Surround-sound psychoacoustics

Criteria for the design of matrix and discrete surround-sound systems

by Michael Gerzon

Mathematical Institute, University of Oxford

There are a number of different mechanisms by which the ears localize sounds, including several low-frequency, mid-frequency and high-frequency mechanisms, as well as information derived from the reverberation of sounds. With only a few transmission channels available, one cannot hope to satisfy them all, but most existing "discrete" and "matrix" systems do not satisfy more than one or two criteria. The approaches associated with the Nippon Columbia UMX system and the NRDC ambisonic system are the only ones so far to adequately allow for several criteria.

When stereo was introduced commercially in the 1950s, it had been subjected to experiments and theoretical studies for 25 years, by Fletcher[1] in the USA, Blumlein[2] in England, and de Boer[3] in the Netherlands. Despite a remarkable anticipation of modern "matrix" four-speaker systems by Blumlein[2] in 1931, virtually no work had been done on four-speaker surround sound before its recent commercial introduction. We are thus only beginning to understand how it works, and it is the object of this paper to describe the fruits of this new understanding. Not surprisingly, hastily introduced commercial systems have proved to be sub-optimal.

Because the mathematical description of surround-sound systems is far from elementary, this aspect is not dealt with here; references[4 to 10] contain such information. In this article the principles of surround-sound psychoacoustics are described, i.e. the relationship between the sound field presented to the listener and what he actually hears.

Lord Rayleigh discovered[11, 12] that the human hearing system appears to use different mechanisms to localize sounds at frequencies below and above 700Hz. Other evidence by Rayleigh[12, 13], Stevens & Newman[14] and Roffler & Butler[15] and others suggests that above about 5KHz, yet other localization mechanisms come into play, relying on the pinnae (the flaps on the ears) to modify sounds from different directions.

To make matters even more complicated, there is considerable disagreement both among theorists and experimenters as to the localization mechanism used within each band of freqnencies, quite contrary results being obtained in different cases[16]. It seems that the ears must nse a number of different methods of sound localization, possibly deciding on a "majority verdict" in the case when different mechanisms would, if used in isolation, give differing results.

In the presence of such contradictory information, the apparent localization of a sound also depends on the experience and expectations of the listener and on the type of attention he is paying to the sound. This can easily be demonstrated by reproducing via a stereo pair of good loudspeakers a sound positioned half-way towards the left speaker, but with the speakers connected out of phase. A suitably positioned listener can then hear the sonnd to be either between the speakers or beyond the left speaker (sometimes, both at once!).

Because most matrix four-speaker systems give highly ambiguous sound position information to the listener's ears, the results obtained will depend on the individual listener. Some listeners will learn to assign sounds to their "correct" positions with experience, and others will not. As a degree of subjectivism is a poor basis for any technology, the general principles behind various different sound localization mechanisms will be examined, with a view to extracting from these common features that can be used in designing surround-sound reproduction systems.

To design surround-sound systems we do not need to understand the full intricacies of the sound processing mechanisms in the ears and brain. As far as engineering is concerned, all we need know is what type of stimulus (i.e. sound field information) is needed to create a given subjective impression, and then we can design apparatus to produce a stimulus of the required type.

However, it is also necessary to have a description of the required stimulus that is simple enough mathematically to handle in detailed calculations. Otherwise we will only be able to design a system by guessing a circuit configuration and then "number crunching" the data in a computer to see whether it will work. As there are many millions of possible system configurations, it is extremely unlikely that such a design procedure would happen to hit upon the best possible result, or even something approximating to it. Such considerations rule out from our account such phenomena as the Haas effect, which says in essence that the earliest arrival of a sound at the ears determines its apparent direction. This is difficult to analyse mathematically, as well as being an unreliable guide to the subjective sound

