the low-cost end of the spectrum of possibilities. So far the discussion has largely replicated the previous example.

Now consider the value implications of $(K–E)$ space as suggested in Fig. 14(c). Take the four corners of $K–E$ space:

- The bottom left corner has coordinates (K^0, E^0). This represents a minimum-endeavour solution with minimum effectiveness achieved consequent on minimum expenditure.
- The top right corner (K^+, E^+) is feasible (it is on the $K = f(E)$ curve) and represents a maximum endeavour solution with maximum effectiveness achieved at maximum cost.
- The top left corner (K^+, E^0) is also feasible, but it represents a complete waste of the available funds as no additional effectiveness has been achieved.
- The bottom right corner (K^0, E^+) is probably unfeasible, but it provides an ideal solution—by genius discover how to achieve maximum effectiveness at no extra cost!

A DM scanning $(K–E)$ space will soon discern that there is a value field increasing from top left to bottom right: the coordinate (K^+, E^0) is the 'worst of all worlds', (K^0, E^+) is the 'best of all worlds'. The other two corner coordinates have intermediate values (their values would be the same if the DM thinks that achieving effectiveness is as important as economizing on costs). Thus the DM is likely to prefer coordinates in $(K–E)$ space the nearer they are to the ideal solution (K^0, E^+). The perceived value field of increasing preference has provided a value criterion for CE analysis that:

- reflects a rational DM's preferences for combinations of K and E in $(K–E)$ space;
- offers a combinatorial rule that combines the contributory attributes in a manner appropriate to

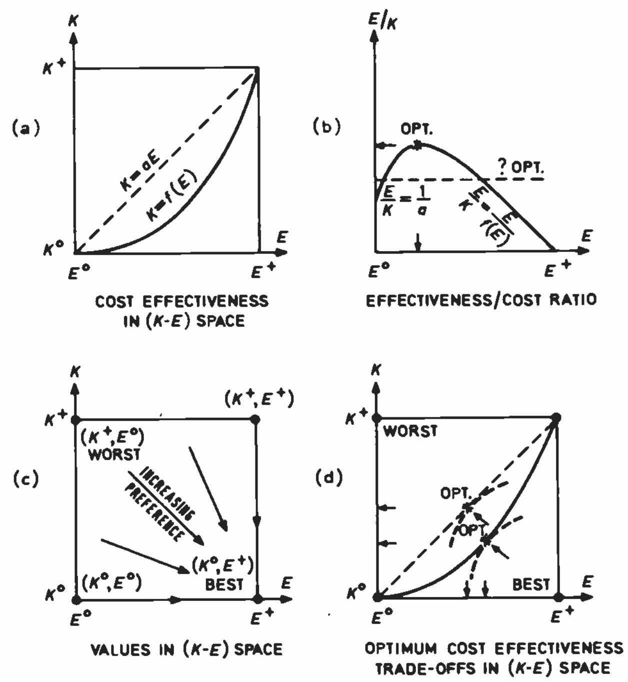


Fig. 14. Comparison of E/K and vector regret criteria.

both the technical nature of the problem and the DM's preferences.

The combinatorial rule in this case is called 'vector regret', i.e. the longer the vector from the ideal point to a coordinate in (K–E) space the greater the DM's regret at being forced to accept that coordinate as a design solution. This criterion is applied in Fig. 14(d) to the two sample cost effectiveness functions. A clear optimum is shown for both the linear and non-linear curves, being the point at which their distances from (K^0, E^+) are a minimum. (Note that the appearance of a preferred location on the straight line diagonal ($K = aE$) normal to the diagonal between the worst and best corners disposes of *addition* as a possible combinatory rule. An additive rule implies that the DM is indifferent to any trade-off between given amounts of the related parameters.)

Appropriateness of the Vector Regret Criterion for CE Analysis

Define a K-dimensioned quality space Q_K resting on a basis formed from the K orthogonal axes $q(\cdot)$ registering the contributed quality from attribute $(\cdot)$. The extreme vertices of this space are

- the origin located at the coordinate $(0, 0, \ldots, 0)$ at which all attributes have their least desirable level: the 'worst of all worlds';
- the ideal point located at the coordinate $(1, 1, \ldots, 1)$ at which all attributes have the most desirable level: the 'best of all worlds'.

Q_K space is a generalization of the two-dimensioned (K–E) space discussed earlier. It can be shown that Q_K space is an Euclidean space in which the rules of vector algebra apply. A quality vector in this space is given by

$$\hat{q} = [q_1, q_2, \ldots, q_k, \ldots, q_K]. \tag{26}$$

Each location in Q_K has an associated worth depending on the weight w_k assigned to the attributes and the combinatorial rule deemed appropriate:

$$W(\hat{q}) = \langle (w_1 q_1) \circ \ldots \circ (w_k q_k) \circ \ldots (w_K q_K) \rangle \tag{27}$$

(see also 24, 57).

In particular

$$W(\hat{q}) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \quad \text{iff } q_k = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}, \quad \forall k \tag{28}$$

Vector regret is the vector distance of $\hat{q}$ from $(1, 1, \ldots, 1)$, and is given by:[25]

$$W(\hat{q}) = 1 - d(\hat{q})$$

where

$$d(\hat{q}) = |\hat{q}^*|^{-1} \left\{ \sum_k [w_K(1 - q_k)]^2 \right\}^{\frac{1}{2}} \tag{29}$$

and

$$|\hat{q}^*| = \left\{ \sum_k w_k^2 \right\}^{\frac{1}{2}}.$$

For a simple Q_2 space

$$d(\hat{q}) = 0{\cdot}707\{(1 - q_1^2 + (1 - q_2)^2\}^{\frac{1}{2}}. \tag{30}$$

The resulting value surface $W(q)$ over the Q_2 space is shown in Fig. 15. The vector regret criterion is now applied to the CE design example of Appendix D. But first consider that the DM may have preferences for variations in E and K as they contribute to worth W_T. Representative quality functions are drawn in Fig. 16(a, b):

- $q(E)$—the quality function for effectiveness relative to technical worth represents a DM who feels that a system with $E < 0{\cdot}9$ has little technical value (within the context of the stated operational requirements);
- $q(K)$—the quality function for whole-life costs indicates that the DM wishes to avoid a high-cost solution if possible.

The locations of the six design alternatives are entered on the quality function.

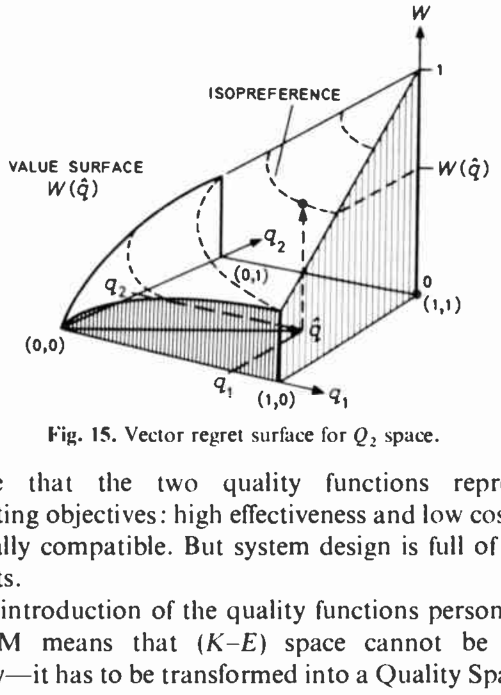


Fig. 15. Vector regret surface for Q_2 space.

Note that the two quality functions represent conflicting objectives: high effectiveness and low cost are not really compatible. But system design is full of such conflicts.

The introduction of the quality functions personal to the DM means that (K–E) space cannot be used directly—it has to be transformed into a Quality Space Q in which the trajectory of a quality vector represents the CE functional $K = f(E)$ as modified by the quality functions. The lower part of Fig. 16 represents the simple construction necessary for the transformation from (K–E) space to Q_2 space. For this simple two-dimensional value problem Q space is the plane surface on a basis of the orthogonal quality axes $q(E)$, $q(K)$. The quality vector is defined by the two coordinates

$$\hat{q} = (q(E), q(K)) \tag{31}$$

whose trajectory is the locus of $K = f(E)$ as transformed into Q. (Follow the dashed arrows from (E, K) on $K = f(E)$ round to the tip of $\hat{q}$ as drawn in Q.) It is assumed that the DM has scaled E and K equally (they

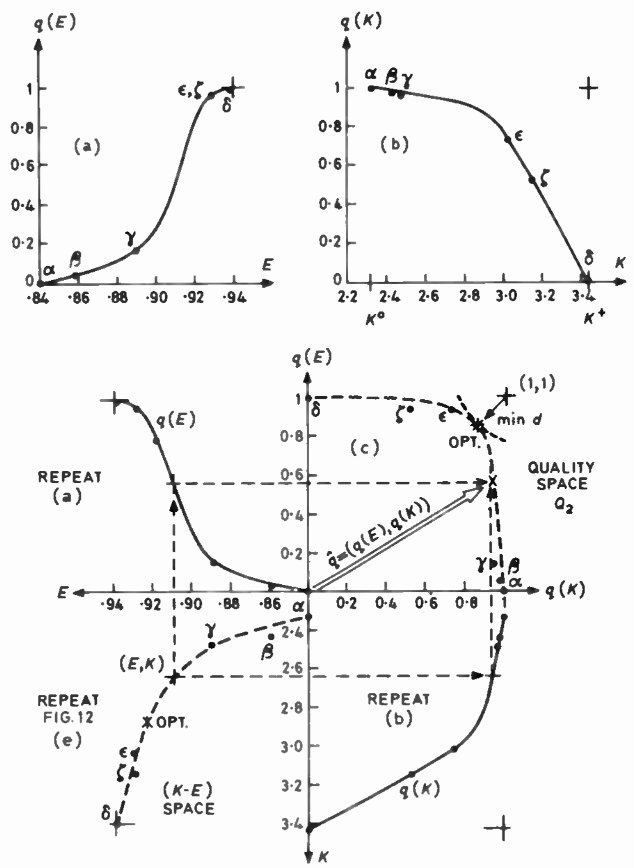


Fig. 16. Transformation of $(K-E)$ space into quality space.

are isoimportant to him).

The locations of the six alternative designs on the trajectory of $\hat{q}$ are indicated. The preference order over the six alternatives using the Vector Regret criterion is given by comparing their distances from the ideal vertex (1, 1). The two preference orders obtained from the E/K and Regret criteria are compared below:

$$\text{Vector Regret:} \quad \varepsilon \succ \zeta \ggcurly \gamma \succ \beta \succ (\alpha, \delta). \qquad (32)$$

$$E/K: \quad (\alpha, \beta) \succ \gamma \ggcurly \varepsilon \succ \zeta \ggcurly \delta. \qquad (25)$$

They are almost the reverse of each other! The regret criterion favours the more costly with-maintenance design alternatives.

Which is the 'right' criterion? The value-free E/K criterion favours low-cost no-maintenance designs because the *curvature* of $K = f(E)$ is such as to find max E/K in the low-cost region. The solution has all the phoney validity of the bargain: e.g. if apples cost 10p each, but 5 apples cost 40p, then the optimum E/K solution is to buy 5 apples. But if the DM only wants one apple to munch during his lunch-hour he will buy just one apple. The E/K criterion does not take the DM's preferences into account. In the case of the design example the E/K criterion ignored:

- the DM's preferences in $(K-E)$ space or Q_2 space;
- the DM's important judgement that only those designs with high effectiveness were 'useful'.

In the particular context of a trade-off between cost and effectiveness the vector regret criterion is arguably close to the way in which a rational DM would perceive the problem in the context of Q_k space. Thus the Vector Regret criterion is better, as it more nearly represents a DM's preferences with respect to the assessment of technical worth. If the DM accepts this argument the technical worth branch of the value criterion is completed:

w_T

$\langle w_E E$ REGRET $w_C C_S \rangle$

$\langle A \times D \times C \rangle$ $\langle C_K + C_O \rangle$

Hence

$$w_T = 1 - \{w_E^2 + w_C^2\}^{-\frac{1}{2}}\{w_E^2(1-q_E)^2 + w_C^2(1-q_C)^2\}^{\frac{1}{2}} \qquad (33)$$

where

$$q_E = q(A[D]C) \qquad (34)$$

$$q_C = q(C_K + C_O). \qquad (35)$$

The worths of the six designs are found by inserting the appropriate values from Fig. 16(a, b) in (34), (35); and assuming equal weights (as in the example), i.e. $w_E = w_C = 0{\cdot}5$.

