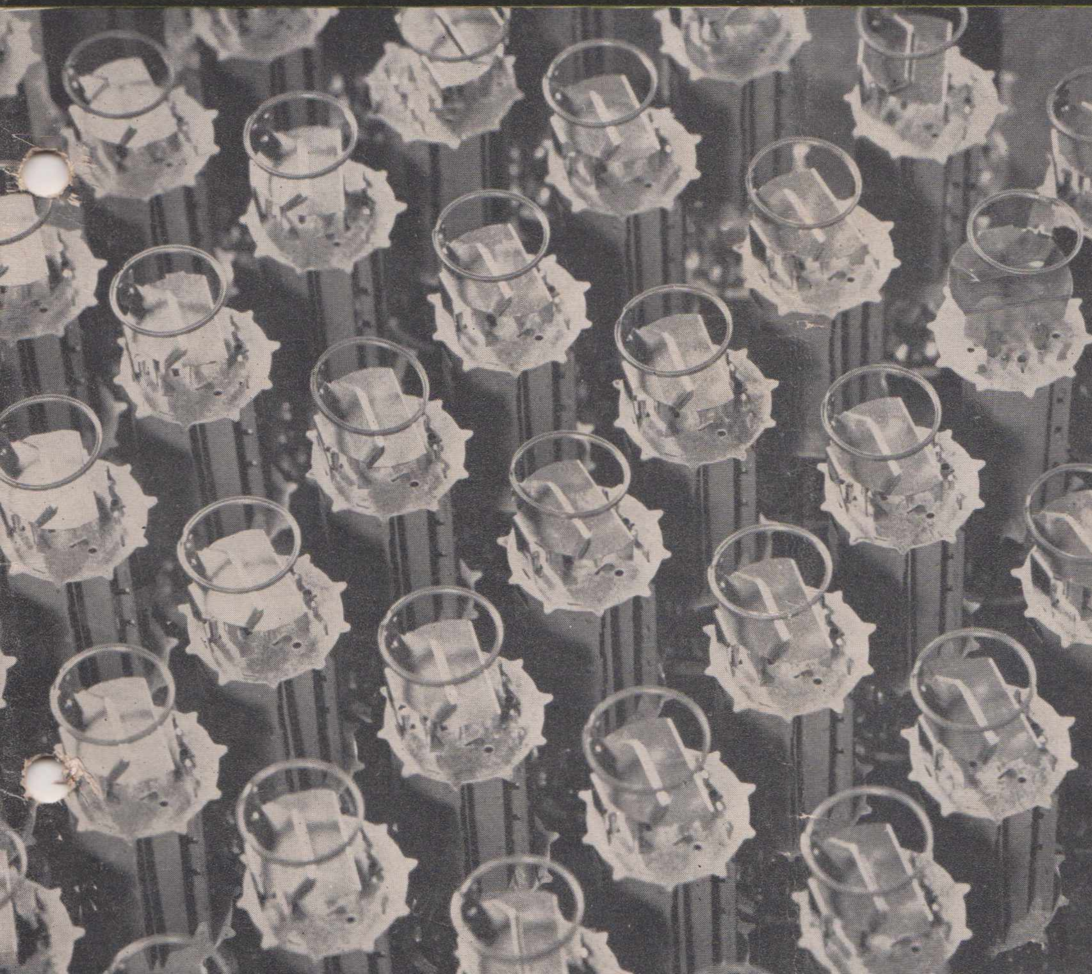
RADIOTRONICS



IN THIS ISSUE

Physical limitations on frequency and power parameters of transistors 2

High Fidelity Topics 11

Microcircuit - Microwatt design techniques for internal medical sensors 17

COVER:
Receiver Valve Assemblies during production process.

Vol. 31, No. 1 January, 1966

1

REGISTERED IN AUSTRALIA FOR TRANSMISSION BY POST AS A PERIODICAL

Physical limitations on frequency and power parameters of transistors

By E. O. JOHNSON

A simple analysis shows that the ultimate performance limits of a transistor are set by the product $Ev_s/2\pi$, where E is the dielectric breakdown strength of the semiconductor material and v_s is its minority-carrier saturated drift velocity. This product, which has a value of about 2×10^{11} volts per second for silicon, emphasizes that a semiconductor material has a maximum capability for energizing the electric charges that process a signal. If the device operating frequency is high, the frequency time period is short and only a small amount of energy can be given to a charge carrier. Consequently, the power and power amplification must be relatively low. At low frequencies the inverse is true, i.e., device physics demands an inverse relation between frequency and power parameters that is independent of the thermal-dissipation arguments commonly given to explain the trade-off between these parameters.

This analysis leads to an effective means for making comparisons between existing devices.

It seems reasonable to suppose that an ultimate limit exists in the trade-off between the volt-ampere, amplification, and frequency capabilities of a transistor. This trade-off should somehow be related to material parameters. A trade-off relation linking these parameters would be useful if it could be derived in a manner that makes it independent of device design details, but yet applicable as a practical yardstick for evaluating device design refinement.

J. M. Early[1, 2, 3] and others[4, 5, 6] have considered this subject with more emphasis on device design details than upon the general physical principles. One objective of this paper is to derive general relations that show the performance limits of transistors independently of design details. Another objective is to demonstrate the use of these relations in comparing existing devices and predicting the trend of future developments.

These objectives will be accomplished first by discussing a basic voltage-frequency relation for semiconductor devices. This relation is extended to include space-charge constraints on device current. Combination of voltage and current constraints leads to a relation which links maximum volt-amperes, device impedance level, and cut-off frequency. Device power gain is introduced by charge-control considerations. This approach leads to a relation linking power gain, dynamic range, and crossmodulation characteristics.

The mode of approach is to establish upper bounds on transistor performance by developing a highly idealized and simplified device model whose performance is not likely to be surpassed by that of any attainable design, no matter how optimized or cleverly conceived. This idealized performance is expressed in terms useful as a yardstick in comparing existing devices and highlighting their shortcomings. This mode of approach should help complement the more usual one in which performance improvements are extrapolated from existing art.

VOLTAGE—FREQUENCY RELATION

The charge-carrier transit-time cutoff frequency f_T of a charge-control type of device[7], such as a transistor, is defined by the relation $f_T = (2\pi\tau)^{-1}$, where τ is the average time for a charge carrier moving at an average velocity v to traverse the emitter-collector distance L. For a given value of L, the transit time τ is minimized when v has its maximum possible value. For a semiconductor this maximum possible value is the saturated drift velocity v_s, which is approximately 6×10^6 centimeters per second for holes and electrons in common semiconductor materials such as silicon and germanium.[8] This limiting velocity is reached at fields of the order of 10^4 volts per centimeter. With the carriers moving at velocity v_s, the transit time can be reduced even further by decreasing the distance L. The lower limit on L, however, is reached when the value of V/L (where V is the applied emitter-collector voltage) becomes equal to the dielectric breakdown field E. This value is approximately 10^5 volts per centimeter in Ge and about twice as much in silicon.[9] It can thus be concluded that the best possible trade-off between the cutoff frequency f_T and the maximum allowable applied voltage V_m is given by:

$$V_m f_T = \frac{E v_s}{2\pi} = \begin{cases} 2\times10^{11}\ \text{V/sec for silicon} \\ 1\times10^{11}\ \text{V/sec for germanium} \end{cases} \quad (1)$$

This relation[10] defines the upper limit on cutoff frequency because the minimum value of V_m must have some value, say one volt, that is sufficiently greater than thermal voltage to insure normal transistor collector action of the base-collector junction.

Practically attainable frequencies are substantially less than the maximum possible frequencies indicated by Eq. (1) because (1) the limiting velocity v_s is not reached in all parts of the charge carrier path, (2) the electric field stress is not geometrically uniform, and (3) the practical technology of small dimensions has its own limitations. In practical circuit applications the maximum applied collector voltage is usually kept to a value of approximately one-half of V_m to provide an adequate safety margin from voltage transients and from the instability effects caused by stray currents flowing to the base from the collector.

With respect to item (1), today's high-performance transistor designs are such that the collector depletion-layer thickness is two or three times the base thickness. Correspondingly, the velocity v_s is reached approximately two-thirds to three-fourths of the carrier path. As a result, v_s is reduced by a factor of about 0.7. With respect to item (2), excluding semiconductor surface effects, the average electric field in the collector-depletion layer ranges from one-half to about one-third of the maximum electric field, depending upon the impurity profile. Taking into account the relatively smaller field in the base region, the combined effect of items (1) and (2) is to reduce performance to a value, which at best, is no greater than about one-fourth of that indicated by Eq. (1). If the effect of the semiconductor surface on junction electric fields is taken into account, performance could be reduced by another factor of perhaps as much as two or three, particularly in higher-voltage devices. However, use of the proper geometrical contour in the region where the junction intercepts the surface of the semiconductor can reduce or eliminate this effect.[11]

Table I lists the V_m and f_T values for a variety of available transistors made from both silicon and germanium and spanning a wide range of frequency, current, voltage, and power capability. The other parameters listed in the table will be considered later in the paper. The factual data in the table was obtained from handbooks and various commercial literature. This data, it should be noted, contains varying degrees of design and manufacturing conservatism. Fig. 1 compares the V_m and f_T values for these transistors with Eq. (1).