Quadraphonic quandary

While this article was written before publication of B. J. Shelley's article *Quadraphonic Quandary* (*Wireless World*, July 1974 pp. 235–6), it does deal with many of the queries he raised on the aims and methods of quadraphonics. You may find it instructive to decide how far his particular criticisms are answered here. But note two points. Firstly, that two of the systems earlier proposed by the author on purely mathematical grounds (two-channel periphony and, via a tetrahedron of speakers, four-channel periphony) are here shown to be inadequate on the type of psychoacoustic grounds suggested by Shelley. And secondly that disagreements among experimenters about quadraphonic psychoacoustics are no new thing; Harwood[16] documented how little agreement there is on ordinary stereo localization. These disagreements may well be due to the conflicting directional cues at the ears inherent in all two-speaker stereo and in badly designed quadraphonic systems.

direction when sounds arrive from all round.

First, what is the aim of surround sound reproduction?

Recreating a sound field

Ideally, one would like a surround-sound system to recreate exactly over a reasonable listening area the original sound field of the concert hall, or in the case of popular or electronic music, a sound field envisaged by the record producer, with many different sounds in different directions at different distances. Unfortunately, arguments from information theory can be used to show that to recreate a sound field over a two-metre diameter listening area for frequencies up to 20KHz, one would need 400,000 channels and loudspeakers. These would occupy 8GHz of bandwidth, equivalent to the space used up by 1,000 625-line television channels!

The best that can be done with the two, three or four channels currently available is as follows. For each possible position of a sound in space, for each possible direction and for each possible distance away from the listener, assign a particular way of storing the sound on the available channels. Different sound positions correspond to the stored sound having different relative phases and amplitudes on the various channels. To reproduce the sound, first decide on a layout of loudspeakers around the listener, and then choose what combinations of the recorded information channels, with what phases and amplitudes, are to be fed to each speaker. The apparatus that converts the information channels to speaker feed signals is called a "decoder", and must be designed to ensure the best subjective approximation to the effect of the original sound field.

In commercial "discrete" practice, the process of assigning positions in the sound field to the available channels, known as "encoding", is done using four channels. Sounds not in the four corner positions are, in this procedure, assigned to just those two of the four channels representing corner directions adjacent to the desired direction. This only handles distant sounds in a horizontal direction, and it is by no means evident that this is the best way of assigning such a sound field to four channels. Similarly, it is not evident, and not in fact true, that feeding these channels directly to a square of speakers gives an optimum recreation of the original sound field.

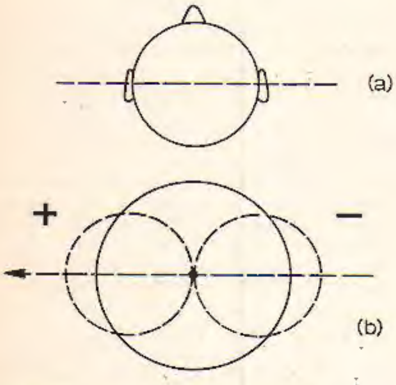


Fig. 1. Omnidirectional and velocity microphones (picture b) receiving the same low frequency information as the human hearing system (picture a).

Thus any surround-sound system gives rise to two distinct but related psychoacoustic questions:

- Is a given method of encoding the sound field *ever* capable of good subjective recreation of the sound field? That is, does the encoding method used permit the possibility of designing *some* decoder giving good results?
- Given a good method of encoding, what is the best design of decoder for use with a given layout of loudspeakers?

Low-frequency localization

The distance between the human ears is half a wavelength of a sound having a frequency of 700Hz. At frequencies appreciably below this, the head offers no obstacle to sound waves, and so the amplitude of sound reaching the two ears is virtually identical[11, 17–19]. The only information available at these low frequencies for sound localization is the phase difference between the two ears, and in 1907 Rayleigh[11] indeed showed that this was used to localize sounds below 700Hz.

There has, however, been disagreement as to how this low-frequency phase difference information is used to deduce sound position. One school of thought, represented by Clark, Dutton & Vanderlyn[20] and Bauer[21], derived a theory assuming that the listener does not move his head, whereas Makita[22], Leakey[23] and Tager[24] assume that the brain uses additional information from variations at the two ears caused by rotations of the head within the sound field.