Then:

Design $(\cdot)$	α	β	γ	δ	ε	ζ
$W(\cdot)$	0·293	0·348	0·404	0·293	0·742	0·619
					*	

In this simple two-dimensioned exercise the same preference order was found by inspection from Fig. 16(c). In a K-dimensioned value problem the topology of Q_k space cannot be visualized, so the preference order is established from the computed worths using equation (29).

The Systems Engineer and Subjective Value Assessment

The example has shown how sensitive design judgement is to the value criterion. Thus great care must be expended on this important part of the system design process. The example has also demonstrated how a value criterion can be built up step-by-step by assessing the DM's preferences over an objectives hierarchy $\hat{G}$. Only the technical worth branch has been explored in the example—but it is sufficient to demonstrate the method. Quite an extensive repertoire of rules and methodologies are available for dealing with the general multiple objective multi-attribute decision problem: the reader is left to explore the literature for himself. But let the reader be assured that the value criterion of equations (33)–(35), with its admission of quality functions to represent the DM's preferences over effectiveness and cost, and its use of the sensitive vector regret criterion in the upper branch, is about as good a criterion as can be constructed for CE analysis.

Many engineers tend to be unhappy with the subjective nature of value criterion design. They prefer the use of straightcut objective value-free criteria such as E/K. But that is to deny the essentially subjective nature of design assessment or decision-making. Whether one is the designer trying to find the best configuration, or the DM trying to determine the best alternative design, the problem of system evaluation is complex and often highly subjective. If the evaluation has to include Technology Assessment, with its careful study of social and environmental impacts, the evaluation has to become increasingly subjective. The systems engineer has an important role to play in these large scale evaluation exercises. The 'wider system' concept has brought in all those additional non-technical considerations as part of the overall integrated system design exercise. Hence, the systems engineer, rather than the specialist designers, must design the systemic value criterion that judges the system overall. In doing so he provides an important aid to the designer, decision-maker and customer when they argue over what the system is worth. The point has been made earlier—other things being equal, the quality of the design and, hence, the quality of the system's subsequent operations depend on the quality of the value criterion.

OVERVIEW

It has been a long haul. But a paper rash enough to call itself 'A Framework for Systems Engineering Design' is not going to be easy going as the systems engineer's job is to provide *complete system assurance*. And that, as we have seen, means:

- the determination of system performance requirements with respect to whole-system whole-life cost-effectiveness;
- the integration of the design and planning activities so that all aspects of the system's future operations (e.g. performance, reliability, support, information, human factors, etc.) are conceived in a balanced and coherent manner;
- the close monitoring of all stages during acquisition to ensure that the operating system will achieve the specified performance within the desired cost-effectiveness boundary.

A framework for systems engineering design must therefore be comprehensive, otherwise it cannot support the systems engineer's proud boast that he at least sees the whole system problem, laterally over all the specialist dimensions, and longitudinally over the future operational life.

The paper has concentrated on providing two interlocking frameworks:

- a framework for the systems design process itself;
- a framework for the evaluation of systems designs.

These frameworks can be seen from two points of view:

- A conceptual structure from which to study and understand the overall systems design problem—as represented by Figs. 1, 2, 3, 6, Table 6.
- A professional methodology that has compartments for all the necessary specialist design and analytical skills and integrates them to provide the necessary systemic dimensions to analysis, design, development and assessment—as represented by Figs. 4, 9, 10, 11.

These frameworks allow the engineering student to realize the full extent of the 'engineering dimension' and to make the necessary connexions between science, applications, practice, economics, management and creative design so that he may graduate as a 'good engineer'. The frameworks may also help the professional engineer already in the field to enlarge his vision and realize a systems engineering dimension in his work. Of course, many engineers and project managers already see things large and see them whole—much of the practice in aviation, chemical, communications, manufacture, transport and power engineering can be seen as systems engineering. But there is a new kind of engineering—call it 'information engineering'—whose practitioners are electronics, computer, control and instrumentation engineers, and who, perforce, find themselves rapidly becoming systems engineers because their activities cause networks of linkages between the various processes and activities in all manner of technological fields. Through these linkages systems—rather than conglomerates—are created. So it is as well that the engineers who cause systems should also be systems engineers.

And at the end of most, if not all, engineering activity there is a market or customer or threat whose technological needs induce engineering developments, and whose economic pressures impose strong constraints on those developments. The whole-system whole-life dimensions in the design frameworks provide a ready bridge for the engineer to cross over from his specialist concerns to study the market's economics and opportunities. Even though the engineer may cause new technological opportunities and markets, he is also always the servant of the customer and the market. No customer for the product—no engineering industry. The focus of the frameworks on matching a system's design to the customer's real needs produces strongly market oriented engineers. And we have all been told over and over again that the UK urgently needs more of that kind of engineer.

REFERENCES

1 M'Pherson, P. K., 'Systems engineering: an approach to whole-system design', *The Radio and Electronic Engineer*, **50**, no. 11/12, pp. 545–58, November/December 1980.
2 Singh, M. G., 'Dynamical Hierarchical Control', pp. 1–9 (North-Holland, Amsterdam, 1977).
3 Hall, A. D., 'A Methodology for Systems Engineering (Van Nostrand, Princeton, 1962).
4 English, J. M., (Ed.), 'Cost-Effectiveness' (Wiley, New York, 1968).
5 Blanchard, B. S., 'Logistics Engineering and Management' (Prentice-Hall, Englewood Cliffs, New Jersey, 1971).
6 Coutinho, J. de S., 'Advanced Systems Development Management' (Wiley, New York, 1977).
7 Sage, A. P., 'Methodology for Large Scale Systems' (McGraw-Hill, New York, 1977).
8 Jelen, F. C. (Ed.), 'Cost and Optimization Engineering' (McGraw-Hill. New York, 1970).
9 Hall, A. D., 'A three-dimensional morphology of systems engineering', *IEEE Trans on Systems Science and Cybernetics*, **SSC-5**, no. 2, pp. 156–60, April 1969.
10 Jenkins, G. M., 'The systems approach', *J. Systems Engng*, **1**, no. 1, pp. 3–49, 1969.
11 Finkelstein, L. and M'Pherson, P. K., 'Information machines and systems engineering', 2nd Feinwerktechnik Meeting, Stuttgart, 1969.
12 M'Pherson, P. K. and Charlwood, F. J., 'A systems engineering approach to the design of multivariable systems', 2nd IFAC Symposium on Multivariable Control Systems, Dusseldorf, 1971.
13 Archer, L. B., 'Technical Innovation: a Methodology' (Science Policy Foundation, London, 1971).
14 Siddall, J. N., 'Analytical Decision Making in Engineering Design' (Prentice-Hall, Englewood Cliffs, N.J., 1972).
15 M'Pherson, P. K., 'System theoretic concepts on systems engineering' in Pichler, F. and Trappl, R. (Eds.), 'Advance in Cybernetics and Systems Research', pp. 72–94 (Transcription Books, London, 1972).
16 Mantell, M. I., 'Scientific method—a triad', in Rapp, F. (Ed.), 'Contributions to a Philosophy of Technology', pp. 115–23 (Reidel, Dordrecht, Holland, 1974).
17 McCrory, R. J., 'The design method—a scientific approach to valid design', in Rapp, F., *op. cit.* **11**, pp. 158–73.
18 Moore, P. G. and Thomas, H. 'The Anatomy of Decisions (Penguin, London, 1976).
19 Ostrofsky, B., 'Design, Planning and Development Methodology' (Prentice-Hall, Englewood Cliffs, N.J., 1977).
20 Kauffman, G. M. and Thomas, H. 'Modern Decision Analysis' (Penguin, London, 1977).
21 BS 4891: 1972, A Guide to Quality Assurance (British Standards Institution, London, 1972).
22 Doggett, R. H., 'Practical applications in airline maintenance', National Conference on Terotechnology in Higher Education, Department of Industry, London, 1979.
23 Cochrane, J. L. and Zeleney, M. (Eds.), 'Multiple Criterion Decision Making' (University of South Carolina Press, Columbia SC, 1973).
24 Starr, M. K. and Zeleny, M. (Eds.), 'Multiple Criteria Decision Making' (North Holland, Amsterdam, 1977).
25 M'Pherson, P. K., 'The Design of Multidimensional Value Criteria', DSS/PKM/187, Department of Systems Science, The City University, London, May, 1979.
26 Miller, R. T., 'A ship design process', *Marine Technology*, October 1975.
27 MacFarlane, A. G. J., 'Dynamical System Models' (Harrap, London, 1970).
28 'WSEIAC (Weapon System Effectiveness Industry Advisory Committee). Final Report, USAF System Command, Andrews Air Force Base, Maryland, 1965.
29 Enthoven, A. C. and Wayne-Smith, K., 'How Much is Enough?' (Harper & Row, New York, 1971).
30 Fisher, G. H., 'Cost Considerations in Systems Analysis' (American Elsevier, New York, 1971).
31 Rau, J. G., 'Optimization and Probability in Systems Engineering' (Van Nostrand-Reinhold, New York, 1970).
32 Howard, R. A., 'Dynamic Probabilistic Systems' (Wiley, New York, 1971).
33 Gupta, S. C., 'Transform and State Variable Methods in Linear Systems' (Wiley, New York, 1966).
34 Kazanowski, A. D., 'A standardized approach to cost-effectiveness evaluation', in English, J. M., 'Cost-Effectiveness', p. 144 (Wiley, New York, 1968).
35 Goldman, A. S. and Slattery, T. B., 'Maintainability: a Major Element of System Effectiveness' (Wiley, New York, 1969).
36 Starr, M. K., 'Systems Management of Operations' (Prentice-Hall, Englewood Cliffs, N.J., 1971).
37 Grouchko, D. (Ed.), 'Operations Research and Reliability' (Gordon and Breach, New York, 1971).
38 Harary, F., 'Graph Theory' (Addison-Wesley, Reading, Mass, 1969).
39 Warfield, J. N., 'Intent structures', *IEEE Trans. on System and Cybernetics*, **SMC-3**, no. 2, pp. 133–40, March 1973.
40 Warfield, J. N., 'Societal Systems: Planning, Policy and Complexity' (Wiley, New York, 1977).
41 Keeney, R. L. and Raiffa, H., 'Decisions with Multiple Objectives' (Wiley, New York, 1976).
42 Bell, D. E., Keeney, R. L. and Raiffa, H. (Eds.), 'Conflicting Objectives in Decisions' (Wiley, New York, 1977).
43 Von Neumann, J. and Morgenstern, O., 'Theory of Games and Economic Behaviour' (Princeton University Press, Princeton, N J, 1953).
44 Luce, R. D. and Raiffa, H., 'Games and Decisions' (Wiley, New York, 1957).
45 Raiffa, H., 'Decision Analysis' (Addison-Wesley, Reading, Mass, 1968).
46 Cetron, M. and Bartocka, B. (Eds.), 'Technology Assessment in a Dynamic Environment' (Gordon and Breach, New York, 1973).
47 Garg, D. P., 'Comprehension of societal impacts via technology assessment', *Trans. Am. Soc. Mech. Engrs: J. Dynamic Systems, Measurement & Control*, **98**, ser. G, no. 1, pp. 57–62, March 1976.
48 Leopold, L. B., *et al.*, 'A procedure for evaluating environmental impact', *Geology Survey*, no. 645, pp. 1–13, 1971.
49 Bishop, A. B., 'Approach to evaluating environmental, social, economic factors', *Water Resources Bull.*, **8**, no. 4, 1972.
50 Buchring, W. A. and Foell, W. K., 'Environmental Impacts of Electrical Generation: A System-wide Approach', RR–76–13, International Institute for Applied Systems Analysis, Laxenburg, Austria, April 1976.
51 Baecher, G. B., Gross, J. G. and McCusker, K., 'Balancing Apples and Oranges: Methodologies for Facility Siting Decisions', RR–75–33, International Institute for Applied Systems Analysis, Laxenburg, Austria, September 1975.
52 Keeney, R. L., 'Energy Policy and Value Trade-Offs', RM–75–76, International Institute for Applied Systems Analysis, Laxenburg, Austria, December 1975.