As expected, the performance of all of the devices falls below the theoretical line. Also, as might be expected, the most recently developed devices tend to lie closest to this line. Silicon devices in almost every case are closer to the line than germanium devices (circles), probably because of the lower value of E for germanium and the fact that the germanium devices plotted in the figure are mostly of older designs. Grown-junction silicon transistors, the oldest devices of all and not shown, lie farthest away from the theoretical curve.

There is a pronounced tendency for performance to fall away from the theoretical curve in the lower frequency domain. This tendency arises from a combination of factors. First, there is substantial difficulty in dealing with the lightly doped semiconductor material necessary for high collector breakdown voltages.[9] This difficulty stems both from the bulk material, itself, and from the relatively greater surface effects experienced with lightly doped material. Second, there is the fact that the greatest amount of engineering effort, particularly developmental, has been applied in the high-frequency, relatively low-voltage, domain. Third, many low-frequency devices are not specifically designed to accommodate high voltages because such voltages are not required in most typical applications.

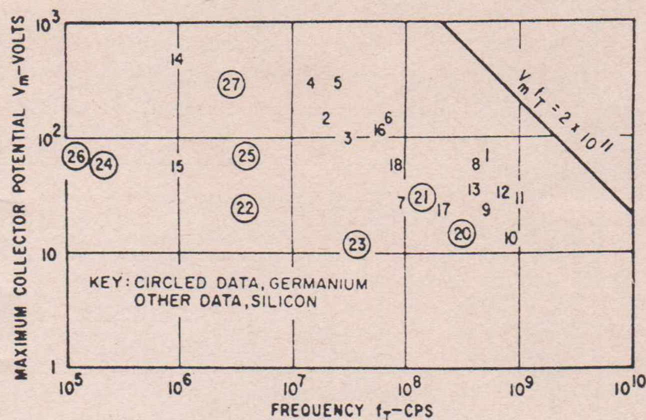


Fig. 1 - Voltage-Frequency Relation for Transistors.

To some degree the locations of points on Fig. 1 are pessimistic, particularly for the germanium alloy devices. The values of their cutoff frequencies quoted in the literature are usually measured at collector potentials several-fold below V_m. If measured at V_m, the values

Table I — DEVICE DATA

Device	Mat'l	Con-struc-tion*	Function	Factual Data: Cutoff Freq. f_T (cps)	Factual Data: Max. Volt. V_m (volts)	Factual Data: Max. Curr. I_m (amps)	Factual Data: Output Cap. C_o (pf)	Calculated Data: $P_m = V_m I_m$ (watts)	Calculated Data: $X_c f_T$ (ohms)	Calculated Data: $I_m X_c$ (volts)	Calculated Data: $(P_m X_c)^{1/2}$ (volts)	Calculated Data: $(P_m X_c)^{1/2} f_T$ (volts/sec)
1. 2N3375	Si	OE	pwr rf	$5x10^8$	65	1.5	10	97.5	32	48	56	$2.8x10^{10}$
2. 2N3265	Si	DDE	pwr rf	$2x10^7$	150	25	900	3750	8.85	221	182	$3.6x10^9$
3. TA2400#	Si	DDE	pwr rf	$3x10^7$	100	25	500	2500	11.6	290	170	$5.1x10^9$
4. TA2496#	Si	EM	pwr sw	$1.5x10^7$	300	10	400	3000	26.5	265	282	$4.2x10^9$
5. TA2301#	Si	EM	pwr aud	$2.5x10^7$	300	0.1	5	30	1270	127	195	$5.0x10^9$
6. TA2529#	Si	EM	video dr	$7x10^7$	150	$5x10^{-2}$	2.5	7.5	910	46	82.6	$5.8x10^9$
7. 3N98	Si	MOS	FET	10^8	30	$15x10^{-3}$	1.5	0.45	1060	16	21.8	$2.2x10^4$
8. 2N2477	Si	DDEP	core sw	$4x10^8$	50	0.5	10	30.0	39.5	20	34.3	$1.37x10^{10}$
9. 2N2938	Si	DDEP	core sw	$5x10^8$	25	0.5	4	12.5	68.5	34	29.2	$1.5x10^{10}$
10. 2N2475	Si	DDEP	logic	$8x10^8$	15	0.5	3	7.5	66.5	33.3	22.4	$1.8x10^{10}$
11. 2N2857	Si	DDEP	rf amp	10^9	30	$2x10^{-2}$	1.8	0.6	88.5	1.8	7.3	$7.3x10^9$
12. 2N2708	Si	DDEP	rf amp	$7x10^8$	35	$2x10^{-2}$	1.5	..7	151	3	10.3	$7.2x10^9$
13. TA2600#	Si	OE	pwr rf	$3.5x10^8$	36	5	25	180	18.2	91	57.2	$2.0x10^{10}$
14. 2N689	Si	D	Thyristor (SCR)	10^6	500	16**	50	8000	3200	51200	5060	$5.06x10^9$
15. 2N1513	Si	SD	pwr sw	10^6	60	8	200	480	800	6400	620	$6.2x10^8$
16. 2N2102	Si	TDP	gen purp	$6x10^7$	120	1	15	120	177	177	145	$8.7x10^9$
17. 2N706	Si	DDP	logic	$2x10^8$	25	10^{-1}	6	2.5	133	13	18.2	$3.6x10^9$
18. 2N696	Si	DDP	pwr sw	$8x10^7$	60	0.5	20	30	100	50	54.8	$4.3x10^9$
19. 2N2016	Si	D	power	$2.5x10^4$	130	10	400	1300	15400	154000	4500	$1.1x10^8$
20. 2N960	Ge	EM	logic	$3x10^8$	15	10^{-1}	4	1.5	133	13	14.1	$4.2x10^9$
21. 2N1177	Ge	AD	rf amp	$1.4x10^8$	30	10^{-2}	2	0.3	570	5.7	13.1	$1.8x10^9$
22. 2N404	Ge	A	logic	$4x10^6$	25	10^{-1}	12	2.5	3310	331	91	$3.6x10^8$
23. 2N1301	Ge	AM	logic	$3.5x10^7$	13	10^{-1}	8	1.3	570	57	17.2	$9.5x10^8$
24. 2N301	Ge	A	power	$2x10^5$	60	10	300	600	2650	26500	1250	$2.5x10^8$
25. 2N2147	Ge	DA	power	$4x10^6$	75	5	400	375	100	500	194	$7.8x10^8$
26. 2N1358(174)	Ge	A	power	10^5	70	15	400	1050	4000	60000	2050	$2x10^8$
27. TA1928#	Ge	DA	pwr sw	$3x10^6$	320	10	300	3200	177	1800	820	$2.5x10^9$

***KEY:**
- **OE** — **Overlay Epitaxial.**
- **DDE** — **Double-Diffused Epitaxial.**
- **EM** — **Epitaxial Mesa.**
- **MOS** — **Metal - Oxide - Semiconductor (Field Effect Transistor).**
- **DDEP** — **Double-Diffused Epitaxial Planar.**
- **D** — **Diffused.**
- **SD** — **Single-Diffused.**
- **TDP** — **Triple-Diffused Planar.**
- **DDP** — **Double-Diffused Planar.**
- **AD** — **Alloy-Diffused.**
- **A** — **Alloy.**
- **AM** — **Alloy Mesa**
- **DA** — **Drift-Alloy.**
- ****** — **Average Current, Peak Current, Ten-fold Higher.**
- -/-/- — **The Type Number Identifies a Particular Laboratory Device Design. The Type Number and Tentative Data are Subject to Change. No Obligations are Assumed for Notice of Change or Future Manufacture of the Product Unless Otherwise Arranged.**

of f_T would be somewhat greater than those listed in Table I. This effect is far less pronounced in the diffused-collector junction devices because their base-collector depletion region tends to expand into the collector region, rather than into the base region as is the case with the alloy devices. Consequently, the effect of f_T is much less.

If device designers can extrapolate today's design perfection to higher frequencies, they will find an upper f_T limit at about 20 gigacycles. This value can be obtained from Fig. 1 by linear extrapolation from point "1", or "11", parallel to the theoretical curve and down to the $V_m = 1$ volt line, which is the nominal minimum for normal transistor operation. At this 20-gigacycle frequency limit the emitter-collector spacing will have to be of the order of 1000 angstroms.