It is possible to construct a "super-theory" including the above two classes of theories as special cases. Essentially, the sum of the waveforms reaching the two ears is the sound pressure that would be at the position of the centre of the listener's head were he absent. This information is the same as that picked up by an omnidirectional microphone (see Fig. 1). The remaining directional information at low frequencies reaching the listener is the difference of the waveforms at the two ears, which is the velocity of the sound field along the ear-axis (see Fig. 1). This is the information picked up by a sideways-pointing velocity or figure-of-eight microphone.

The fixed-head theories thus assume that the information picked up by an omnidirectional and by a sideways-facing velocity microphone is all that is available to the brain. The assumption that no use is made of amplitude differences at the two ears amounts to assuming that components of the velocity microphone information that are 90° out of phase with the omnidirectional information are not used in deducing the direction of sounds. The "moving head" theories assume that the velocity microphone information may point in any direction, but still assume that 90° out-of-phase velocity microphone information is not used.

It is not difficult to compute the "omnidirectional" and "velocity microphone" information produced by a quadraphonic reproduction system, and hence to calculate whether the useful information at low frequencies reaching the ears is the same as for live sounds (see Fig. 2).

Such calculations reveal that, for low frequencies, no existing two-channel matrix encode/decode system reproduces all the useful information as it occurs in live sounds, although the Cooper/Nippon Columbia BMX system[5] satisfies the hypotheses of Makita and Leakey. More remarkably, conventional discrete four-channel sound also does not satisfy low-frequency criteria other than those of Makita and Leakey. This is because phantom inter-speaker sound images with this system give too large an omnidirectional component of the sound field[25], which causes front-centre and side-centre sounds to be very poorly localized[26].

The poor positioning of phantom images suggests that discrete four-channel systems should not be used as a standard of excellence by which other systems are judged. There are better ways of representing the set of possible directions around the listener via four loudspeakers[8, 26]. The National Research and Development Corporation has recently been developing, with the author, a two-channel decoding apparatus for BMX or RM-encoded sounds, to feed four loudspeakers so as to satisfy the low frequency criteria shown in Fig. 2, and also the mid-high frequency criteria described later.

The three-channel system discovered

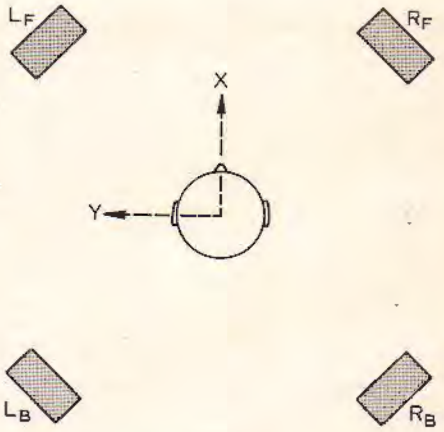


Fig. 2. Low-frequency quadraphonic localization information available to the ears.

Omnidirectional information:

$\Omega = L_B + L_F + R_F + R_B$

x-velocity information:

$X = \mathrm{Real}(-L_B + L_F + R_F - R_B)$

y-velocity information:

$Y = \mathrm{Real}(L_B + L_F - R_F - R_B)$

For "live" sounds we must have

$\Omega^2 = \frac{1}{2}(X^2 + Y^2)$.

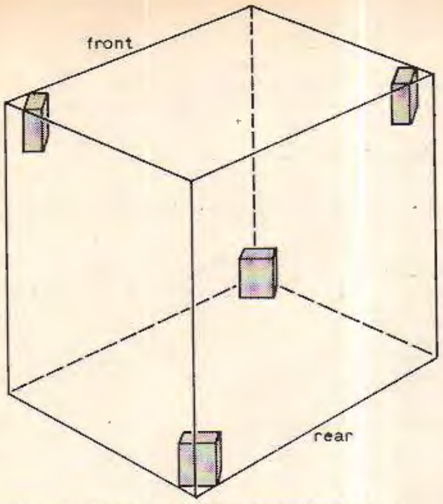


Fig. 3. Tetrahedral loudspeaker layout shown embedded in a cube.

independently by the author[10], Gibson et al[27], Eargle[28], Madsen (unpublished) and Cooper[5], is capable of correct low frequency results, as is the four-channel QMX system[5] and the tetrahedral with-height system of the author[6, 10, 29], which is reproduced via the speaker layout of Fig. 3. It is also possible to design a decoder for discrete recordings so as to satisfy all low-frequency requirements.