APPENDICES

Appendix 1: State Transition Models

The feasible state transitions for the system of Fig. 7 may be represented by Fig. 17.[28,31,32]

The probability that the system is in state z_j at t is

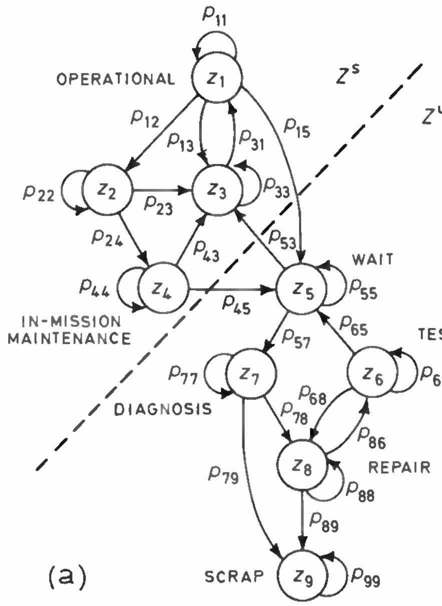
$$P_j(t) = \sum_j p_{ij} P_i(t_0); \qquad \sum_j P_j(t) = 1, \tag{37}$$

e.g.

$$P_1(t) = p_{11}P_1(t_0) + p_{31}P_3(t_0) + p_{51}P_5(t_0)$$
$$P_2(t) = p_{12}P_1(t_0) + p_{22}P_2(t_0).$$

Hence the operational state probability difference equation is

$$\mathbf{P}(t) = [P]\mathbf{P}(t_0) \tag{38}$$



	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9
z_1	p_{11}	O	p_{31}	O	O				
z_2	p_{12}	p_{22}	O	O	O				
z_3	p_{13}	p_{23}	p_{33}	p_{43}	p_{53}		O		
z_4	O	p_{24}	O	p_{44}	O				
z_5	p_{15}	O	O	p_{45}	p_{55}	p_{65}	O	O	O
z_6					O	p_{66}	O	p_{68}	O
z_7		O			p_{57}	O	p_{77}	p_{78}	p_{79}
z_8					O	p_{68}	O	p_{88}	O
z_9					O	O	O	p_{89}	p_{99}

$= [P]$

(b)

Fig. 17. Feasible state transitions for system of Fig. 7.
(a) State transition diagram. (b) State transition matrix.

where **P** is the column vector

$$[P_1, P_2, \ldots, P_j, \ldots, P_n]^T. \tag{39}$$

The state transition probabilities p_{ij} are functions of the associated reliabilities R and maintainabilities M. Consider the three-state system of Fig. 18. P_{aa} is the reliability of state $z_a = 1 - p_{ab}$.

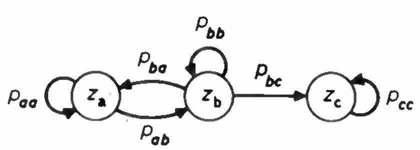


Fig. 18. Three-state system.

Assuming exponential failure and repair distributions

$$p_{aa} = R_a = 1 - e^{\lambda a \tau} \tag{40}$$

where λ_a is the failure rate of state a, and $\tau = t - t_0 \rightarrow$ small. Hence

$$p_{aa} \simeq 1 - \lambda_a \tau,$$
$$p_{ab} \simeq \lambda_a \tau,$$
$$p_{ba} \simeq \mu_b \tau, \quad \text{etc.}$$

The state probability equation for one transition interval τ is given by

$$\mathbf{P}(\tau) = [P]\mathbf{P}(0) \tag{41}$$

where

$$[P] = \begin{bmatrix} 1-\lambda_a\tau & \mu_b\tau & 0 \\ \lambda_a\tau & 1-(\lambda_b+\mu_b)\tau & 0 \\ 0 & \lambda_b\tau & 1 \end{bmatrix}. \tag{42}$$

Hence

$$\frac{\mathbf{P}(\tau)-\mathbf{P}(0)}{\tau} = \begin{bmatrix} -\lambda_a & \mu_b & 0 \\ \lambda_a & -(\lambda_b+\mu_b) & 0 \\ 0 & \lambda_b & 0 \end{bmatrix} \mathbf{P}_0 \tag{43}$$

which may be written as the differential equation

$$\dot{\mathbf{P}} = [Q]\mathbf{P}_0 \tag{44}$$

where $[Q]$ is the matrix on the r.h.s. of (43).

Equation (44) is a linear differential equation which may be solved using standard techniques:[33]

$$\mathbf{P}(t) = [D]\mathbf{P}(0) \tag{45}$$

where $[D]$ is the solution matrix for the homogeneous equation. It is also the *dependability matrix* with elements d_{ij} which are the solution probabilities of being in state $z_j(t)$ given the initial state probabilities $\mathbf{P}(0)$.

Note that the state transition model is structured according to the design configuration of the operational, reliability and maintenance system $\$_O$, $\$_R$, $\$_M$. It is quantified by obtaining the values of the elements in the failure and repair rate vectors λ and μ which are derived from the actual design and policy parameters being considered.

The elements c_j of the capability vector have to be assessed by considering the likely performance of the operational system in each z_i state.[4] For example, reverting to the original state transition diagram:

$C =$	= (e.g.)	
c_1	1	fully operational
c_2	0·9	operational but degraded
c_3	0	on standby
c_4	0·7	operational under maintenance
c_5	0	non-operational states
c_6	0	
c_7	0	
c_8	0	
c_9	0	

$c_1 = 1$ represents performance in the 'as good as new' state.

Point Availability, that is the probability that the system is in an operational state at some time t, is the sum of the probabilities of the satisfactory states.

$$A(t) = \sum P_k(t) \qquad (47)$$

where

$$P_k(t) = p(z(t) = z_k \in Z_S).$$

In the steady state, long term or inherent availability reduces to:

$$A(\infty) = \frac{\text{MTTF}}{\text{MTTF} + \text{MDT}} \qquad (48)$$

where MTTF is the system mean time to failure, which is a function of reliability design and policy on in-service repair. MDT is the system mean down time, which is a function of the overall maintenance and logistic system designs.

As a trivial demonstration of an effectiveness calculation, assume a simple two state system (z_1 = up, z_2 = down) with and without repair.

Let

$$\mathbf{A} = [0{\cdot}7 \;\; 0{\cdot}3], \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 0{\cdot}9 \\ 0{\cdot}1 \end{bmatrix}.$$

Suppose the solution gives the following dependability matrices, then from equation (19):

$$E \text{ (no repair)} = [0{\cdot}7 \;\; 0{\cdot}3] \begin{bmatrix} 0{\cdot}8 & 0{\cdot}2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0{\cdot}9 \\ 0{\cdot}1 \end{bmatrix} = 0{\cdot}55,$$

$$E \text{ (with repair)} = [0{\cdot}7 \;\; 0{\cdot}3] \begin{bmatrix} 0{\cdot}8 & 0{\cdot}2 \\ 0{\cdot}4 & 0{\cdot}6 \end{bmatrix} \begin{bmatrix} 0{\cdot}9 \\ 0{\cdot}1 \end{bmatrix} = 0{\cdot}64.$$

The system with repair has a higher effectiveness and is therefore technically better as it will have a higher probability of meeting its objectives during operations. But this is not the final answer as the life cycle costs with and without repair must be examined to see if the increase of effectiveness is worth the added cost.

Appendix 2: Structuring and Weighting a Simple Objectives Tree

For illustration, the following ten (not inclusive) design objectives are suggested for a portable instrument with visual read-out:

A	Good *Stability*	F	*Robustness*
B	Fast *Response Time*	G	*Reliability*
C	Low *Dither*	H	*Maintainability*
D	Low *Error*	J	*Readability*
E	Low *Drift*	K	*Portability*

Subordination Matrix

Labelling a subordination matrix for the objectives will yield the structure of the objectives tree and the levels of importance of the objective elements. The entry (0, 1) indicates that the assessing analyst or decision-maker (does not, does) think that the row element is subordinate to the column element.

	A	B	C	D	E	F	G	H	J	K	Row sum	
A	0	1	0	0	0	0	0	0	0	0	1	
B	0	0	0	0	0	0	0	0	0	0	0	Top element
C	0	0	0	1	0	0	0	0	1	0	2	
D	0	0	0	0	1	0	0	0	0	0	1	
E	0	0	0	1	0	0	0	0	0	0	1	
F	0	0	0	0	0	0	1	0	0	0	1	
G	0	0	0	1	1	0	0	0	0	0	2	
H	0	0	0	1	1	0	0	0	0	0	2	
J	0	0	0	1	0	0	0	0	0	0	1	
K	0	0	0	0	0	0	0	0	0	0	0	Top element
	×		×			×		×		×	× = Bottom element (Column sum = 0)	

Graph of Subordination Matrix

(by inspection for this simple case)

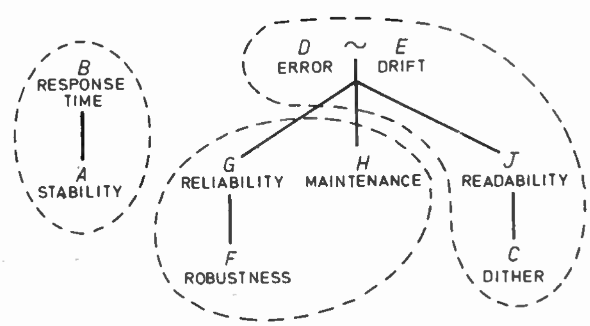


Fig. 19. Graph of subordination matrix.

Clusters of like elements in Fig. 19 have been labelled for Dynamics, Availability, Accuracy. To obtain an Effectiveness format: group Dynamics, Accuracy and Portability under Capability.

Objectives Tree with Weights by Direct Scoring

Each element is now scored to obtain weights to signify its importance relative to other elements in its branch at its level, and to deduce the relative importance of design efforts.

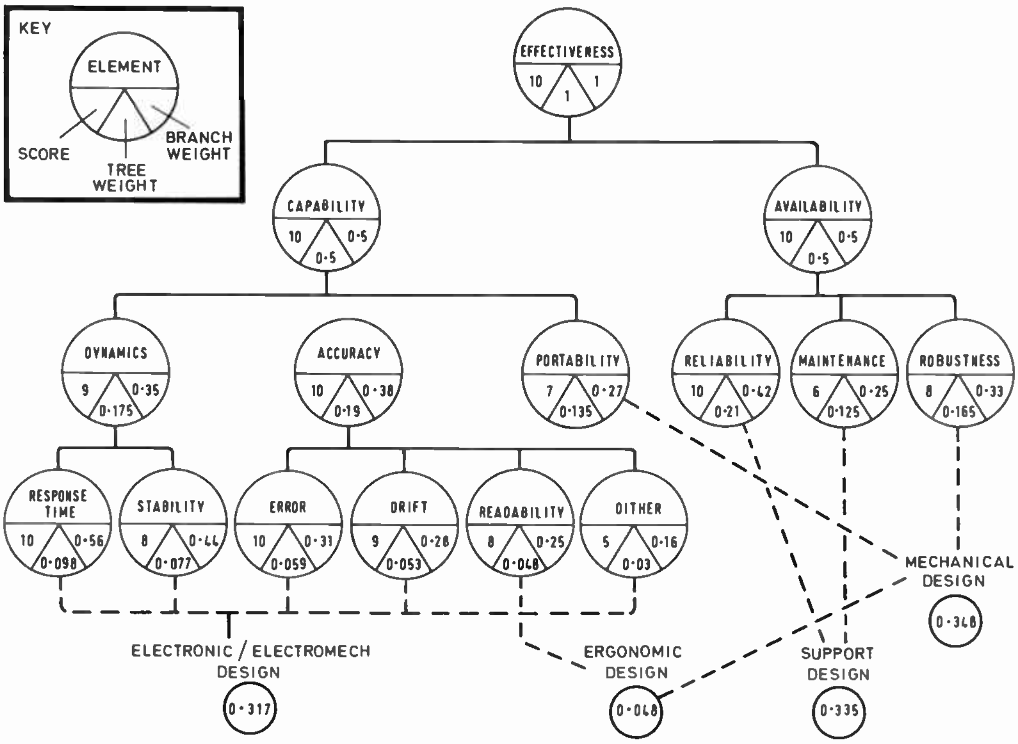


Fig. 20. Objectives tree with branch and tree weights for the elements.

Score out of 10 (for very high).

- Branch weight for an element is its score divided by sum of scores of the branch at that level.
- Tree weight for an element is its branch weight multiplied by the tree weight of the successive higher branches from which it hangs.

The final version of the objectives tree is given in Fig. 20, the insert showing the significance of the entries in the boxes.

Appendix 3: Evaluation and the Loading Matrix

Consider the general loading matrix L^i_{NJK} in which the system A_i ($\$_i$, $\hat{\alpha}_i$) is represented by N performance functions Y^i_n, there are J possible states of nature N_j to be taken into account, and evaluation has to be conducted with respect to K attributes g_k. The superscript i will be omitted, except where it is essential.