The thyristor or SCR (point "14") is included for comparison purposes that will become clearer later. The developmental field-effect transistor (point "7") is included because it is a charge-control device and can be directly compared with the conventional bipolar transistor.[7]

CURRENT—FREQUENCY RELATIONS

The load current I through a charge-control device is basically defined by:

$$I = \frac{Q}{\tau} \qquad (2)$$

where Q is the total mobile charge in the emitter-collector region flowing to the collector, and τ is the average charge transit time as defined previously. If τ is considered as a fixed quantity, then the ratio I/Q is also a constant:

$$\frac{I}{Q} = \frac{1}{\tau} = \text{constant} \qquad (3)$$

and I will increase linearly with Q. In a transistor the largest possible value of I is attained when Q is approximately equal to the total fixed charge Q_f in the emitter-collector space. If Q is somehow made to exceed Q_f, the current I becomes space-charge-limited, as in a vacuum tube, and the controlling action of the base is lost. This condition sets an absolute upper limit on I.

In practice this limit is not reached because the mobile charge density first exceeds the fixed charge density in local regions of the emitter-collector path, for example, at the collector edge of base. This effect leads, as is well known, to the frequency-reducing phenomenon of base widening. It can also lead to another deleterious effect, an increase in the electric field in the collector depletion layer which reduces the maximum voltage that the device will tolerate before field breakdown occurs. The maximum rated current I_m of a device may be well below the maximum possible current for other reasons: power dissipation, second-breakdown effects, drop-off in current gain due to decreased emitter efficiency, etc.

If the maximum current is defined as the current I_m that causes the onset of significant base widening, Eq. (3) can be written in terms of V_m and a capacitance approximately equal to the usually quoted value of the collector-base capacitance C_o. Accordingly,

$$\frac{I_m}{C_o} = \frac{V_m}{\tau_C} \qquad (4)$$

where τ_C approximates the collector depletion-layer transit time. The ratio I/C is ideally invariant with collector area, doping level, or parallel circuit configurations of devices. For example, an increase in the doping level in the base produces a higher value of C_o and allows a proportionally larger value of I_m. Eq. (4) has some interesting consequences. First of all, if V_m is replaced by its value from Eq. (1), the following relation is obtained:

$$\frac{I_m}{C_o} = \frac{Ev_s}{2\pi f_T \tau_C} = Ev_s \qquad (5)$$

because $2\pi f_T \tau_C$ = approx. $2\pi f_T \tau = 1$ is equal to unity. The quantity I_m/C_o, a convenient measure of device speed ability in switching voltage, has a maximum attainable value that is equal to Ev_s.

Second, if C_o is converted to reactive impedance X at the frequency f_T by use of the relation $C_o = 1/2\pi f_T X$, the following relation is obtained:

$$(I_m X) f_T = \frac{Ev_s}{2\pi} \qquad (6)$$

This relation gives the most optimistic tradeoff between current and frequency that one can expect from a transistor. A comparison of the devices listed in Table I with Eq. (6) is shown in Fig. 2. In marked contrast with Fig. 1, the data does not fall away from the theoretical curve in the low-frequency domain. Existing devices thus appear better at handling high currents than they are at handling high voltages. Devices having an emitter geometry which maximizes the emitter periphery-area ratio (points "1", "9", "10", and "13", for example) minimize current-crowding effects and so tend to be closer to the theoretical curve.

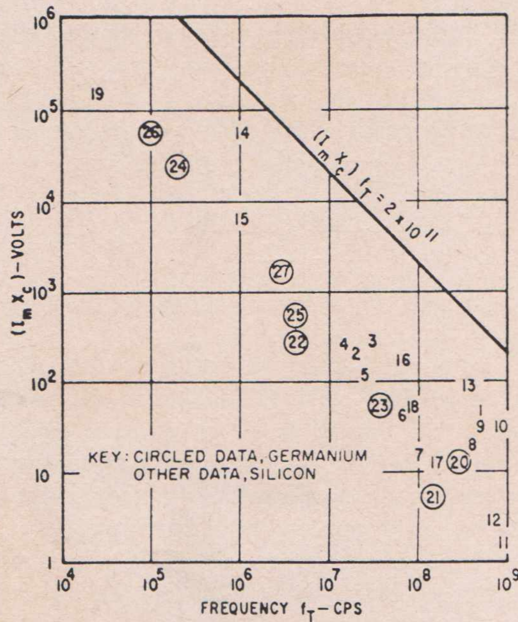


Fig. 2 - Current-Frequency Relation for Transistors.

Device current performance falls below the theoretical curve for reasons in addition to those listed above for the current itself. One reason is that current-crowding effects at the emitter periphery cause the portions of the collector under the center of the emitter to be unused in handling current. Portions of the collector that are off-set from the emitter, such as those regions under base contacts, are also useless for handling load current. For these reasons the actual collector capacitance tends to be several-fold larger than is necessary for handling the rated value of collector current. Package and other stray capacitances, particularly with high-frequency devices, also have the same deleterious effect.

As might be expected, the thyristor or SCR (point "14") is closer to the theoretical curve than any of the other devices. This device is not limited to a fixed I/C ratio because the maximum collector current can be almost arbitrarily large. Collector-current space-charge effects are cancelled by the counter-current of carriers of opposite sign. The counter-flow of carriers of opposite sign creates a dense solid-state plasma that can carry currents not directly limited by the density of fixed charge in the device. Accordingly, a thyristor is inherently superior to a transistor in its volts-per-second switching capability. This superiority is further enhanced by the absence of current-crowding effects; base current is supplied from the collector, rather than through the lateral base resistance as in a transistor. The price one pays for the inherent switching superiority of the thyristor is that current turn-off action becomes cumbersome and relatively slow.

VOLT-AMPERE—FREQUENCY RELATIONS

The relation between volt-ampere product, impedance level, and frequency for a transistor is obtained by multiplying Eq. (1) and (6). This combination gives:

$$(P_m X)^{1/2} f_T = \frac{E v_s}{2\pi} \qquad (7)$$

where the volt-ampere product, $V_m I_m$, is replaced by P_m. J. M. Early[1,2] has shown that the left-hand side of this relation can be derived by simple intuitive arguments. One interesting conclusion from Eq. (7) is that, for a given device impedance, the volt-ampere ability of a device must necessarily decrease as the device cutoff frequency f_T is increased.[12]

This decrease must occur, not because of the power-dissipation arguments usually stated, but because a given semi-conductor material has a limited volts-per-second capability for the charge carriers traversing it. For a device designed to operate at a low frequency, the time period available for energy transfer to the charge carriers is relatively long. Accordingly, energy transfer and power capability can be relatively large. For high-frequency devices the inverse is true. More specifically, the maximum electric field that a charge carrier in the device can be subjected to is the breakdown field E. A charge carrier can traverse this field at a maximum velocity equal to v_s. The product Ev_s is the maximum rate at which a carrier can acquire volts of energy. If time is long, then the energy transfer can be large, and vice versa.

The power-dissipation argument, sometimes inferred as a principle, holds that device size decreases with frequency capability and that this smaller size leads to decreased power dissipation capability and hence to a decreased ability for handling output power. As a principle this argument is fallacious; Eq. (7) infers that, in principle, an arbitrarily large value of P_m can be achieved, for a given f_T, by connecting devices in parallel. The price paid for this approach is a proportionate decrease in the impedance X. In principle, these devices can be physically arranged in such a distributed array that the heat-dissipation capability can be made almost arbitrarily high. Indeed, practical power output may be limited more by the problems of economics, circuitry, and device uniformity than by heat dissipation, per se.

Eq. (7) points out that when the best power-frequency capability is desired for a transistor it should be designed to operate at relatively high current. This choice is traded for a decrease in impedance level, but not for a decreased frequency capability, as would be the case if the device were designed for low current and high voltage. Circuit requirements and practical device-design problems limit how far one can proceed in the direction of high current.

Fig. 3 compares the performance of the devices in Table I with Eq. (7). Again, as in Fig. 1, and for the same reasons, performance tends to diverge more from the theoretical at the lower f_T values. Also, the superiority of silicon devices compared to germanium devices is strongly implied. The most recently developed silicon power-transistor designs using a large emitter periphery-area ratio are closest to the theoretical curve for the reasons given previously.