It is well known that velocity microphones give an exaggerated bass for very close sounds. Because the ears use velocity microphone information to localize sounds, close loudspeakers modify the directional effect at the ears. In particular, 90° out-of-phase velocity components caused by phase shifts are converted to phase differences between the ears. This causes the very low frequencies of phase-shifted sounds to be rotated around the listener. This effect has been observed by Bauer et al[30] via two speakers, but can be removed electronically. The degree of the effect is inversely proportional to loudspeaker distance.

Statistical methods may be used to apply the above theory to listeners not placed in the centre of the loudspeaker layout. The details are involved, but give results somewhat similar to the mid-high frequency theory of sound localization described next.

Mid-high frequency localization

Above 700Hz, the wavelength of sound is sufficiently small that the phase relationships between the loudspeakers are no longer of primary importance in sound localization. Under these conditions, what matters is the directional behaviour of the energy field around the listener. It is possible to show that, because of the positive nature of energy (in the mathematical sense), one can only exactly recreate the energy field of a live sound source through a small number of loudspeakers if the sound happens to be at the position of one of these. Thus at mid and high frequencies, not all of the ear's localization mechanisms can be satisfied in a practical reproduction system.

However, it is possible to analyse the directional energy field into omnidirectional and vector components analogous to those used for the sonnd amplitude field at low frequencies. If one assumes that the effect of head movement is nsed by the brain, these sound energy components can be used to estimate the probable subjective mid- and high-frequency sound direction. For a sound reproduced through several speakers, this direction may be calculated as the direction of the sum of vectors, one pointing at each speaker, each having as length the energy of the sound from that speaker. Calculations using this theory indicate that various four-speaker sound reproduction systems give the mid-high frequency sound localizations shown in Fig. 4, which agrees well with experimental data[26].

Note that if the number of channels equals the number of speakers (as for "discrete" and QMX via four speakers), then phantom inter-speaker sounds are drawn toward the nearest speaker. Cooper[31, 32] has called this the "detent" effect, but it is not significant for his BMX (two-channel) or TMX (three-channel) systems. A similar "pull" by the speakers is found for tetrahedral with-height reproduction (Fig. 3), but not when a cube of speakers is used.

The ratio of the length of the above-defined energy vector to the total reproduced energy should ideally be unity; iu practice the larger it is the better defined the sound image—it is this that makes TMX better than two-channel BMX.

This mid-high frequency theory holds only so long as the ears do not have too great a directionality in their response to sounds. The data of Sivian & White[17] and Rolls[19] on the ear's directionality show that above about 5kHz a new theory is needed.

Localization above 5KHz

In 1907, Rayleigh[11] found that when the head was stationary the ability to distinguish front from rear relied entirely on high frequencies. This has been confirmed by Stevens & Newman[14] and Roffler & Butler[15], who showed that the ears could localize sounds in the plane of symmetry of the human head quite accurately despite the two ears receiving the same sound waveform! This ability disappeared when the pinnae were masked. Conversely, many workers have found that dummy head recordings (which incorporate the effect of the pinnae's acoustic obstruction) give good spatial localization when reproduced either via headphones or via loudspeakers with the pinnae masked[33]. Perhaps using the ultimate "purist" microphone technique, Edmund Rolls of Oxford University has made similar recordings using microphones inside the ears of real heads!