States of Nature and Expected Value

The range of each performance function will fall between a minimum level y^0_n and y^+_n, perhaps as prescribed by the system specification:

$$y^0_n \leqslant y_n \leqslant y^+_n . \qquad (49)$$

It is to be expected that some at least of the performances will be sensitive to the prevailing state of nature. Assume that an appropriate scenario analysis has provided occurrence probabilities for each state of nature:

$$p(N(t_f) = N_j) = p_j; \qquad \sum_{j=1}^{J} p_j = 1. \qquad (50)$$

Then the expected value of each performance function is given by:

$$E(Y_n) = \sum_j p_j(y_n | N_j) \qquad (51)$$

where $(y_n|N_j)$ is the level of Y_n in the circumstances of N_j.

Risk and Utility Functions

The expected values of the performance function take the uncertainty of future performance into account: they introduce the element of risk into evaluation. The expected values of the design achievements will be called design *outcomes* (in line with Decision Analysis). The DM now has to assess the loading of each design outcome $E(Y_n)$ onto each attribute g_k. The loading has two aspects:

- the value contribution to the attribute taking the element of risk into account;
- the importance of the outcome to the realization of the attribute.

The assessment of value contributions under risk is a subjective matter depending on the DM's attitude towards risk in the context of the particular desirable attribute g_k. Utility functions are employed here to represent the DM's perception of the changing desirability of an outcome with respect to the attribute depending on the level of outcome in the range (y^0_n, y^+_n). Utility theory has a strong axiomatic basis and provides a methodology for 'measuring' the DM's risk attitudes.[43–45]

The determination of a utility function depends on the DM's response to a sequence of lotteries in which he is offered the chance of 'winning' y_n^+ with a probability of π or y_n^0 with a probability of $(1-\pi)$, and then finding what *certain* value of y_n he would accept in lieu of the lottery. (See Fig. 21.)

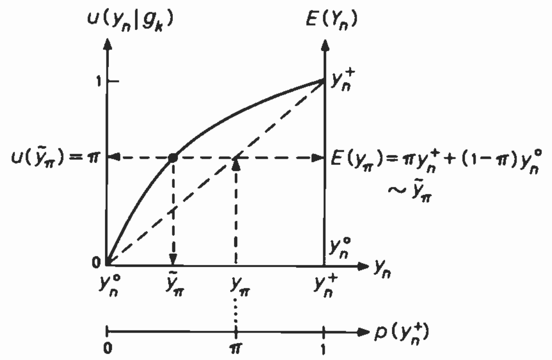


Fig. 21. Response of the decision maker to a sequence of lotteries.

The lottery has the expected value:

$$E(y_n^+, \pi, y_n^0) = \pi y_n^+ + (1-\pi)y_n^0 = y_\pi. \qquad (52)$$

In Fig. 21, the DM has responded by saying that he would accept the lottery (in the context of g_k) for $y_n \leqslant \tilde{y}_\pi$ and accept the certainty $\tilde{y}_\pi < y_\pi$ for $y_n \geqslant \tilde{y}_\pi$. Thus to this DM the 'utility' of $\tilde{y}_\pi$ is the same as the expected value of the lottery, which is somewhat greater than the expected value of $\tilde{y}_\pi$. By repeating the process a utility function $u(y_n|g_k)$ can be drawn over the range (y_n^0, y_n^+). A typical risk-averse function is drawn.

By convention

$$u(y_n^0) = 0, \qquad u(y_n^+) = 1. \qquad (53)$$

Negative utilities can be avoided by rescaling so that complete avoidance of the undesirable outcome has unit utility. Thus a utility function:

- maps the range of *any* risky outcome onto the real line between (0, 1);
- represents the DM's personal attitude towards the risky element y_n in the context of a particular attribute g_k;
- provides a preference function such that:

$$y_1 \succ y_2 \quad \text{iff } u(y_1) > u(y_2). \qquad (54)$$

A utility function is not a statement of general validity: it is personal to the DM whose preferences were mapped in a risky context.

Loading Functions

Even if there is no risk, the DM is likely to have subjective preferences with respect to an outcome's contribution to an attribute as it varies over its range. Take for an example the case of a design achievement of Mean Time To Failure (MTTF) loading the attribute Reliability (R). There is a simple relationship between MTTF and R:

$$R = \exp\left(-\frac{T}{\text{MTTF}}\right);$$

where T is a reference time. The relationship between R and MTTF is plotted in Fig. 22. Suppose that the design specification stipulates a minimum acceptable reliability, and that the designer reckons that the current state of art defines a maximum MTTF beyond which it would be unreasonable to strive. The range to explore is therefore between MTTF_{min} and MTTF_{max}. Clearly MTTF_{min} is 'just good enough', and MTTF_{max} is 'as good as I can get', which may be translated into a loading scale between the values (0, 1). Using a similar lottery procedure as that for determining a utility function, (but changing the context to satisfaction with possible design achievements), the DM can be 'measured' for his preferences over (y^0, y^+) with respect to the satisfaction of the reliability attribute.

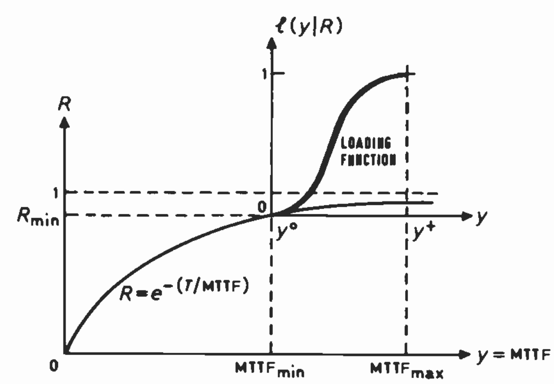


Fig. 22. Relationship between reliability and mean time to failure.

Quality

The term loading is introduced to denote a kind of riskless utility. It is to be expected that the DM might produce different shapes for $u(y|R)$ and $l(y|R)$ depending on the uncertainty he felt in achieving the higher levels of y. The end result in either case is a scalar function on the real line between (0, 1) whose form represents the DM's preferences with respect to a particular outcome—attribute combination. The term *quality* is now introduced to act as a label for the perceived value contributions resulting from the evaluation process *whatever* the underlying circumstances (risky, certain, objective or subjective). Quality is defined as a scalar on the real line between (0, 1) such that:

$$(Y_n^i \succ Y_n^{i+1}|g_k) \quad \text{iff } q(y_n^i|g_k) > q(y_n^{i+1}|g_k) \qquad (55)$$

with

$$\left.\begin{array}{c} 0 \leqslant q(\cdot) \leqslant 1 \\ \text{and} \\ q(y_n^0|g_k) = 0, \quad q(y_n^+|g_k) = 1; \quad \forall n, \forall k \end{array}\right\} \qquad (56)$$

Multi-attribute Quality

The Loading Matrix has $N \times K$ cells (Fig. 5). The DM is now confronted by the problem of combining up to N qualities for each attribute column. If a column has more than one non-zero cell two questions arise:

- how should the qualities be scaled to represent the relative importance of the contributions?;
- how should the qualities be combined to provide an aggregate quality contribution for that column?

Introduce scaling factors s_{nk} to denote the relative importance of the design outcome y_n with respect to attribute g_k. The s_{nk} will be non-negative real numbers. The quality of a particular design A_i with respect to the attribute g_k can now be written as

$$q(A_i|g_k) = \langle (s_{1k}q_{1k}) \circ \ldots \circ (s_{nk}q_{nk}) \circ \ldots \circ (s_{Nk}q_{Nk}) \rangle \quad (57)$$

where $\langle (\cdot) \circ (\cdot) \circ \ldots \circ (\cdot) \rangle$ is some appropriate combinatorial rule between the terms in the brackets that has to be determined.

Columns with only one non-zero cell imply a one-to-one correspondence between design outcome and attribute. A diagonal loading matrix results if the 1 : 1 relationship is carried between the outcome and attribute sets. This should be regarded as a special case, but designers who have only ever operated within the technical level often assume that the set of design targets they are aiming for is the same as the set of desirable systemic attributes. Consequently there is some confusion in separating out the design outcomes (achievements) which represent 'what the system will be', and the system attributes which represent 'what the system *ought* to be'. Nor should it be assumed that the scaling factor is unity for the single non-zero cell in a column. A unity scaling factor implies both maximum importance, and direct relevance. For example, MTTF is not a complete indicator of Reliability: other factors enter such as quality of operational procedures and maintenance.

The problem of determining scaling factors and the appropriate combinatorial rule between multiple quality functions does not disappear with a diagonal Loading Matrix. A similar problem will always occur when constructing the value criterion (upper part of Fig. 11) where the input qualities from the loading matrix have to be combined into a single overall worth.

Scaling Factors

Subjective scaling factors and 'spot' loading factors can of course be obtained from direct assessment (i.e. scoring) by either the DM or an expert group. A more reliable if more tedious method is to trade-off between the design outcomes loading an attribute to find the DM's points of indifference relative to the levels of associated outcomes.[41] Consider the set of outcomes

$$Y = \{y_1, y_2, \ldots, y_n, \ldots, y_N\}$$

loading one attribute g_k. Introduce the notation

$$Y = \{y_1, y_2; y_{\overline{12}}\} \quad (58)$$

where $y_{\overline{12}}$ is the complement of $\{y_1, y_2\}$ in Y.

Assume for convenience that an initial direct assessment of the DM's preferences has suggested that the outcomes have the preference order

$$y_1 \succ y_2 \succ \ldots \succ y_n \succ \ldots \succ y_N$$

with respect to g_k.

We have

$$Y^+ \succ Y \succ Y^0 \quad (59)$$

as

$$q(y_1^+, y_2^+, \ldots, y_n^+, \ldots, y_N^+) = 1$$

and (60)

$$q(y_1^0, y_2^0, \ldots, y_n^0, \ldots, y_N^0) = 0.$$

Trade-off between (y_1, y_2), assuming that $y_{\overline{12}}$ is fixed at $y_{\overline{12}}^0$, by asking the DM to find the level of y_1 at which he is indifferent between $(y_1^0, y_2^+; y_{\overline{12}}^0)$ and $(y_1, y_2^0; y_{\overline{12}}^0)$. These two points define the ends of an isopreference curve in consequence space (Fig. 23).

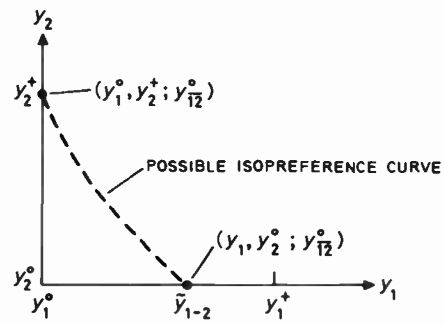


Fig. 23. Trade-off of outcomes.

We have

$$(q(y_1^0, y_2^+; y_{\overline{12}}^0) = s_2) \sim (q(\tilde{y}_{1-2}, y_2^0; y_{\overline{12}}^0) = s_1 q(\tilde{y}_{1-2})), \quad (61)$$

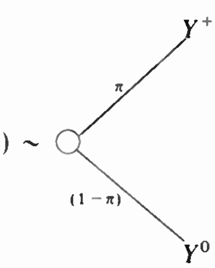
i.e.

$$s_2 = s_1 q(\tilde{y}_{1-2})$$

By a similar process

$$s_3 = s_2 q(\tilde{y}_{2-3})$$
$$\vdots$$
$$s_n = s_{n-1} q(\tilde{y}_{n-(n-1)})$$
$$\vdots \quad (62)$$

Equations (62) provide $n-1$ simultaneous equations so all the scaling factors $s_2 \ldots s_n \ldots s_N$ can be solved for in terms of s_1. To find s_1 offer the DM the following lottery:



(63)

The value of probability π at which he shifts from the certainty equivalent to accepting the lottery gives s_1, because

$$q(y_1^+, y_2^0 ; y_{12}^0) = s_1$$

and (64)

$$E(Y^+, \pi, Y^0) = \pi.$$

If $\pi \gg 0{\cdot}5$ the DM prefers y_1 much more than any single other outcome, i.e. y_1 is dominant.

Note that the scaling factors found by this method do not necessarily sum to unity.

Estimating the Capability Elements

A special problem of evaluation in the Loading Matrix is the definition of the system's capability. While the scalar values of availability and dependability can result directly from analysis of the state-transition model representing probable operational behaviour (Fig. 9, Appendix 1), the elements c_r in the capability vector require more complex analysis. The capability vector is given by equation (18):

$$C = [c_1, c_2, \ldots, c_r, \ldots, c_R]^T. \quad (18)$$

The construction of a capability element is shown in Fig. 12—which has obvious similarities with the valuation structure of Fig. 11. The capability c_r is a measure of the *overall* performance of the system relative to its 'as good as new' design performance c_1 when it is in an operational state z_r. Thus c_r is the conflation of the weighted utilities of each design achievement y_n with respect to the attainment of the system's overall performance. The individual contributions are given by

$$(s_{nc} \cdot u_n | z_r) = s_{nc} u(\zeta_{nr} \cdot y_n) \quad (65)$$

where s_{nc} is the scaling factor for the nth performance outcome y_n with respect to overall system capability;
$u_n = u(y_n)$ is the utility of y_n relative to capability;
ζ_{nr} is the degradation factor reducing the outcome y_n to a degraded value corresponding to its ability to perform in state z_r.