RELATIONS INVOLVING AMPLIFICATION PROPERTIES

The maximum available power gain G_p of a charge-control device, such as a transistor, is given by:[7]

$$G_p = G_I^2 \frac{Z_o}{Z_{in}} = \left(\frac{f_T}{f}\right)^2 \frac{Z_o}{Z_{in}} \quad (8)$$

where G_I is the current amplification ($G_I = 1/2\pi f\tau = f_T/f$) and Z_o and Z_{in} are the output and input impedances, respectively. If the idealizing assumption is made that no electrode series resistances exist, Eq. (8) can be approximated by:

$$G_p = \left(\frac{f_T}{f}\right)^2 \frac{C_{in}}{C_o} \quad (9)$$

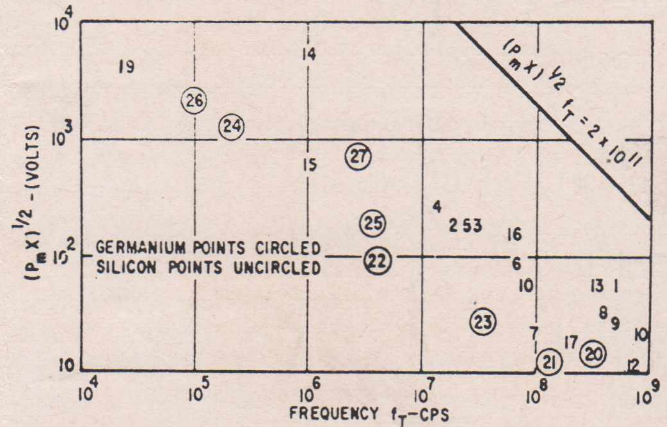


Fig. 3 - Power-Frequency Relation for Transistors.

where C_{in} is the input capacitance and C_o is the output (base-collector) capacitance. Furthermore, if it is assumed that the emitter diffusion capacitance C_d dwarfs the emitter transition capacitance, then the maximum value of the input capacitance is given by:

$$C_{in} = C_d \cong \frac{Q_m}{V_T} = \frac{I_m \tau_b}{V_T} \quad (10)$$

where Q_m is the maximum total carrier charge flowing to the collector, τ_b is the carrier base transmit time, and V_T is the thermal voltage ($KT/_e$ volts). Other quantities have been previously defined. The output capacitance C_o is found from Eq. (4):

$$C_o = \frac{I_m \tau_c}{V_m} \quad (11)$$

The maximum capacitance ratio is thus defined by Eq. (10) and (11), and Eq. (9) can be rewritten as follows:

$$G_p = \left(\frac{f_T}{f}\right)^2 \frac{V_m}{V_T} \frac{\tau_b}{\tau_c} \cong \left(\frac{f_T}{f}\right)^2 \frac{V_m}{V_T} \quad (12)$$

The frequency ratio accounts for the input-output charge amplification of the device; the voltage ratio accounts for the energy step-up per carrier. That is, in the ideal case an energy of V_T is required to place a charge on the base control electrode; the charges which flow to the collector as a consequence ideally pick up an energy equal to V_m.

Eq. (12) can be written in a form consistent with the other performance trade-off relations by use of Eq. (1):

$$(G_p V_T V_m)^{1/2} f = \frac{E v_s}{2\pi} \qquad (13)$$

In this equation, G_p is evaluated at some operating frequency f sufficiently close to f_T for Eq. (8) to be valid. The value of V_m corresponds to that used in Eq. (1). Eq. (13) emphasizes the fact that power amplification, like volt-ampere performance, basically depends upon the volts-per-second capability of the semiconductor material.

The actual performance of a transistor falls far short of that predicted by Eq. (13). First of all, the qualifications on practically attainable electric fields and average carrier velocities noted for Eq. (1) apply. These qualifications reduce the right-hand constant to an effective value equal to about one-fourth, or less, of the theoretical value. Second, the practical working value of applied collector voltage is about one-half of V_m (see Table II). Third, Eq. (10) over-estimates the value of C_d, and hence G_p, by perhaps as much as two times in a device operating at very high carrier densities. This decrease stems from an electric-field enhanced carrier velocity in the base. Fourth, this same electric field in the base doubles the input-energy requirement per carrier. Fifth, another factor of at least two is added because of ohmic base resistance; the energy lost in the resistance of a series RC circuit is equal to the potential energy acquired by the charges on the capacitor, when the charging time is a few-fold or more larger than the RC time constant. The energy loss in the base resistance can be very much greater when the driving-signal time constant is less the RC time constant of the circuit.

TABLE II - DEVICE POWER-GAIN DATA

Device	Material	Construction*	Factual Data					Calculated Data
			Operating Freq. (f)	Power Gain (G_p)		Voltage		$(V_m V_T G_p)^{1/2}$
						Operating	Max. (V_m)	
			(cps)	(db)	(ratio)	(volts)		(volts)
1. 2N1631	Ge	DA	1.5×10^6	47.7	6.0×10^4	12	34	232
2. 2N1425	Ge	DA	4.6×10^5	51.0	1.3×10^5	12	24	280
3. 2N1180	Ge	DA	1.0×10^7	35.0	3.2×10^3	12	30	15.5
4. 2N1638	Ge	DA	2.6×10^5	61.5	1.4×10^6	11	34	1100
5. 2N384	Ge	DA	5.0×10^7	21.0	1.3×10^2	12	40	11.4
6. 2N1066	Ge	DA	5.0×10^7	26.0	4.0×10^2	12	40	20
7. 2N1177	Ge	DA	1.0×10^8	14.0	2.5×10^1	12	30	4.3
8. 2N175	Ge	A	2.0×10^4	43.0	2.0×10^4	4	10	71
9. 2N139	Ge	A	4.6×10^5	37.0	5.0×10^3	9	16	45
10. 2N408	Ge	A	4.6×10^5	37.8	6.0×10^3	9	20	55
11. 2N2873	Ge	M	1.8×10^8	21.6	1.5×10^2	12	35	11.4
12. 2N2708	Si	DDEP	2.0×10^8	22.0	1.6×10^2	15	35	11.8
13. 2N2875	Si	DDEP	4.5×10^8	19.0	8.0×10^1	6	30	7.8
14. 3N98	Si	MOS	6.0×10^7	10.0	10	20	30	34.6

***KEY: DA — Drift Alloy.**
A — Alloy.
M — Mesa.
DDEP — Double-Diffused Epitaxial Planar.
MOS — Metal Oxide Semiconductor — Field Effect Transistor.
As noted in text, the voltage corresponding to V_T for the MOS is approximately the pinch-off voltage. Voltage use in calculation was 4.0 volts.

Transition capacitance accentuates the problem because it increases the time constant of the transistor input circuit. The effect of all these factors is to make the right-hand constant in Eq. (13) at least thirty-fold less than the theoretical constant.

A plot of data in Table II is shown in Fig. 4 along with the theoretical curve corresponding to Eq. (13). The points closest to the theoretical curve correspond to the most recently developed devices and are roughly a hundred-fold removed. The next closest group of points, "1" through "7", correspond to the next most recently developed devices, the alloy-drift types. The farthest removed group of points, "8", "9", and "10" for the alloy devices, correspond to the oldest device types.

One reason for the tendency for the points to diverge increasingly from the theoretical curve at lower frequencies has to do with the previously discussed factors involving V_m which apply to the results in Fig. 1. Another reason is that circuit-instability problems tend to limit the amount of power gain sought by the device designer.

If a speculative extrapolation to higher frequencies is made with Fig. 4, as was done for the data in Fig. 1,

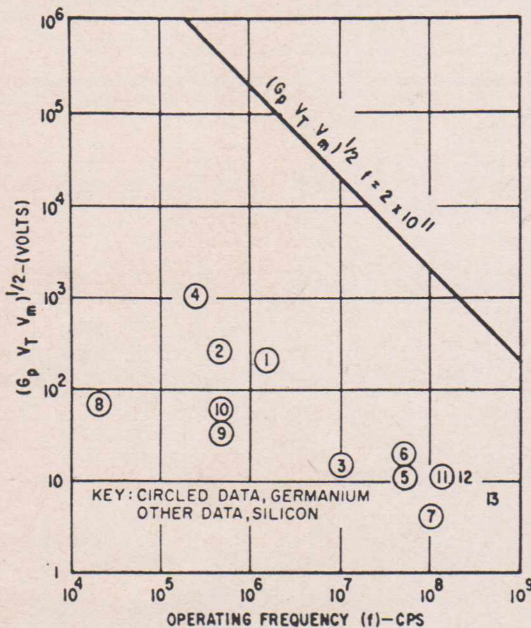


Fig. 4 - Power-Gain and Frequency Relation for Transistors.

at an operating frequency of 20 gigacycles the value of $(G_p V_T V_m)^{1/2}$ is somewhat less than unity and G_p is thus limited to a maximum value of about ten.

Eq. (12) and (13) suggest that one must trade power gain for the energy required to place a charge on the control electrode.[7] For the bipolar transistor this input energy has been noted as V_T volts; for the field-effect transistor this energy can be associated with the pinch-off voltage V_p, or the G_m/I ratio, and is ten to a hundred times greater than V_T. Both V_p and V_T approximately define the useful dynamic input voltage range of field-effect and bipolar devices. Simple consideration shows that the larger this dynamic input-voltage range, the smaller tends to be the curvature components on the transfer characteristics. Consequently, the smaller will be the tendency for cross modulation phenomena in the device, particularly when the device operating point is swept along the transfer characteristic by agc action.[13]

Thus, it can be concluded that a bipolar transistor will tend to have more power gain at a given operating frequency than a field-effect transistor when both devices are constructed to have the same value of f_T. On the other hand, the field-effect transistor will have a greater dynamic range and, correspondingly, a greater ability to avoid cross-modulation interference effects. Point "14" in Fig. 4 shows that the field-effect transistor holds its own against the bipolar transistor when the dynamic range factor is taken into account.