The pinnae localization mechanism is not well understood, but appears to rely on the fact that sounds from each direction arrive inside the listener's ear with a distinctive colouration. Thus, if we can reproduce that colouration in a

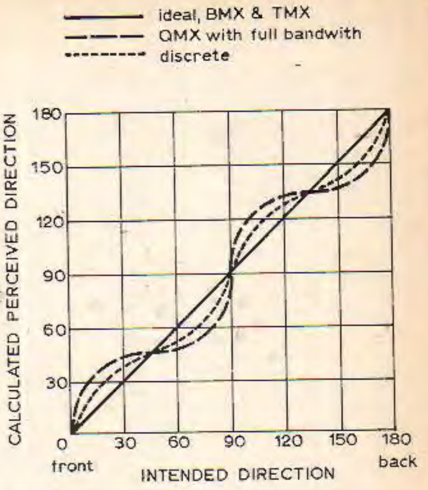


Fig. 4. Perceived localization vs intended direction of sounds in degrees, according to the mid-high frequency theory of this paper, for various systems via a square of speakers as in Fig. 2. Triangles indicate speaker positions. QMX data only applies for a full bandwidth system. Compare with Figs 19 and 20 of reference 26.

recording, we can reinforce the sense of direction created; to the author's knowledge, this has not yet been done in surround-sound recordings.

Reverberation to aid localization

It is possible to locate sounds more accurately in a moderately reverberant room than when there is no reverberation. Although the mechanism is not understood, it is found that correctly recorded reverberation also aids sound localization during reproduction[34], although poor artificial reverberation makes the sound image more indistinct. The author has computed the distribution of reverberation energy around the listener given by various recording techniques[34], and it is found that the most accurate sound localization is obtained when the energy is uniformly distributed, and not concentrated too much in any one direction.

Thus if a surround-sound system is to work optimally, it must be capable of capturing all nuances of reverberant sound and of reproducing these uniformly around the listener. Certain popular commercial matrix systems assign the original sound field to the two available channels in such a discontinuous manner[8, 9] that these criteria cannot be satisfied. "Variable matrix" or "logic" decoders, which work by pushing the whole sound field towards those directions in which the sound is momentarily strongest, clearly cannot reproduce those nuances of reverberation needed by the ears to localize sounds. The "detent" effect of discrete reproduction (Fig. 4) also prevents uniformly distributed reverberation.

Acknowledgment

This article is a revision of a paper by Michael Gerzon given at the 1974 Festival du Son, Paris. (Published in French in Conférences des Journées d'Etudes 1974 du Festival du Son—Editions Radio.)

References

Abbreviations JAES and JASA mean Journal of the Audio Engineering Society, and Journal of the Acoustical Society of America, respectively.