The outcome y_n will be the expected value $\tilde{y}_{nj}$ (eqn. (5)) as appropriate. Then the capability element c_r is given by the conflation:

$$c_r = \langle (s_{1c} u_1 | z_r) \circ \ldots \circ (s_{nc} u_n | z_r) \circ \ldots \rangle \quad (66)$$

and an appropriate combinatorial rule has to be found. It is often assumed that an additive rule can be assumed for combining multiple utilities (e.g. Ref. 4):

$$c_r = \sum_n s_{nc} \cdot u_n | z_r ; \qquad \sum_n s_{nc} = 1. \quad (67)$$

The simple additive rule is appropriate for small perturbations of performance level about a reference, and it is also valid for performance functions that are strictly independent of each other over the whole of their range. But the property of strict independence cannot be assumed (e.g. for a vehicle design are the utilities of range and speed independent?). Thus the problem of finding the combinatorial rule appropriate to a multi-attribute utility function has also to be solved during the construction of the capability elements.

Appendix 4: Illustration of Cost-effectiveness Analysis

Consider a simple two-state system in which:

state z_1: operating at design capability;
state z_2: operating at reduced capability due to internal malfunction.

Six designs are to be evaluated within the same operational and environmental contexts. For each design high and low cost concepts for availability, dependability and capability are under consideration. The six designs incorporate the high/low cost concepts as follows:

CONCEPT		DESIGN					
		α	β	γ	δ	ε	ζ
Availability:	low	×	×	×	×		
	high					×	×
Reliability:	low	×	×	×		×	×
	high				×		
Maintainability:	low	×	×	×	×		
	high					×	×
Capability:	low	×	×	×			
	high				×	×	×

Thus α, β, γ are low cost designs, δ is intermediate and ε, ζ are mostly high cost designs.

The effectiveness E of the system is given by equation (19):

$$E = A[D]C = [a_1, a_2] \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}. \quad (68)$$

In particular d_{11} is the system reliability (probability of not failing during mission), and d_{21} is the maintainability (probability of a successful repair during the mission).

The costs K of the system are given by:

$$K = f(K(A), K(D), K(C), K(P)) \quad (69)$$

where $K(A) = k(a_1)$, the cost of the support system that provides an availability at start of mission of probability a_1;
$K(D) = k(d_{11}) + k(d_{21})$, the cost of designing in a reliability of d_{11}, and a maintainability of d_{21};
$K(C) = k(c_1, c_2)$, the cost of designing a system with relative performance c_1 in state z_1, and c_2 in state z_2;
$K(P) = f(p_2)$, the penalty cost payable depending on the proportion of mission time spent in state

z_2, where p_2 is given by

$$[a_1, a_2]\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} = [p_1 \quad p_2]. \qquad (70$$

The relative costs of $K(A)$, $K(D)$, $K(C)$ are shown in Fig. 13(a, d). Designs $\alpha, \beta, \gamma, \delta$ do not provide in-mission maintenance. The high cost availability for designs ε, ζ is coupled to the need to provide efficient and close-up logistic support to the in-service repair capability.

Assume:

$$\left.\begin{aligned} K(D) &= k(d_{11})+k(d_{21}) \\ K(C) &= \{k(c_1)+k(c_2)\} \\ K(P) &= (p_2)^2\{K(A)+K(D)+K(C)\} \end{aligned}\right\} \qquad (71)$$

i.e. penalties increase as the square of the downtime, with the asset value of the system being wasted completely if it is not operational for any period of the mission

$$\max k(d_{11}) = k(c_1),$$

i.e. max reliability is provided by two operational systems in parallel

$$k(d_{21}) = 0 \quad \text{for } \alpha, \beta, \gamma, \delta$$

$$\max k(d_{21}) = \tfrac{1}{2} \max k(d_{11}) \quad \text{for } \varepsilon, \zeta.$$

Design ζ is taken as the reference design for costs.

Let

$$K_s = K(A)+K(D)+K(C) \qquad (72)$$

for any design. Then the cost relative to ζ's cost is given by

$$K_s = k(a_1)+\{1{\cdot}82k(d_{11})+0{\cdot}5k(d_{21})+1\} \times k(c_1)+k(c_2) \qquad (73)$$

and

$$K = (1+p_2^2)K_s. \qquad (74)$$

Table 9 lists the data for the Effectiveness/Cost calculation: the attribute values and costs for each design are as labelled in Fig. 13(a, d). Figure 13(e) then indicates the calculated location of each design in (K–E) space, together with a representative function $K = f(E)$ for this family of designs. Figure 13(f) plots the E/K ratio for each design against cost K. Using the E/K criterion as an index of merit the low cost design concepts α, β, γ form a cluster of significantly 'better' designs than the high cost designs δ, ε).

Appendix 5: Notation

Abbreviations

CE	Cost-effectiveness
DA	Design Assurance
DM	Decision-maker/Designer
DT	Downtime
ISD	Integrated System Design
LD	Logistic Design
MD	Maintenance Design
MDT	Mean Downtime
MTTR	Mean Time To Repair
NPV	Net Present Value
PD	Performance Design
QA	Quality Assurance
RD	Reliability Design
SA	Systems Assurance
SD	Support Design
SDM	Subsystem DM

Symbols

$A = \{A_i\}$	Alternatives
A	Point availability
A_x	Inherent availability

Table 9

DESIGN	PROBABILITY OF STATE Z_2 p_2	EFFECTIVENESS E $[a_1 \; a_2]\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22}\end{bmatrix}\begin{bmatrix} c_1 \\ c_2\end{bmatrix} = E$	ATTRIBUTE COSTS $K(A)$ $k(a_1)$	$K(D)$ $k(d_{11})$	$k(d_{12})$	$K(C)$ $k(c_1)$	$k(c_2)$	COST RELATIVE TO ζ K_s	K_p	K	EFFECTIVENESS COSTS E/K	Order
α	0·12	$[0{\cdot}84 \; 0{\cdot}16]\begin{bmatrix} 0{\cdot}90 & 0{\cdot}10 \\ 0 & 1\end{bmatrix}\begin{bmatrix} 0{\cdot}95 \\ 0{\cdot}50\end{bmatrix} = 0{\cdot}84$	0·55	0·55	0	0·9	0·5	2·3	0·03	2·33	0·361	1=
β	0·20	$[0{\cdot}87 \; 0{\cdot}13]\begin{bmatrix} 0{\cdot}92 & 0{\cdot}10 \\ 0 & 1\end{bmatrix}\begin{bmatrix} 0{\cdot}95 \\ 0{\cdot}50\end{bmatrix} = 0{\cdot}86$	0·59	0·57	0	0·9	0·5	2·34	0·09	2·43	0·354	3
γ	0·15	$[0{\cdot}90 \; 0{\cdot}10]\begin{bmatrix} 0{\cdot}95 & 0{\cdot}05 \\ 0 & 1\end{bmatrix}\begin{bmatrix} 0{\cdot}95 \\ 0{\cdot}50\end{bmatrix} = 0{\cdot}89$	0·66	0·62	0	0·9	0·5	2·42	0·05	2·47	0·360	1=
δ	0·08	$[0{\cdot}93 \; 0{\cdot}07]\begin{bmatrix} 0{\cdot}99 & 0{\cdot}01 \\ 0 & 1\end{bmatrix}\begin{bmatrix} 0{\cdot}96 \\ 0{\cdot}70\end{bmatrix} = 0{\cdot}94$	0·74	1	0	1	0·6	3·42	0·02	3·44	0·273	6
ε	0·12	$[0{\cdot}91 \; 0{\cdot}09]\begin{bmatrix} 0{\cdot}95 & 0{\cdot}05 \\ 0{\cdot}20 & 0{\cdot}80\end{bmatrix}\begin{bmatrix} 0{\cdot}97 \\ 0{\cdot}70\end{bmatrix} = 0{\cdot}93$	0·82	0·62	0·5	1	0·6	2·98	0·04	3·02	0·308	4
η	0·11	$[0{\cdot}95 \; 0{\cdot}05]\begin{bmatrix} 0{\cdot}90 & 0{\cdot}01 \\ 0{\cdot}70 & 0{\cdot}30\end{bmatrix}\begin{bmatrix} 0{\cdot}96 \\ 0{\cdot}70\end{bmatrix} = 0{\cdot}93$	1	0·55	1	1	0·6	3·10	0·04	3·14	0·296	5

$\mathbf{A} = [a_r]$	Availability vector
a_r	Probability system is in state z_r
$\mathbf{B}$	Vector of resource inputs
$\mathbf{C} = [c_r]^T$	Capability vector
C_{AQ}	Acquisition cost
C_{Eth}	Ethical costs
C_{DT}	Downtime penalty
C_F	Failure or repair cost
C_I	Inventory cost
C_K	Capital costs
C_{KC}	Capital charges
C_M	Maintenance cost
C_O	Operational cost
C_{OS}	Operating system costs
C_R	Reliability cost
C_{RL}	Realization cost
C_S	System whole-life costs
C_{SS}	Support system cost
C_I	Inspection costs
$[D]$	Dependability matrix
d_{ij}	Element in $[D]$
E	Effectiveness
$E(\ )$	Expected value of ( )
$f(\ )$	Function of ( )
$G = \{G_k\}$	Objectives
$\hat{G}$	Objectives hierarchy
$\vec{g}$	Set of design goals
g_k	kth goal in $\vec{g}$
I	Spares inventory
K	System whole-life cost
$K(\ )$	Cost of ( )
$K(P)$	Penalty cost
L^i_{NJK}	Loading matrix
l^i_{njk}	Loading factor
$\underset{\sim}{M}$	Vector of control inputs
$N = \{N_j\}$	Environmental states
$\hat{N}$	Vector of environmental states
n_F	Number of failures
n_I	Number of inspections.
$\mathbf{P} = [P_j]$	State probability vector
p_j	Probability system is in state z_j
p_{ij}	State transition probability from z_i to z_j
p_j	Probability environment N_j will occur
$p(\ )$	Probability of ( )
Q	Quality space
$\hat{q}$	Quality vector
q_{nk}	Quality of nth design outcome relative to kth design goal
$q(\ )$	Quality of ( )
R	Reliability
$\$_i$	Design configuration for A_i
$\$_M$	Maintenance structure
$\$_O$	Operational structure
$\$_R$	Reliability structure
s_n	Scaling factor for y_n
s_{nk}	Scaling factor
T	Mission time, total time
T	Inspection interval
t_f	Future time
t_0	Present time
$u(\ )$	Utility of ( )
$V(\ )$	Value of, value criterion
$W(\ )$	Worth of ( )
w_C	Commercial worth
w_N	Environmental worth
w_S	Social worth
w_T	Technical worth
$\hat{X}$	Vector of functional state variables
$Y_i = \{Y^i_n\}$	Performance functions
$\bar{\bar{Y}}^i = \{\bar{y}^i_n\}$	Expected performance levels
y^i_n	Level of Y^i_n
y^0_n, y^+_n	Min, max levels of y^i_n
$\bar{y}^i_j$	Expected level of y^i_n given N
$\tilde{y}$	Certainty value, indifference value of y in lottery
$y_{\overline{12}}$	Complement of y_1, y_2 in set $Y = \{y_1, y_2, \ldots, y_n, \ldots, y_N\}$
$Z = \{z_r\}$	Operational states
Z_S	Set of satisfactory states
Z_U	Set of unsatisfactory states
$\alpha, \beta, \gamma, \delta, \varepsilon, \zeta$	Labels for alternative designs
$\hat{\alpha}_i$	Alterables in design A_i
$\delta(\ )$	Small change in ( )
λ	Failure rate
μ	Repair rate
π	Probability of stock-out
π	Probability in lottery
σ	Demand rate for spares
τ	Increment of time
ζ_{nr}	Degradation of y_n given z_r
iff	'if and only if'
[ ]	Matrix or vector
$(\underset{\sim}{\ }), (\hat{\ })$	Vector
$(F\|H)$	Event F given event H
$\succ$	Preference
$>$	Greater than
$\forall(\ )$	Universal quantifier
$\langle \circ \ldots \circ \rangle$	Conflation of terms in $\langle\ \rangle$.