CONCLUSIONS

A simple analysis of transistors shows that the product Ev_s, where E is the semiconductor breakdown field and v_s is the maximum carrier drift velocity, is the ultimate measure of transistor volt-ampere, power gain, and frequency performance. This product emphasizes that a given semiconductor material used in a transistor has a definite and fixed capability for imparting volts of energy per second to a charge carrier. For this reason the energy transfer to a charge carrier must necessarily decrease with frequency, as must the volt-ampere and power-gain performance. However, if the designer is willing to pay the price of a decreased impedance level, a transistor or transistor array can, in principle, be made to give an almost arbitrarily high volt-ampere performance at any operating frequency within the frequency limits of the device.

For a given device cutoff frequency f_T, there is a trade-off between power gain and dynamic range. The bipolar transistor will tend to have a superior gain at the given value of f_T than the field-effect transistor. On the other hand, the field-effect transistor will tend to have a superior dynamic range and ability to avoid cross-modulation interference.

The means for illustrating the various trade-off relations can be plotted in a simple manner that compares transistors having a very wide range of parameter values.

REFERENCES

1. J. M. Early; "Maximum Rapidly-Switchable Power Density in Junction Triodes", *IRE Trans. Electronic Devices,* Vol. ED-6, pp. 322-325, July, 1959.
2. J. M. Early; "Speed and Semiconductor Devices", *IRE International Conv. Record,* March, 1962.
3. J. M. Early; "Structure-Determined Gain-Band Product of Junction Triode Transistors", *Proc. IRE,* pp. 1924-1927, December, 1958.
4. J. M. Goldey and R. M. Ryder; "Are Transistors Approaching Their Maximum Capabilities?", *International Solid-State Circuits Conference,* February, 1963.
5. R. L. Pritchard; "Frequency Response of Grounded-Base and Grounded-Emitter Junction Transistors", *AIEE Winter Meeting,* January, 1954.
6. L. J. Giacoletto; "Comparative High-Frequency Operation of Junction Transistors Made of Different Semiconductor Materials", *RCA Rev.,* Vol. 16, pp. 34-42, March, 1955.
7. E. O. Johnson and A. Rose; "Simple General Analysis of Amplifier Devices with Emitter, Control, and Collector Functions", *Proc. IRE,* pp. 407-418, March, 1959.
8. E. J. Ryder; "Mobility of Holes and Electrons in High Electric Fields", *Phys. Rev.,* Vol. 90, pp. 766-769, June, 1953.
9. J. Shields; "Breakdown in Silicon pn Junctions", *J. Electron, Control,* Vol. 6, pp. 130-148, February, 1959. J. Hilibrand has called the author's attention to the decrease in breakdown field as the impurity density is decreased. This effect is described in detail by C. D. Roots, et. al., "Avalanche Breakdown Voltages of Diffused Silicon and Germanium Diodes", *IRE Trans. Electron Devices,* Vol. ED-7, No. 4, pp. 257-262, October, 1960.
10. First pointed out to author by W. M. Webster. Referred to by E. O. Johnson, "Whither the Tunnel Diode", *AIEE-IRE Electron Devices Conference;* October, 1962.
11. R. L. Davies and F. E. Gentry, "Control of Electric Field at the Surface of P-N Junctions", *IEEE Trans. Electron Devices,* Vol. ED-11, pp. 313-323, July, 1964.
12. Varactor action in the collector circuit is not included in this argument.
13. First pointed out to the author by J. Olmstead and J. Avins of RCA.

With Acknowledgement to RCA

OPERATION MULTIFLASH WINNER VISITS AWV FACTORY

Mr. Barry Mitchell, winner of the recent AWV Multiflash competition visited this company's Rydalmere plant late in November. These pictures were taken during his conducted tour which he described as most interesting and instructive.

High Fidelity Topics

An article produced at irregular intervals, consisting of a series of notes dealing with audio topics of general interest, many of them inspired by readers' letters.

By B. J. SIMPSON
A.W.A. Patents Department

Loudspeakers and Amplifier Power

Questions are often asked regarding the use of speakers where the rated power handling capability of the speaker is considerably less than the maximum power output of the amplifier with which it is to be used. I suppose the general answer to this question would be to the effect that this arrangement is quite all right as long as the amplifier is used in such a way as to prevent more power flowing into the speaker than it is capable of handling, and this is generally interpreted in terms of keeping the gain control at a fairly low setting.

However, as in the case of many questions, the more we consider the import of the matter the less it is possible to provide such a simple answer. For example, if a 10-watt speaker is used with a 10-watt amplifier, we know that even under peak signal or maximum drive conditions, very little more than 10 watts can reach the speaker. The clipping which will take place in the amplifier above the rated power output level will limit the power fed to the speaker to a sufficiently small overload as to be considered safe for the speaker, and will naturally and usually exhibit a warning of overload by gross distortion of the signal. It has been assumed here that the same "type" of watts were being talked about when discussing the speaker and amplifier ratings; where a mixed bag of rms, peak, music, peak music and so on is involved, appropriate adjustments will require to be made.

If we now consider the case of the 10-watt speaker used with a 30-watt amplifier, it could be argued that this could be considered safe if the volume control is kept at a low setting, at least until the matter is studied a little further. It will be possible under these conditions, for example, to feed a signal into the amplifier at an average level that would be considered safe, but wherein the peak values could go to about four times the rated input of the speaker. Now it is true that these peaks are transitory, and are generally taken care of in the calculation of mean values. However, there is some reason to feel that the subjection of a speaker to such large transients relative to the rated power input is undesirable and may lead to damage. This thought is based on experience rather than on manufacturers' figures, which do not specifically cover such eventualities.

But the matter is even more serious than this. There is the ever-present danger of a complete overload of the speaker, either by inadvertent misadjustment of the gain control, or by certain types of circuit failure or malfunction. By way of example, one case that was brought to my notice recently, and which to some extent inspired these remarks, illustrates one way in which this practice can be dangerous to equipment.

In this case a speaker rated at about 10 watts, and incidentally an expensive high-quality unit, was used with an amplifier that was capable of delivering a generously-rated 30 watts rms. In fact I know that the particular amplifier is in fact capable of delivering nearly 40 watts rms before appreciable clipping sets in. The first symptom of trouble was generally poor-quality reproduction. The sad owner of the equipment should of course, as we who are gifted with after-sight will know, have switched off quickly to find out what was wrong. Instead he persisted for some time, in fact until the speaker voice coil went open-circuit and brought the experiment to a close. The audio level being used was obviously far below the rated 10 watts of the speaker, based on the known operating conditions at the time, so why did the speaker burn out? No prize is offered for the solution to this query, the answer to which was arrived at by use of a little logic.

Well, it is quite clear that an excessive amount of power fed to the speaker burnt out the voice coil, but this was certainly not audio power, the average level of whch would have been well within the rated capability of the speaker. It follows then that the excess power must have been outside the audio range. Conceivably in some types of direct-coupled semi-conductor amplifiers, the excessive power could have been developed by a dc current through the coil. In many cases this case of a dc current would also cause a catastrophic failure in the amplifier itself; in the event, it was possible to discount this possibility in the case under consideration, which makes things a little easier.

The possibility of the power being at a sub-audio frequency was also quickly discounted, and it was deduced that the amplifier was oscillating at a supersonic frequency. This in fact proved to be the case. The amplifier was oscillating in the 150 to 200 Kc region, the oscillation being sufficiently strong to drive the amplifier to overload. This meant that something of the order of 40 watts continuous was being delivered to the speaker coil, whilst no very obvious indication of this was available to the user, except that the quality of the audio signal to which he was trying to listen was poor. This poor quality would be expected in an amplifier which was already clipping.

It is in fact probable that something less than 40 watts was going into the speaker winding, because the speaker impedance would have been fairly high at the frequency range indicated, and matching of the speaker to the amplifier would have been far from optimum. However, there was still enough energy getting into the speaker to do the damage.

In the case in point the fault was a drift of one component well outside the stated tolerance, which in turn led to instability around the feedback loop. It will be clear that under other circumstances, a similar state of affairs could arise in a newly-built amplifier due to poor layout, or inefficient grounds, wiring error or one of the many similar pitfalls for the unwary.

One point arising from all this is that it seems unwise to build the higher-powered and more sophisticated types of amplifier, unless facilities are available for checking these things out. For example, I never do any work at all on an amplifier without an oscilloscope connected, and I watch it constantly. The same applies when the amplifier is finally installed, until reasonable tests appear to indicate that the system is completely stable. In the case under discussion, for example, the CRO would have indicated immediately the presence of the disastrous oscillation.