1. Fletcher, H. Stereophonic sound film system—general theory, *JASA*. vol. 13, 1941, pp. 88–99.
2. Blumlein, A. D. British Patent 394,325 (1931).
3. de Boer, K. Stereophonic sound production, *Philips Tech. Rev.*, vol. 5, 1940, pp. 107–44.
4. Shorter, G. Four-channel stereo, *Wireless World*, vol. 78, 1972, pp. 2–5, 54–7. See also *Wireless World Annual*, 1975, pp. 84–9.
5. Cooper, D. H. & Shiga, T. Discrete-matrix multichannel stereo, *JAES*, vol. 20, 1972, pp. 346–60.
6. Gerzon, M. A. Periphony: with-height sound reproduction, *JAES*, vol. 21, 1973, pp. 2–10.
7. Scheiber, P. Analyzing phase-amplitude matrices, *JAES*, vol. 19, 1971, pp. 835–9.
8. Fellgett, P. B. Perspectives for surround sound, *Hi-Fi Sound Annual*, 1974.
9. Fellgett, P. B. Japanese regular matrix, *Hi-Fi News*, Dec., 1972.
10. Gerzon, M. A. Principles of quadraphonic recording (in two parts), *Studio Sound*, Aug. & Sept., 1970.
11.* Strutt, J. W. (Lord Rayleigh). On our perception of sound direction, *Phil. Mag.*, vol. 13, 1907, pp. 214–32.
12.* Strutt, J. W. Our perception of the direction of a source of sound, *Nature*, vol. 14, 1876, pp. 32, 33.
13.* Strutt, J. W. Acoustical observations—1, *Phil. Mag.*, 1877, pp. 456. 457.
14. Stevens, S. S. & Newman, E. B. Localization of actual sources of sound, *Amer. J. Psychol.*, vol. 48, 1936, pp. 297–306.
15. Roffler, S. K. & Butler, R. A. Factors that influence the localization of sound in the vertical plane, *JASA*, vol. 43, 1968, pp. 1255–9.
16. Harwood, H. D. Stereophonic image sharpness, *Wireless World*, vol. 74, 1968, pp. 207–11.
17. Sivian, L. J. & White, S. D. *JASA*, vol. 4, 1933, pp. 296–8.
18. Wiener, F. M. On the diffraction of a progressive sound wave by the human head, *JASA*, vol. 19, 1947, pp. 143–6.
19. Rolls, E. (private communication).
20. Clark, H. A. M., Dutton, G. F. & Vanderlyn, P. B. The stereosonic recording & reproducing system, *I.R.E. Trans. on Audio*, 1957, pp. 96–111.
21. Bauer, B. B. Phasor analysis of some stereophonic phenomena, *JASA*, vol. 33, 1961, pp. 1536–9.
22. Makita, Y. On the directional localization of sound in the stereophonic sound field, *EBU Review*, part A no. 73, 1962, pp. 102–8.
23. Leakey, D. M. Some measurements on the effects of interchannel intensity and time difference in two-channel sound systems, *JASA*, vol. 31, 1959, pp. 977–87.
24. Tager, P. G. Some features of physical structure of acoustic fields of stereophonic systems, *JSMPTE*, vol. 76, 1967, pp. 105–10.
25. Fellgett, P. B. Directional information in reproduced sound, *Wireless World*, vol. 78, 1972, pp. 413–7.
26. Kohsaka, O., Satoh, E. & Nakayama, T. Sound-image localization in multichannel matrix reproduction, *JAES*, vol. 20, 1972, pp. 542–8.
27. Gibson, J. J., Christensen, R. M. & Limberg, A. L. R. Compatible f.m. broadcasting of panoramic sound, *JAES*, vol. 20, 1972, pp. 816–22.
28. Eargle, J. M. Multichannel stereo matrix systems: an overview, *JAES*, vol. 19, 1971, pp. 552–9.
29. Gerzon, M. A. Experimental tetrahedral recording (in three parts), *Studio Sound*, Aug., Sept. & Oct. 1971.
30. Bauer, B. B., Gravereaux, D. W. & Gust, A. J. Compatible stereo-quadraphonic (SQ) record system, *JAES*, vol. 19, 1971, p. 641.
31. Cooper, D. H. Proposal for QMX discrete/matrix carrier-channel disc, privately circulated report, July 15, 1972.
32. Cooper, D. H., Shiga, T. & Takagi, T. QMX carrier-channel disc, *JAES*, vol. 21, 1973, pp.614–24.
33. Sennheiser Kunstkopf-Stereofonie, 45 rev./min. record, Sennheiser Electronic, 3002 Bissendorf (1973).
34. Gerzon, M. A. Recording techniques for multichannel stereo, *British Kinematography, Sound & Television*, vol. 53, 1971, pp.274–9.

*Refs 11–13 are in: Lord Rayleigh, Scientific Papers, Dover Publications, New York.