Manuscript first received by the Institution on 28th May 1980 and in final form on 29th July 1980.
(Paper No. 1971/ACS 82)

UDC 681.325.57

Indexing Terms: Computers, Two's complement multiplier, Cellular array

A two's complement cellular array multiplier

K. Z. PEKMESTZI, B.E.E.*

and

Professor G. D. PAPADOPOULOS, B.E.E., M.S.E.E., Ph.D., M.I.E.E.E.†

SUMMARY

A cellular array for the multiplication of signed binary numbers is presented. The implementation is based on the direct addition of the partial products. The addition of the negative partial products is performed with a new method which yields a fully cellular array. Furthermore, the proposed array multiplier has an iterative cell interconnection pattern and it is well suited for large-scale integration.

* *Electronics Department, Nuclear Research Center Demokritos, Aghia Paraskevi Athens, Greece.*

† *School of Engineering, Applied Electronics Laboratory, University of Patras, Patras, Greece.*

The Radio and Electronic Engineer, Vol. 51, No. 2, pp. 94–96, February 1981

1 Introduction

Iterative multiplication arrays that perform the multiplication by the conventional shift-and-add algorithm have been proposed by Hoffmann *et al.*,[1] by Dean[2] and De Mori.[3] The De Mori multiplier is faster but does not yield an iterative cell interconnection pattern. To overcome this disadvantage Guild[4] proposed a full iterative multiplier that is based on the direct partial product sum. These arrays, however, can be used only for the multiplication of positive numbers.

For the direct multiplication of signed numbers, represented in two's complement notation, three methods have been proposed for the manipulation of the negative terms. Booth's algorithm[5] is the first method that, in conjuction with the sign extension technique, has been used in several array multipliers.[6–8] The second method is based on a procedure that handles the negative terms by means of their complements. This method has also been used for the construction of array multipliers.[9,10] In the third method, proposed by S. Pezaris, the negative terms are added on a digit level using modified adders to handle the addition of positive and negative digits. This method employs three types of cells and an array based on the carry-save technique.[11] Recently, parallel array multipliers have been constructed in integrated circuits.[12,13] These multipliers are also based on the carry-save addition technique.

In this paper a two's complement cellular array multiplier is proposed which handles the negative terms on a digit level. The basic structure of the multiplier is the Guild array which has been modified in order to permit signed digit additions.

2 The Proposed Array Multiplier

Let us represent the numbers to be multiplied in two's complement notation both with N digits as shown below,

$$X = -x_{N-1}2^{N-1} + \sum_{i=0}^{N-2} x_i 2^i \quad (1)$$

$$Y = -y_{N-1}2^{N-1} + \sum_{j=0}^{N-2} y_j 2^j \quad (2)$$

The product is

$$X\ Y = \sum_{i=0}^{N-2} x_i 2^i \sum_{i=0}^{N-2} y_j 2^j - x_{N-1} \sum_{j=0}^{N-2} y_j 2^j - y_{N-1} \sum_{i=0}^{N-2} x_i 2^i + x_{N-1} y_{N-1} 2^{2N-2} \quad (3)$$

The above equation defines the negative partial products that need special treatment. These terms are incorporated in the Guild array shown in Fig. 1 for $N=4$, where the configuration of a typical cell is shown at the top of this Figure. The cells are distinguished by a

0033–7722/81/020094+03 $1.50/0

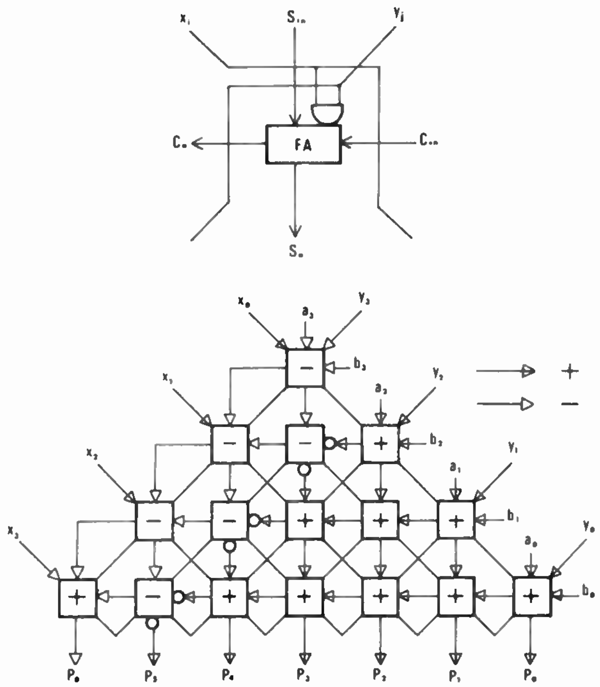


Fig. 1. The proposed two's complement cellular array multiplier. The basic cell used is shown separately at the top of the Figure.

positive or negative sign according to the sign of the partial product $x_i y_j$. It must be emphasized, however, that this distinction does not mean any change in the basic cell structure.

The presence of negative terms implied negative digits that propagate through the $(-)$ cells of the array. From the canonical form of the array it can be shown that the $(-)$ cells have their S_{in} input always negative while the C_{in} input can take either sign. Thus, there are two distinct cases of sign combinations for the inputs S_{in}, C_{in}, which are shown in Fig. 2(a) and 2(b). These sign combinations are indicated in Fig. 1 by arrows of two kinds corresponding to positive and negative signs.

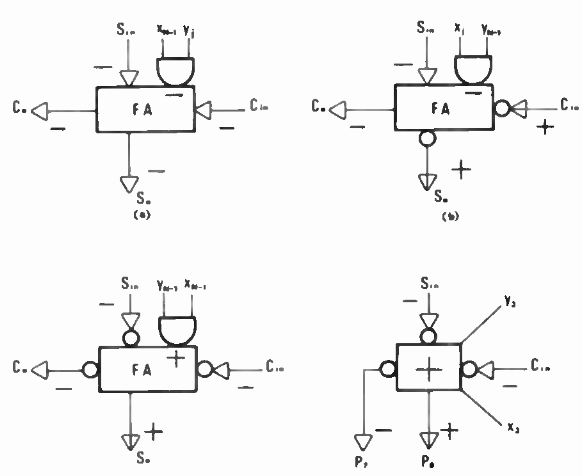


Fig. 2. Modifications of the basic cell for signed-digit additions. (a) The type of $(-)$ cells at the left boundary of the array. (b) The type of $(-)$ cells at the diagonal next to the left boundary of the array. (c), (d) The left corner cell and its specific form for $N=4$.

The $(-)$ cell of Fig. 2 (a) employed at the left boundary of the array of Fig. 1 corresponds to the terms $y_{N-1}x_i$ for $i=0$, N-2 and performs the addition,

$$(-y_{n-1}x_i - S_{in} - C_{in}) = (-S_o - 2C_o) \qquad (4)$$

This addition is implemented with the direct use of the basic cell, as can be verified by changing all the signs in eqn. (4). The $(-)$ cell of Fig. 2(b) placed at the diagonal next to the left boundary of the array of Fig. 1 corresponds to the terms $x_{N-1}y_j$ for $j=0$, $N-2$; it employs inverters at C_{in} and at S_o of the basic cell in order to perform the addition,

$$(-x_{N-1}y_j - S_{in} + C_{in}) = (S_o - 2C_o) \qquad (5)$$

The $(+)$ cells do not accept negative weighted digits, except for the corner cell that handles the term $x_{N-1}y_{N-1}$. This cell must perform the following addition,

$$(x_{N-1}y_{N-1} - S_{in} - C_{in}) = (S_o - 2C_o) \qquad (6)$$

The implementation of this addition is shown in Fig. 2(c), where the outputs S_o and C_o are the product digits P_{2N-2} and P_{2N-1}, respectively. The digit P_{2N-1}, however, is redundant information because the product can always be represented with $2N$-1 digits, except for the case that both X and Y are equal to -2^N. Consequently, the output S_o is the m.s.b. of the result and from the basic cell of Fig. 2(c) we obtain.

$$S_o = (x_{N-1}y_{N-1}) \oplus \bar{S}_{in} \oplus C_{in} \qquad (7a)$$

But this equation is equivalent to the following,

$$S_o = (x_{N-1}\, y_{N-1}) \oplus S_{in} \oplus C_{in} \qquad (7b)$$

Equation (7b) shows that P_{2N-2} can be obtained by using the basic cell without inverters, as shown in Fig. 1.

If the product is required to have $2N$-1 digits in order to include the case where both X and Y are equal to -2^N, the corner cell must be modified to have three inverters at S_{in}, C_{in} and at C_o, as shown in Fig. 2(c). For the specific case where $N=4$, the corner cell takes the form shown in Fig. 2(d).

The proposed array also performs the function $XY+A+B$. The two numbers A and B are in two's complement notation and are applied through the additive inputs at the boundary of the array, as shown in Fig. 1.

The multiplication time with the proposed array is $(2N\text{-}1)T_p$, where N is the word length of X and Y and T_p is the propagation delay of a full adder. The multiplication time can be reduced if the N left-most cells at the bottom of the array are replaced by a gated carry look-ahead adder. If $N=2^b$ the multiplication time after

Table 1

Comparison of the proposed two's complement cellular array multiplier with previously proposed realizations ($N \times N$) multiplication).

Multiplier type proposed by:	Multiplication time	Number of cells	Gates per cell	Cellularity / Type of circuit	Iterative interconnection pattern
J. C. Majithia and R. Kitai[6]	$(2N-1)T_p$	$N^2 + \frac{N(N-1)}{2}$	13	ALMOST add-subtract no operation cell	YES
S. Bandyopadhyay *et al.*[7]	$(3N-2)Tp$	N^2	13	ALMOST Add-subtract-no operation cell	YES
C. I. Toma[8]	$2NT_p$	$\frac{N(N+1)}{2}$	16	ALMOST Add-subtract-shift-no operation cell	YES
I. D. Deegan[9]	$2NT_p$	$N(N+1)$	10	ALMOST Add-no operation cell	NO
S. D. Pezaris[11]	$(2N-1)T_p$	$\frac{(N-1)N}{2}$	20	YES Three types of cells	NO
Proposed multiplier based on the Guild array	$(2N-1)T_p$ or $(N+b)T_p$	N^2	10	YES Add-no operation cell	YES

this modification becomes $(N+b)T_p$. For example, a 16-bit by 16-bit multiplication requires $20T_p$. Using emitter coupled logic this multiplication can be performed in 40 ns.

The main characteristics of the proposed two's complement cellular array multiplier are compared in Table 1 with those of previously proposed realizations. As can be seen from this Table only the proposed multiplier has fully cellular structure with iterative interconnection pattern. Furthermore, it offers fast multiplication with rather simple circuitry.

3 Conclusions

A two's complement cellular array multiplier has been described. The proposed multiplier incorporates the fast Guild array with a small modification that permits two's complement number multiplication. The main advantages of the proposed array are its cellular structure and its iterative interconnection pattern. These features suggest that the proposed multiplier is more suitable for integrated circuit realization than the previously proposed multipliers mentioned in Table 1.

4 References

1 Hoffman, J. C., Lacaze, B. and Csillag, P., 'Multiplieur parallele à circuits logique iteratifs', *Electronics Letters*, **4**, p. 178, May 1968.

2 Dean, K. J., 'Design of a full multiplier', *Proc. IEEE*, **115**, pp. 1592–4, November 1968.

3 De Mori, R. 'Suggestion for an i.c. fast parallel multiplier', *Electronics Letters*, **5**, pp. 50–1, February 1969.

4 Guild, H. H., 'Fully iterative fast array for binary multiplication and fast addition', *Electronics Letters*, **5**, p. 263, May 1969.

5 Booth, A. D., 'A signed binary multiplication technique', *Quart. J. Mech. Appl. Math.*, **4**, pt. 2, pp. 236–40, 1951.

6 Majithia, J. C. and Kitai, R., 'An iterative array for multiplication of signed binary numbers', *IEEE Trans, on Computers*, **C-20**, pp. 214–6, February 1971.

7 Bandyopadhyay, S., Basu, S. and Choudhoury, A. K., 'An iterative array for multiplication of signed binary numbers', *IEEE Trans on Computers*, **C-21**, pp. 921–2, August 1972.

8 Toma, C. I., 'Cellular logic array for high-speed signed binary number multiplication', *IEEE Trans. on Computers*, **C-24**, pp. 932–5, September 1975.

9 Deegan, I. D., 'Cellular multiplier for signed binary numbers', *Electronics Letters*, **7**, pp. 436–7, July 1971.

10 Gibson, J. A. and Gibbard, R. W., 'Synthesis and comparison of two's complement parallel multipliers', *IEEE Trans. on Computers*, **C-24**, pp. 1020–7, October 1975.