For those who wish, and this is the proper way to do it, the amplifier is subjected during bench testing to capacitive loads, say on the lines of the tests laid down in the recommended procedures of the Audio Group of the British Radio and Electrical Manufacturers' Association.[1] But use these methods with caution in certain types of transistorized amplifiers, or further disaster will result. Some configurations, such as the popular series-connected single-ended class B arrangements, may be severely damaged if they are asked to feed appreciable output to a load whose impedance (at the frequency in use) is considerably lower than the rated load impedance of the amplifier.

To get back to the question of speakers, what can be recommended for the condition stated? One thing that could be done would be to provide a partial dummy load for the amplifier. In the case described, where a 15-ohm speaker was used on a 15-ohm output, three 15-ohm resistors and the speaker could be connected in a series/parallel arrangement so that the speaker would receive only one quarter of the power fed into the network from the amplifier. The amplifier would still "see" the correct load, and the power available to the speaker would be limited to something that it could handle safely. A higher power consumption in the amplifier would be involved in practice to produce sound levels comparable with those of the original arrangement, but a little more power is much cheaper than buying new speakers. In any case, one is paying a power penalty whenever an amplifier considerably more powerful than required is used.

Where the amplifier has multiple output taps, a variation of the method just described may be used. For example, the 15-ohm speaker could be used with a 15-ohm resistor in parallel and the whole arrangement connected to the 8-ohm output terminals. This would allow only about one half of the amplifier output to reach the speaker. Naturally the resistor(s) used in these arrangements must be capable of handling their appropriate share of the output power. Many similar arrangements would be possible and will readily occur to readers.

Fusing Speaker Lines

In the type of situation which has just been discussed, and in the case where the output configuration of the amplifier is such that there is a possibility however slight of a malfunction causing a large dc current to flow through the speaker, fuses have been suggested and used. This would seem to be a simple and effective method of protecting the speaker, and the breaking current of the fuse is easily calculated. Where the required value falls between two standard values, then the lower value is indicated as affording the maximum protection, remembering that it is easier to replace a fuse than a speaker.

There are two disadvantages associated with the use of fuses in this type of application, but neither of them is probably strong enough to prevent the use of a fuse. One disadvantage is that when the fuse blows, the amplifier is left without load. Now we realise that this is most undesirable in high power installations and may cause consequential damage. In home systems, however, where the amplifier is generally operated well below maximum output, the amplifier should tolerate this, at least until the user can get to the switch. The other point is that the fuse constitutes a non-linear element outside the feedback loop. Here again, it is assumed that in home systems, the normal operating level of the amplifier is well below the calculated output at which the fuse is expected to blow, so that changes in the resistance of the fuse over the normal operating region should be so small as to be negligible.

I suppose if one wanted to be a perfectionist, or the amplifier was expected to be operated at or near maximum rating for a substantial part of its life, one could go inside the amplifier and put the fuse inside the feedback loop. Some readers may remember that this was done in one of the designs published in these pages.[2]

Two further ways with fuses are possible, both of which are intended to overcome at least in part the disadvantage mentioned that the blowing of the fuse will leave the amplifier without a load. Both methods aim to provide a load for the amplifier when the fuse blows, which is not the rated load, but which may be considered sufficient to avoid damage at the ouput transformer or other components. The first idea is to connect across the output terminals of the amplifier a resistance equal to about 5 to 10 times the nominal speaker impedance, and then to connect in parallel with this resistance the speaker and an appropriate fuse connected in series. The resistance then provides a load when the fuse blows. See Fig. 1(a).

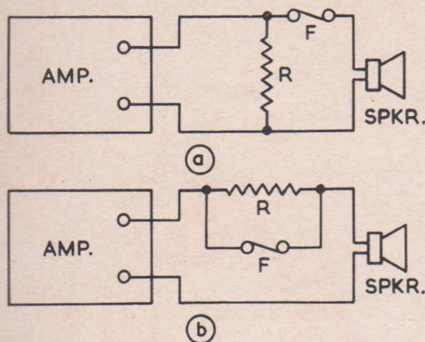


Fig. 1

An alternative which is a little more sophisticated is to connect a fuse in series with the speaker as earlier described, and then to connect a resistance in parallel with the fuse. The resistance would in this case be chosen so as to limit the power dissipated in the speaker under worst conditions to something well within its ratings, but a resistance value of between 3 and 5 times the nominal speaker impedance will generally be suitable. In this case, a low level signal is still available in the speaker when the fuse has blown, assuming that the amplifier is otherwise operative, and provides an indication of the state of the apparatus. See Fig. 1(b).

With many types of semiconductor amplifier, the removal of the load impedance under operating conditions will not result in damage. In such cases measures to protect the amplifier itself, as opposed to the speaker, are not required.

Speaker Ratings

It is usual to find that a speaker manufacturer gives little information about the power that a speaker will handle except a bare power rating figure. Frequently one is even left to wonder whether this figure is an rms or a peak rating, or is one of those disreputable ratings evolved by our brothers overseas. In any case it is most probable that the figure quoted was taken at about 400 to 1,000 cps, and information regarding permissible power input versus frequency is not available.

Now it is recognised that there may be some practical difficulties in quoting such figures, for they can be influenced by amplifier characteristics, baffling and other matters. We also know that the impedance versus frequency characteristic of a speaker is far from being linear. However, in the present state of the art I feel that some effort should be made to provide figures measured under some set of standard conditions, which figures could then be translated into other figures depending on the way the speaker was to be used.

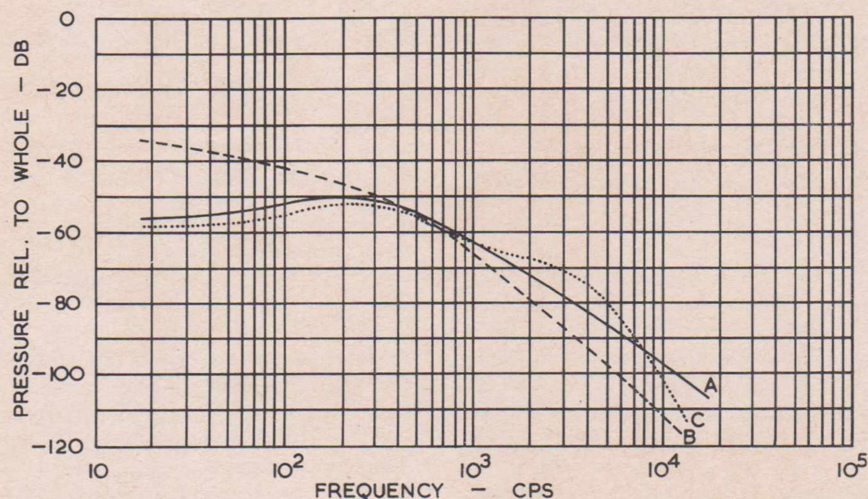
Not all these points are significant in the case of complete speaker and enclosure assemblies, but this is a case where the presentation of more information is greatly simplified, as we know the conditions under which the speaker is being used. However, we are likely to find a bland statement that a unit will handle, say, 20 watts rms. But we peek inside at perhaps a 3-inch tweeter and all reason tells us that whilst the woofer may handle 20 watts, the tweeter certainly will not. What then is the answer? The makers are remarkably reticent about such points, but intelligence suggests that the power rating quoted is applicable in the low or mid-range below the tweeter cutoff.

The makers know just as well as we do that there is in fact much less energy in a typical musical programme in the region over which the tweeter will become operative than there is below the tweeter cutoff point, and this fact is being relied upon in the design of the unit. Then why not say so? Whilst not perhaps of great value to the person who merely buys a speaker and takes it home to listen to it, further data, besides enhancing the standing of the manufacturer, would be of great help to experimenters and others interested in forwarding the art.

Moir[3] has reported that a large orchestra may have frequency components between 15 cps and 20 Kc, with peak amplitudes fairly constant over the range 50 cps to 15 Kc, but wherein the peak amplitudes in the bass region are almost entirely due to the drums and those in the treble region are almost entirely due to the cymbals. Significantly he reports that above about 500 cps, rms amplitudes are approximately inversely proportional to the frequency. Work done by the Bell Telephone Laboratories [4, 5] has provided a great deal of information on this subject, and has been used to prepare Fig. 2, which shows the average pressure spectra for (a) a 75-piece orchestra, (b) a cinema organ and (c) male speech, plotted against frequency.

In justice to all it must be stated that the speaker manufacturers are well aware of the problems involved in determining speaker ratings. A recent article by Mr. K. Russell, Technical Manager of Wharfedale Wireless Works Ltd.,[6] ("Audio", Sept. 1965), dealt with the protection of speakers against overload and enumerated the more significant aspects of the problem. It would not be the right thing to do to paraphrase Mr. Russell's article here, but for all those interested his remarks are

Fig. 2



well worth reading. Incidentally he suggests what appears to be a novel speaker protection circuit intended to protect the speaker against most of the damaging types of input that may occur. The circuit is interesting but is not cheap, and is admitted by the author to be not the complete answer.