Integrated injection logic

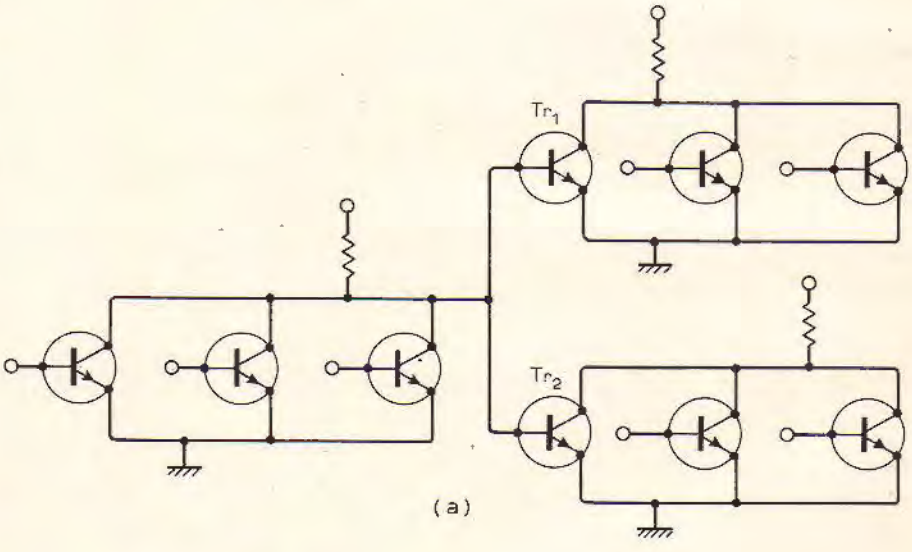
The development of new techniques in circuit integration has apparently been concentrated in the field of m.o.s. devices, and the amount of information appearing in the technical press about m.o.s. has tended to obscure the latest arrival on the bipolar logic field—integrated injection logic (i^2.l. for short). Its characteristics are impressive and it seems set to take over from conventional t.t.l. circuitry when packing density and low power dissipation are the essential requirements of a system.

As a result of the elimination of passive components in the basic gate and a reduction in the number of devices per gate, up to 3000 gates can be fabricated in one chip—an increase by a factor of ten over t.t.l. chips. The speed of i^2.l. is lower than that of t.t.l. (delay around 30ns instead of 10ns) but the speed-power product is only about 0.4pJ or less for i^2.l., compared with 100pJ. Cost is lower than in i.cs using the m.o.s. technology, particularly so as the same chip can coutain both digital and analogue circuits.

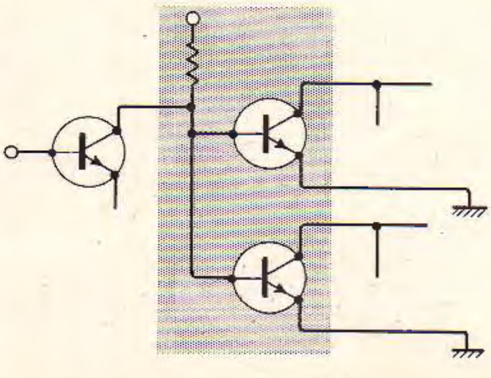
The circuit takes the form of a radically rationalized direct-coupled-transistor-logic (d.c.t.l.) element. In the diagram at (a), a typical d.c.t.l. gate (on the left) is shown driving one input of two other gates. Rearranging the interface gives (b) in the drawing, which can be further simplified by replacing the base resistor by an active current source and by substituting a multi-collector transistor for those with common bases. The result is (c), where the input emitter is termed the injector, the whole circuit being contained within the area of a t.t.l. multi-emitter input transistor. The combining of the two base emitter junctions of the interface gives protection against the effect, when junction voltages on different chips differ, of one gate monopolizing the current output from the previous gate, starving others connected in parallel.

The basic gate can operate at a current of around 1nA and a logic swing of 0.6V, which means interface circuits are needed between i^2.l. and other logic systems or linear devices. Variations of voltage and current can be obtained for different applications.

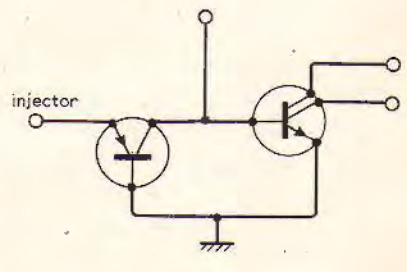
The new logic family can be used in a similar range of work as other l.s.i. systems. It was originated by Philips at Eindhoven, Netherlands, and at about the same time, but independently, by IBM at Boblingen.



(a)



(b)



(c)