11 Pezaris, S. D., 'A 40ns 17 bit by 17 bit array multiplier', *IEEE Trans. on Computers*, **C-20**, pp. 442–7, April 1971.

12 Molver, G. W., Miller, R. W. and O'Shaughnessy, O., 'A monolithic 16 × 16 digital multipliers', Digest of Technical Papers, 1974 IEEE Intl Solid-State Circuit Conference.

13 TRW, USA Patent No. 3,283,170.

Manuscript first received by the Institution on 22nd January 1980 and in final form on 24th July 1980.
(Paper No. 1972/Comp 198)

UDC 621.395.625.3: 621.396.712.2

Indexing Terms: Tape recorders, Sound recording and reproduction, Radio broadcasting

Synchronization of 6·3 mm tape recorders using EBU time and control code

W. R. HAWKINS, C.Eng., M.I.E.E.*

C. HENOCQ, C.Eng., M.I.E.R.E.*

and

R. J. TAYLOR, B.Sc.(Eng.), C.Eng., M.I.E.E.*

Based on a paper presented at the IERE Conference on Video and Data Recording held in Southampton in July 1979

SUMMARY

The paper discusses the addition of EBU time and control code to the twin-track format on 6·3 mm tape recorders. The code is carried on a centre track 0·8 mm wide using a recording characteristic of zero microseconds so that the system can read the code accurately over a wide range of tape speeds. To preserve the waveshape, the recording amplifier has phase-corrected equalization; and to minimize noise, the equalization of the reproducing chain is split into two parts using head loading for high-frequency compensation followed by a low-frequency circuit in the amplifier.

* *Designs Department, British Broadcasting Corporation, Broadcasting House, London W1A 1AA.*

1 Introduction

The difficulties encountered during television productions make post-production sound mixing and editing essential. In this operation various sound sources available on magnetic tape are mixed to produce the soundtrack of the programme which has already been edited as far as the picture component is concerned. These operations have become known in the BBC as SYPHER because they use a multi-track audio recorder synchronized to a U-matic format helical scan v.t.r. (*SY*nchronized *P*ost-production dubbing using *H*elical scan and *E*ight-track *R*ecorders).[1] The 80-bit EBU time and control code[2] (timecode) was recorded on audio tape recorders in the BBC for the first time especially for this application so that they could run in synchronism with the v.t.r.

It is fairly convenient to allocate one track of a 51 mm multi-track recorder to timecode, but when only one or two tracks are required it is more convenient to use 6·3 mm tape (especially for sound effects). Before this can happen it is essential to find space for a dedicated timecode track.

Very similar equipment is required for the simultaneous broadcast of stereo sound on radio channels and programmes on the television network. It would be particularly advantageous to use 6·3 mm tape here so that the same equipment could be used for both normal and simultaneous broadcasts (Simulcast). In a practical situation the television recorder (v.t.r.) could be housed in the television area replaying its programme to the network, and the radio recorders could be housed in the radio area with their own independent controls. The only link would be a feed of EBU timecode from the v.t.r to the audio tape machines. Dedicated equipment at the receiving end would ensure that synchronism was maintained using this signal and the timecode reproduced from the audio recorder. If necessary the link could be broken and the sound and television broadcasts continued independently.

This paper describes a suitable standard for the recording and disposition of the timecode track together with the necessary electronics and editing controls. If generally accepted, such a standard would promote the design of suitable equipment and permit programme interchange.

2 Current Requirements For Timecode Working

The timecode waveform can be recorded and replayed on almost any audio recorder so that the off-tape signal can be decoded at normal play speed without any additional waveform processing. Some distortion is inevitably introduced but usually the timing distortion or interference is not excessive and the waveform can be decoded. However, this only applies at the normal play speed and a small range of speeds either side. For some applications this may be acceptable but for reasons given

0033-7722/81/020097+06 $1.50/0

below it is often necessary to adapt the machine to replay a useable timecode at spooling speeds.

The timecode signal can be used to control a number of functions such as searching and parking at preselected cue points, synchronizing a slave machine with a master and the generation of commands to other equipments. Searching and parking require the timecode to be read from the tape over a wide range of tape speeds, from 50 times normal play speed to $\frac{1}{5}$th play speed, so making possible the automatic parking of the tape to within 0·1 second approximately (2 television pictures). This wide range of tape speed requires the machine to be fitted with a wideband replay amplifier which has a 6 dB/octave characteristic over 10 octaves (100 Hz-100 kHz) or so. A more complex equalization is not effective because the replayed signal occupies different parts of the band as the speed is varied.

More accurate parking than this is possible but usually not necessary unless accurate starting of the machine from a sound cue is required. This could require positioning the tape to within $\frac{1}{300}$ s of tape time and may only be feasible by manually rocking the tape to find the precise position.

In addition, the tape must engage either continuously or intermittently with the replay heads while searching for the selected park point. Fortunately most professional machines have the facility for bringing the tape into contact with the heads while spooling or can be easily modified to do so.

Almost all twin-channel recorders in the BBC Radio Service are stereo-format machines whilst those in the Television Service are twin-track. Both Services use multi-track audio machines and where timecode is required the highest numbered track is used for it. The twin-track machines use one of the audio channels for audio and the other for timecode recording.

The narrow guard-band between standard stereo audio tracks is not wide enough to accommodate a satisfactory third channel even though it be used for timecode. The problem lies mainly with track crosstalk but there is also the difficulty of inserting effective screens into the normal audio head stacks. Both problems are helped by a wide guard-band between tracks. Fortunately, the twin-track format provides sufficient space for a 0·8 mm track with 0·65 mm guard bands either side (Fig. 1).

If such a standard is acceptable both existing and future machines would have to be fitted with a complete set of timecode heads, amplifiers and function controls as well as audio heads made to the twin-track format.

The advantage of this proposal is that the machines retain much of the simplicity and ease of operation of the basic 6·3 mm tape machine. It has the disadvantage that stereo tapes recorded with timecode on these modified twin-track machines cannot be replayed on conventional recorders. The converse is possible, but even so care will

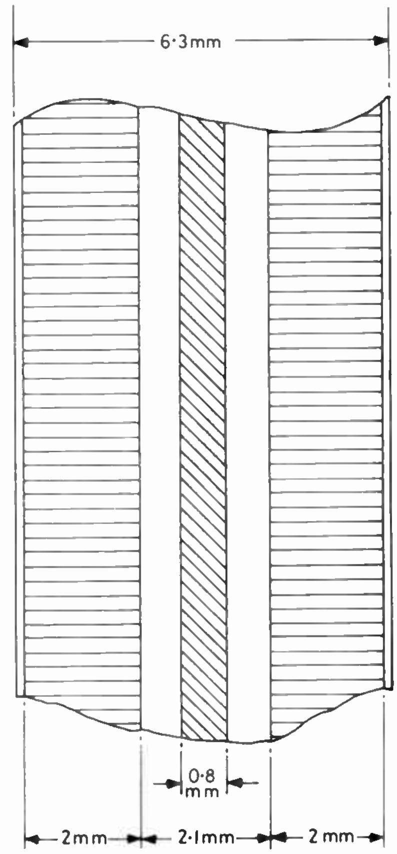


Fig. 1. Nominal dimensions of audio and timecode tracks.

have to be taken to keep tapes recorded on the two formats quite separate from each other.

3 New Operational Facilities

A machine fitted with timecode should have independent control of record for the timecode and audio tracks so that timecode can be recorded before, together or after the audio tracks have been recorded. The simplest form of control on modified twin-track machines is a Timecode Record switch which enables the timecode chain to erase and record when the machine is in the Play or Record modes. With the switch in the Off position, the audio tracks alone may be played or recorded, because usually the timecode signal is left untouched once it has been recorded.

Insert electronic editing is a requirement on multi-track machines used in the Television Sound Dubbing suites and such machines should be able to insert new audio material into existing programmes without noticeable clicks, gaps or overlap. However, the signal which puts the machine into Insert Record has to be in advance of the start of the Insert to allow for the spacing between the Erase and Record heads. The timing of this operation is derived from the timecode signal using control equipment with pre-programmed offsets.

An offset can also arise between the programme material and the timecode because the same head is not being used to record and replay the programme material. This will occur when mixing the outputs from a number of replay heads and recording the composite signal on another track. In this case, an offset between the material on the newly recorded track and the original timecode has been introduced, equal to the spacing between the record and replay heads. For example, if the spacing between the record and replay heads is 76 mm and the tape speed is 19 cm/s, then the offset is 0·4 s (10 television pictures).

If synchronization of this sound source is subsequently required the offset will have to be allowed for in the replay synchronizing equipment.

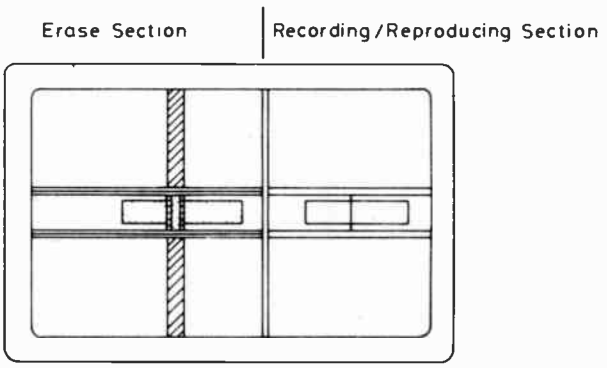


Fig. 2. Combined timecode erase and recording/reproducing head.

4 The Timecode and Audio Tracks

A successful system must meet the following conditions:

(a) As mentioned before, it must be possible to record the timecode together with the audio, or before or after the audio has been recorded.
This condition is best satisfied by the addition of a completely separate timecode channel, with erase, record and reproducing facilities. It is also necessary to ensure that the whole width of the tape is erased by suitable arranging the widths and dispositions of the individual erase heads.

(b) The cross-talk produced by the addition of the timecode channel must not degrade the noise performance of either of the two audio channels.
A poor performance could be caused by crosstalk from the timecode track. It can be kept to a minimum by using bias to record the timecode signal, thereby reducing the recorded level by some 20 dB compared to that produced by a saturated recording.

(c) The signals recorded on the audio tracks must not crosstalk sufficiently into the timecode channel to cause mis-reading of the information contained within the timecode.
Track crosstalk at low frequencies is produced by the magnetic flux spreading into the reproducing head from adjacent tracks, whilst at high frequencies the crosstalk is caused by parasitic coupling in the multi-track heads.

The use of a separate head for recording and reproducing timecode will virtually eliminate the high frequency crosstalk but unless the head is very carefully constructed it is impossible to eliminate the low frequency crosstalk from the audio tracks into timecode head.

The big drawback to using a separate head for timecode is that only a combined record/reproduce head can be accommodated in most head blocks and this means timecode cannot be monitored whilst recording it. However, it is possible to build an erase section within the combined assembly and so the amount of mechanical work required on the head block is reduced to adding a single component. Furthermore, the presence of the erase head close to the recording point reduces the complexity of the timecode offset arrangements, should it be required to erase and re-record small sections of the timecode track. Figure 2 shows the construction of the combined erase record/reproducing head used in the prototype equipment.

5 The Recording Characteristic

The performance of any sound or video tape recorder is made consistent by the careful setting-up of input and output voltages using test tapes and measuring equipment. This degree of standardization has not yet been achieved with the EBU timecode because not all of the signal parameters have been specified.

In formulating the first timecode proposals it was considered that a conventional cue audio channel was satisfactory for the timecode waveform. However, group delay is a parameter which is unspecified in an audio channel but of great importance in a digital signal such as timecode. Excessive group delay distortion will cause poor waveshape and therefore poor resolution; correcting this distortion will provide a good waveshape over a wide range of replay speeds.

In the overall chain, two kinds of losses can be identified, circuit losses and wavelength losses. Circuit losses usually approximate to minimum phase shift characteristics and so conventional equalization provides phase equalization as well as amplitude equalization. Wavelength losses on the other hand usually have no phase shift so phase-corrected equalization must be used to compensate for them.

Typical examples are:

	Recording	*Reproducing*	*Comment*
Circuit effects	Head losses	Head losses, Flux rate of change	Phase equalization fairly easy
Wavelength effects	Record process, Coating thickness	Gap losses, Recording characteristic correction	Phase equalization more difficult

The terms are self-explanatory with perhaps the exception of:

Flux rate of change losses: These losses arise because a flux is recorded, but magnetic heads respond to a rate-of-change of flux and hence the head output rises 6 dB/octave. Tape reproducing amplifiers have a constant 6 dB/octave loss as one of their correction characteristics.

Recording characteristic correction: This conforms to a simple RC network top lift with an ultimate steady slope of 6 dB octave. It corrects the losses due to the thickness of the oxide coating and is usually in the range 30 to 100 µs (+6·5 dB to +16 dB at 10 kHz relative to 1 kHz).