It is of course fortunate that this problem does not intrude in practice more often, but there is reason to feel that it may become more acute in the future. We know that quite modest amplifiers with rms ratings of some 10 watts or so are capable of delivering peaks of more than 5 or 6 times that rating. We are entitled to speculate on the performance under similar conditions of modern amplifiers that at least have solid-state power supplies with their improved regulation, and those all-solid-state units which are also capable of sustaining their high power output well over the audio spectrum. For example, speakers today are being subjected to the possibility of inputs at the full rated power in the region below 20 cps, something unheard of a few years ago. It would be most interesting to hear readers' comments on the subject.

Home Building

No, this is not going to deal with methods of building a house, but is in response to a regular flow of letters about the relative merits and demerits of home-building amplifiers and the like. I think one has to be quite honest about this, and I will endeavour so to be. In the first place it must be recognised that in the main, constructional articles are intended for those whose hobby is construction and experiment, and who can therefore be expected to possess at least some facilities for testing and evaluating equipment. Such articles are rarely intended for the person whose sole venture into practical electronics is to be an amplifier to play records, at which point his interest starts and finishes.

It has to be appreciated that it is virtually impossible to design and build a number of anything without faults intruding into at least some of them. In full scale manufacture, one of the big problems is component tolerances. For the home builder, there is not only this type of problem, but he is vulnerable in many other areas as well, including the mercifully rare but intensely annoying dud component. Without at least rudimentary test facilities, he would not even be able, except perhaps by clever deduction, to overcome such a simple contretemps. Home building is fine for those who want the sport, who can get themselves out of the occasional hole, and who recognise and cheerfully accept the fact that nothing man makes is entirely perfect. It can be a stimulating and intensely rewarding effort for those so inclined, but a source of bitter disappointment for others.

As far as costs are concerned, let us face the fact that in most cases, building a unit at home from a published design will often work out very little if anything cheaper than buying a comparable factory made unit. Very often the published designs, because they are for those who may be short of test equipment or may lack other facilities, have to be held down to something that can readily be put together and is essentially reproducible.

Occasionally one finds a situation where home construction can pay large dividends to those who can undertake it, not so much by way of costs but in other ways. For instance, at the time when this magazine started publishing transistorized amplifiers with all their now-proven advantages, not only could one not buy comparable units on the local market, but there were still those prophets of doom who were telling us that transistors would never be any good for high fidelity.

So if you enjoy constructing and have the facilities, do so by all means. But if you have previously built little or no electronic apparatus, and just want an amplifier to sit down and listen to, go out and buy a reputable make from a reputable dealer and take it home and enjoy it.

The Third Channel

Mention has often been made of a third channel for stereo, and has often been coupled with the conversion of an existing mono system to stereo or with a reduction in the size and cost of the speakers of a stereo installation. In the first case, the idea has been to use the large high quality speaker already forming part of the existing mono system either as a "combined bass" speaker to be used with two small low-cost left and right channel speakers which are to be added and which carry only the higher frequency portions of the programme material, or as a wide-band third channel speaker again augmented with two new low-cost speakers. The idea in any event was to save cost and perhaps also size.

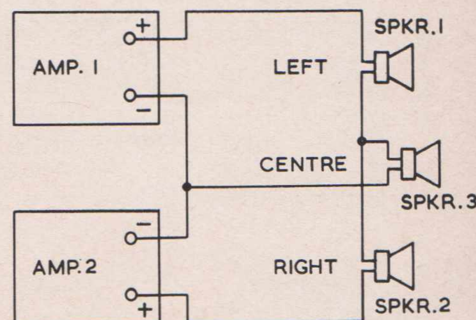


Fig. 3

Some time ago a third channel arrangement was suggested by C. G. McProud, a well-known writer on audio topics, and is shown in Fig. 3. In this arrangement a third speaker is connected into the common return lead from the left hand and right hand speakers. This third speaker receives the sum of the two channels and is therefore a monophonic signal, which is used to "fill in the middle" of the stereo reproduction. It will be clear that this arrangement will tend to decrease the channel separation and will blur the stereo image, it will, however, produce a better distribution across the room.

An improved but more complex system has recently been suggested by Mr. D. Hafler, President of the Dynaco Corporation (USA).[6] The improved system consists of additional modification as shown in Fig. 4, that is, additional to the arrangement shown in Fig. 3. Fig. 4 shows the additional modification inserted between the control unit or units as the case may be and the main amplifier or amplifiers. The alternative is mentioned here in the way of control unit and main amplifier, as this arrangement, like all third channel systems, finds a great deal of its application to cases where an existing mono system is being upgraded to a stereo system, and the likely procedure

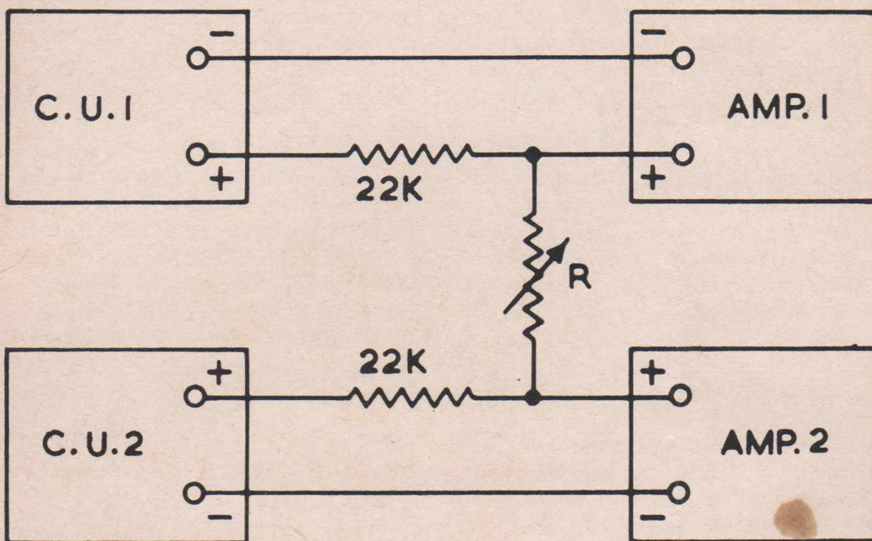


Fig. 4

will be the addition of a second control unit and main amplifier system similar to the already existing one. However, any other arrangement lends itself to the modification except possibly "floating output" systems. Further, the modification could be inserted at any convenient point, depending on circumstances.

The modification consists of the addition of three resistors, one being variable to permit adjustment of cross talks between the two channels. The modification converts the system into a bridge configuration. It this is not immediately clear, then the redrawn complete arrangement of Fig. 5 will serve to show just what is being done. To adjust the system, connect only one side of the stereo pickup to one input, say the left, channel of the control unit(s). Then put on a mono record and adjust resistor R until the minimum output is heard in the right hand speaker. If the conditions are now reversed, minimum output should in turn be heard from the left hand speaker. If this is not so, then it is likely that the balance control needs adjustment, or that the gain, bass, treble and loudness controls are not correctly balanced. All this depends of course on the arrangement of the particular control unit and amplifiers, and the positions of the controls named in the circuit. A likely value for R is 20K.

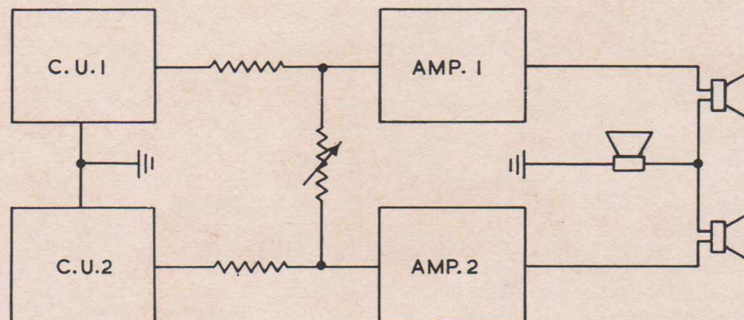
However, having established the required conditions, it will now be found that the right hand channel is played by the right hand speaker and the centre speaker, whilst the left hand channel is being played by the left hand speaker and the centre speaker. The subjective effect is something akin to having extended the speaker system right across the room. For those who like experimenting, this is really something to try on a wet weekend. For those who have a good mono system and are seeking lowest cost ways to upgrade to stereo, this is something worth considering. For the purist, I suppose one must admit that no arrangements of this kind are as good as the classic approach, which also in its purest state requires extensive experimentation and tailoring of the speaker system to suit the room. On the other hand it can produce results acceptable to most, is far cheaper, and is less demanding on one's time.

New NAB Tape Standards

The National Association of Broadcasters Recording and Reproducing Standards Committee set to work in 1961 to update standards as they then existed, and new tape standards were presented in April of this year at the Twelfth Annual Spring Convention of the Audio Engineering Society in Los Angeles. Many sections of the standards laid down would interest the average recordist and experimenter very little, as they relate to such matters as reel dimensions, tape dimensions and similar matters outside his control.