If the recording characteristic is made zero microseconds then this amount of correction is transferred from the reproducing amplifier to the recording amplifier and two attendant advantages occur. (Fortunately gap losses are relatively small and their equalization can be ignored.)

(a) The replay chain can now be equalized over a wide bandwidth (essential for variable speed replay) because it need be only a constant 6 dB/octave.

(b) The replay chain can be conventionally calibrated using a voltage injection method at the reproducing head. Usually a low-value resistor is connected in series with the head. Once a chain is calibrated in this way a test tape can be made.

In a practical case the replay equalization can be shared between the reproducing head and the reproducing amplifier. (Fig. 3.) The head can be loaded by a preset resistor to produce a high-frequency roll-off with an accurate turnover frequency; the amplifier can best provide the low-frequency correction because of the very high gain provided by integrated-circuit operational amplifiers. In this way, a 6 dB/octave characteristic from 10 Hz to 250 kHz can be achieved without amplifier instability.

A zero microsecond recording characteristic will require considerable pre-emphasis in the recording amplifier, in some cases up to 18 dB, which suggests tape overload and waveform crushing may occur. However, the effect is not as serious as might be feared at first sight because two factors ease the situation. First, the edges of the EBU timecode waveform have a sine-squared shape with a rise-time of 50 µs, and secondly, the amplitude of the component spectrum falls with frequency. The two effects are shown in Table 1, where the components of 1 kHz and 2 kHz square wave are tabulated. The reference level is 0 dB peak-to-peak.

The timecode waveform uses bi-phase modulation to carry a 2 kbit/s data channel. It is convenient to consider the cases where all the bits are 0 or all are 1, and the waveform becomes a 1 kHz or 2 kHz square wave, even though the waveform specification precludes this happening.

Table 1 Amplitude of components

Square wave – 1 kHz	f_1	f_3	f_5	f_7	f_9	
Frequency	1	3	5	7	9	kHz
Amplitude of components	+2	−8	−12	−15	−17	dB
Rise-time effect	0	−½	−1½	−3	−5	dB
Resultant	+2	−8½	−13½	−18	−22	dB

Square wave – 2 kHz	f_1	f_3	f_5	f_7	
Frequency	2	6	10	14	kHz
Amplitude of components	+2	−8	−12	−15	dB
Rise-time effect	0	−2	−6	−12	dB
Resultant	+2	−10	−18	−27	dB

And so, by carefully choosing the recorded level, a very good waveshape can be reproduced with only small impairment due to noise. We have found that by

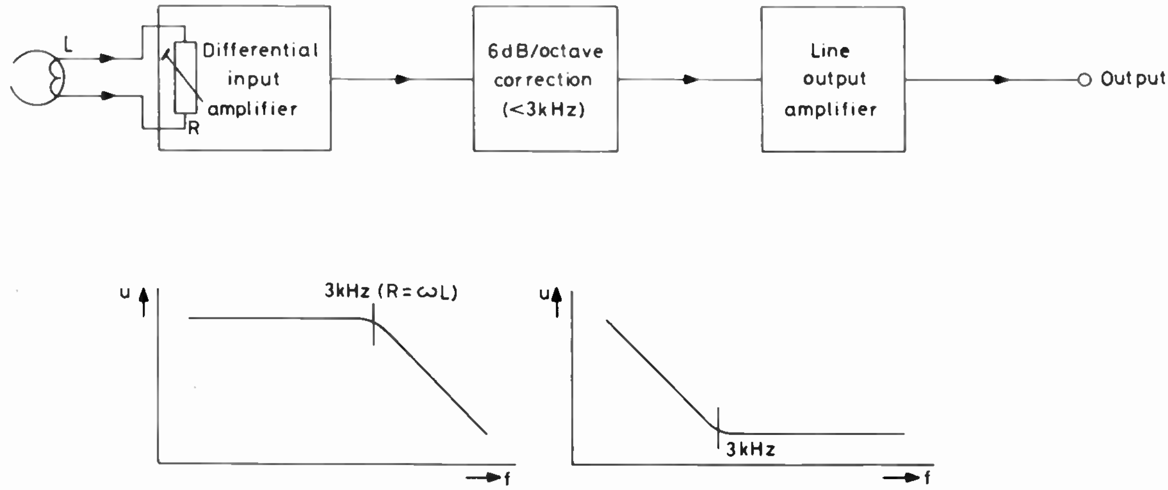


Fig. 3. Timecode reproducing chain.

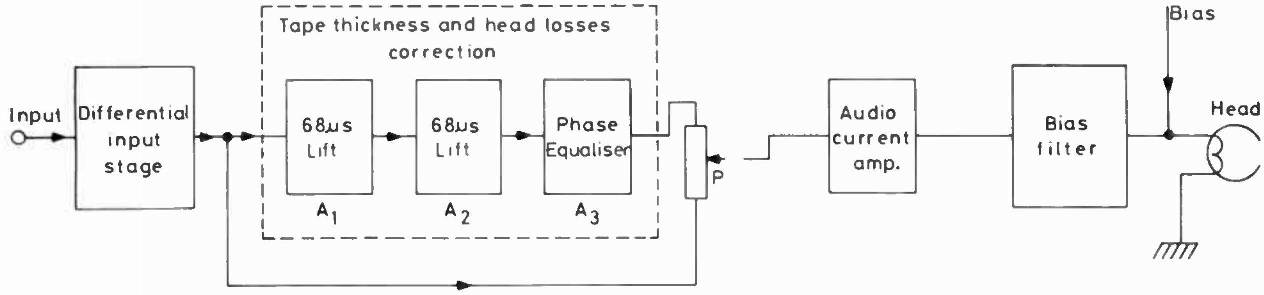


Fig. 4. Timecode recording chain (zero microsecond recording characteristic)

recording only 6 dB below our maximum audio level at 38 cm/s and 19 cm/s, the waveshape remains adequate at hand inching (÷20) speeds and good up to fast spooling (×60 speeds). Moreover, because of the replay chain characteristic, the amplitude remains sensibly constant over the whole range.

The record phase-corrected equalization is conveniently achieved by two operational amplifiers each providing a cell of high-frequency lift (68 µs = 12 dB at 10 kHz) followed by a single all-pass amplifier capacitor-resistor circuit. (Fig. 4) The overall gain $A_4 = A_1A_2A_3$

Let

$$A_1 = A_2 = A_0(1+j\omega T_1) \qquad (1)$$

and

$$A_3 = A_0\frac{1-j\omega T_1}{1+j\omega T_1}. \qquad (2)$$

Therefore

$$A_4 = A_0^3(1+\omega^2T_1^2). \qquad (3)$$

Because of the phaseless nature of (3) it is easy to adjust its characteristic using the potentiometer P.

Let its adjustment fraction be n where $(0 < n < 1)$

The output $= 1\ (1-n)+n(1+\omega^2 T_1{}^2) = 1+n\omega^2 T_1{}^2$

The phase-corrected high-frequency lift can thus be continuously adjusted from zero to maximum in a convenient manner.

6 The Synchronizing Equipment

The essential features of a synchronizer are the decoding and comparison of the two timecode signals from master and slave machines, and using this error signal to control the capstan motor speed. Once the slave is synchronous with the master, the system must be able to 'flywheel' through timecode drop-outs from the slave and master signals and use the flywheel to remain synchronous without following the wow and flutter of the master.

Such a simple system would require the slave machine to be brought into synchronism manually, say to within one or two seconds of tape time difference, so that its capstan control could then achieve synchronism in less than 5 seconds.

However, the synchronizer can be extended to control automatically the slave machine's spooling motors so that it will achieve and maintain synchronism whatever the transport mode of the master. Even though the phase error between the master and slave may be a few seconds of tape time when the slave is following the master at fast spooling speeds, yet when the master stops the slave will park within 2 or 3 pictures of the master parking point. Accurate synchronization is generally only necessary when the machines are in the play mode and the phase difference is then within ±1 ms of tape time.

In BBC equipments, offsets are stored in the synchronizer and selected as appropriate to allow for timecode errors arising from head spacing on the machines.[3] Further, so that common designs of synchronizer can be used in a number of applications, the interface between the synchronizer and machine generally takes the form of a special unit designed to suit the machine. The functions controlled by the synchronizer are Play Mode, Stop, Fast Forwards, Rewind and Tape Lift Defeat, which allows the tape to be put back into contact with the heads when the synchronizer requires timecode in the spooling mode (Search Mode).

Control of the spooling speed is usually arranged by switching between Fast Forwards and Rewind with a variable mark/space ratio. The technique allows quite accurate parking and control, but it is essential that the machine has solid state switching control of these functions.

Simple synchronizers such as this, each controlling one machine which follows a master machine in all modes, have been in operational use at the BBC Television Centre for five years, but more complex systems are coming into service using a synchronizer for each machine with microprocessor-controlled data entry.

7 Conclusions

The work described in this paper has demonstrated how a successful timecode centre track can be incorporated into the normal twin-track audio format on 6·3 mm tape to provide a useful facility for post-production dubbing in television and in simultaneous broadcasts (Simulcasts). Even though the proposed format is incompatible with the conventional stereo format it should be possible to use machines with this facility for normal replay to radio networks.

In this way the radio broadcaster will be given much of the editorial and operational freedom that he seeks on Simulcasts and the exploitation of numerical control on audio tape machines could lead to more productive forms of audio tape editing.

However, the work has revealed how essential it is to have versatile operational controls so that the recording and reproducing operations can be carried out in any order on the audio and timecode channels, and to have a full width-of-tape erase facility. Moreover, the synchronizer which controls the machines should have an adequate number of remote controls with special twin-machine facilities for Simulcasts on radio networks to facilitate automatic changeovers.

Operational trials of the equipment based on these ideas have yet to be undertaken and may of course reveal a need for some additional features.

8 Acknowledgements

The authors acknowledge the work of their colleagues in pioneering the SYPHER and Simulcast operations in the BBC and they wish to thank the Director of Engineering, British Broadcasting Corporation, for permission to publish this paper.

9 References

1 Eden-Eadon, J., 'Sypher—a modern method of dubbing on videotape' *Brit. Kinematograph Sound Television Soc. J.*, 59, no. 10, pp. 306–12, October 1977.

2 'EBU Standards for Television Recordings', Tech. 3097-E, Second edition, June 1980. European Broadcasting Union.

3 'Tape Machine Synchronizer' RD4L/506, BBC Designs Department, EDI 10349, October 1979.

Manuscript received by the Institution in final form on 30th June 1980.
(Paper No. 1973/Comm 211)

Contributors to this issue

Brian Lee graduated in physics at Bristol University in 1967 (1st hons.) and until 1973 was with Marconi Space & Defence Systems, Stanmore, working with the design of microwave hardware and, in particular, solid-state sources. Since 1973 he has been with British Aerospace Dynamics Group at Bristol, being initially involved with the computer simulation of the e.m.c. performance of large deployments of equipments, which included the development of propagation prediction methods for ground based antennas from v.h.f. to microwaves, and is currently a group leader for systems studies.

Bill Hawkins joined the BBC in 1947 as Recording Engineer in Radio operations and maintenance and has been Head of Recording Section in the BBC Designs Department since 1975. Recently he has been concerned with videotape editing, and the development of equipment for the subtitling of programmes on CEEFAX for the deaf viewer. Because of his current commitment to both video and audio systems he is particularly concerned to ensure that equipment using the EBU timecode signal can be made common to videotape, audiotape and television film systems.

Cecil Henocq (Member 1958, Graduate 1953) recently retired from the BBC, having spent over 30 years in the Research and Designs Departments. He has worked on virtually every aspect of professional sound recording in broadcasting as well as on the VERA recorder, one of the earliest attempts at video tape recording. During the past few years he has been a member of numerous national and international committees, including an NAB group responsible for their latest specification for broadcast cartridges. He remains chairman of ISO committee TC60, and at present is engaged in consultative work in the USA.

Professor Philip M'Pherson is Head of the Department of Systems Science at the City University, London. A fuller biographical note was published in the November/December 1980 issue of the Journal with his previous paper.

Ray Taylor joined the BBC in 1959 as a graduate apprentice after graduating from Imperial College with a B.Sc. (Eng) degree in the same year. He has worked in the Recording Section of Designs Department since 1961 on a wide variety of equipment development. This has included work on the early colour videotape recorders, vertical aperture correctors and electronic character generation. For the past eight years he has been concerned with the development of Timecode equipment for videotape editing and sound post-production using the EBU Time and Control Code.

Biographical notes on **Professor George Papadopoulos** and **Mr. K. Z. Pekmestzi** were published in the November 1979 issue of the Journal.