Firstly, the new standards are unique in that they recognise two different operating philosophies, the use of high quality studio equipment and the use of lower quality equipment for portable and location work. Further, 7½ ips is now recognised as the preferred speed, and 15 ips and 3¾ ips as supplementary speeds. There are several other interesting features of the new standards, but this is not the place to attempt an exhaustive evaluation of the new provisions. Those sufficiently interested will doubtless secure their own copies and form their own conclusions.

Fig. 5



New Frontiers in AF Amplifiers

Several interesting things are happening at the moment in the world of audio amplifiers, all of them as one may well imagine in the field of solid-state amplifiers. First I suppose one must recognise the swing to silicon transistors for use in hi fi and entertainment equipment generally, largely made possible by the production of suitable devices at comparatively low prices. This has allowed designers to use silicon units and avail themselves of the advantages arising therefrom, including in many amplifiers a large saving in the size and cost of heat sinks for the output stages.

Since the introduction of the solid-state amplifier into the home, there has been a great deal of discussion and argument about the respective merits and demerits of the different circuit configurations available. But I do not think anyone would doubt that by and large the most popular arrangement has been the series-connected single-ended class B circuit, even though this arrangement generally used a driver transformer and is very susceptible to short circuits across the load terminals. The popularity of the circuit has been amply demonstrated by the large number of commercial manufacturers who have used it, and have spent a great deal of time and trouble trying to provide means to render this type of amplifier short circuit proof. The problem of short circuit protection is covered by another article now being prepared for these pages, so I will not get involved in it at this time. However, in the more recent types of amplifier coming to notice, and in which silicon transistors are now the order of the day, a similar circuit is being used, with the exception that direct coupling to the output stage is being used, doing away with the driver transformer. This generally necessitates feeding the load through a large value of capacitor, as the load connection point is otherwise above ground. This is no disadvantage, particularly when one considers that the new breed of amplifier is proof against open circuit and short circuit loads. There is probably no longer a valid reason for retaining any transformer in the signal circuits of a modern amplifier, except in such specialised cases as automobile radios and the like; certainly not, I would suggest, in hi fi equipment being designed today.

There is another thing first thought of in about 1930, if not before, which is today being heralded as the latest word, and the amplifier to end all amplifiers, at least by some. This is the so-called class D pulse-width-modulated amplifier, also known as the two-state amplifier. This arrangement is based on the use of transistors as switches instead of continuously-controlled amplifiers. It involves the use of a supersonic frequency generator, the output of which consists of or is formed into a pulse train wherein the width of the pulses is modulated by the audio signal. The resulting pulse train, modulated in width responsive to the frequency and amplitude of the audio signal, is then used to switch the output transistors. From the output stage, the energy in the pulse width modulation is recovered in amplified form in the load, using a low-pass filter to reject the supersonic "carrier".

A considerable literature on this type of amplifier has built up during this year, although interest really started in the idea back in the late 40's, practical low cost realisations of the idea becoming possible with the introduction of the transistor. The main claims to fame for this arrangement seem to be based on size and cost, coupled with high output efficiency. It is true that high efficiencies can be achieved, but a little arithmetic will show that some of the wilder claims certainly cannot be met. The basic premise is quite sound, that transistors work well as switches, and dissipate very little power in that service. This is because, with the exception of the very rapid transition from one state to the other, the transistor is either in the "on" condition, passing current but with a very small voltage across it, or in the "off" condition, with full supply voltage across it but passing negligible current.

However, some of the examples that have so far appeared have been very disappointing. I think this arises from lack of reproducibility in the design and/or a failure on the part of the designer to see or appreciate some of the problems involved. Much of the literature that has appeared has been contradictory, a symptom of an imperfect understanding of the problem. An analysis of the literature to date appears to show some workers blithely ignoring requirements that the majority of other workers agree on as rather fundamental. So the present state of the art is a little confused, and a cautious approach would seem to be a good idea, at least for a while. In the meantime, the famous Mr. N. Crowhurst has turned up with what appears to be a class BD amplifier, or is it a class DB? I must confess I am not quite sure, but the basic idea seems to be an amplifier that operates in class B for low signal amplitudes and undergoes a transition to class D at high signal amplitudes. Time will tell.

REFERENCES

1. *"Audio Frequency Standards", "Radiotronics", Oct. 1962.*

2. *"Transistors in Hi Fi", "Radiotronics, Feb. 1963, and "Super-Radiotron 75-watt Hi Fi Stereo System", idem, Aug. 1963.*

3. *"High Quality Sound Reproduction", J. Moir, Chapman and Hall. Second Edition.*

4. *"Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras", L. Sivian, Dunn and White, J. Acoust. Soc. Ameri., Vol. 2 No. 1, Jan. 1931.*

5. *"Speech Power and its Measurement", L. Sivian, Bell System Tech. Jour., Vol. 3 No. 4, Oct. 1929.*

6. *Reported in "Technical Talk", The Gramophone, Aug. 1965.*

MICROCIRCUIT-MICROWATT DESIGN TECHNIQUES FOR NEW INTERNAL MEDICAL SENSORS

Until recently, internal medical sensors have been limited to those transducers that were not only small enough, but also extremely sensitive in response. Now, with modern miniature semiconductor devices and integrated circuits, it is becoming practical to combine a relatively insensitive transducer with some electronic gain inside the body to achieve the needed overall system sensitivity with reduced size. Work on such microcircuit-microwatt sensors is reviewed herein.

Sensors for measuring physiological functions have been with us for many years. The measurements have, in general, been mechanical or acoustical and relied upon the senses of the doctor for their detection and interpretation—for example, the stethoscope. Today, with modern electronics, such basic medical information may be obtained not only with much better fidelity but in many cases directly from the source. Small microphones[1] can be (and are) inserted into arteries and veins, floated into the heart, and used for listening for valve and other defects.

But such devices obviously must be extremely small and very sensitive. As might be expected, doctors anxious to use such sensors often want from designers twice the sensitivity, half the size, or both. In the past, most of the effort has been concentrated on devising transducers with high sensitivity and small size. But with up-to-date transistors and integrated circuits, it is now sometimes possible to use the combination of a small but relatively insensitive transducer with some electronic gain to produce the required sensitivity together with reduced size. Examples of completely electronic transducers only recently available are: sensitive tunnel-diode pressure transducers,[2] thermistors, varactor diodes, and semiconductor strain gauges. By adding thermocouples, variable-reluctance inductors, and temperature-sensitive capacitors to these standards, there are a variety of ways in which the job could be done.

F. L. HATKE and L. E. FLORY*

Astro - Electronics Applied Research Laboratory Astro-Electronics Division, DEP, Princeton, N.J.

Telemetry from within the body

RCA Laboratories' work in this field has concentrated on obtaining physiological information from within the human body without any connecting wires. The pioneering work in this field occurred almost simultaneously in the U.S.[3] and in Europe[4-8] where investigators, spurred on by the invention of the transistor, developed active (battery-powered) telemetering capsules. These units broadcast pressure, temperature, etc. from the alimentary tract. Unfortunately, the experiments were limited to a few days or weeks by the capacity of the battery. Many new and more interesting experiments became apparent if the capsule could be made smaller and its life, inside the body, could be made infinite.

With these additional requirements, a second,[9-11] or passive system, was devised incorporating more complexity on the outside of the body and less on the inside. This system, with but two passive electrical components inside the body, made it possible to reduce the capsule volume approximately 50%.

Passive telemetering system

In the passive system, energy is supplied to the capsule from the outside,

* Mr. Flory is a *Fellow* RCA Laboratories.

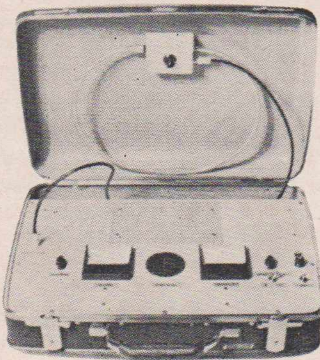


Fig. 1—A completely transistorized passive system including transmitter and receiver.

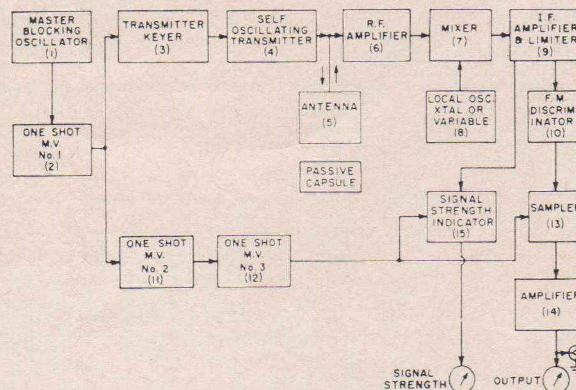


Fig. 2—Transistorized 2-Mc passive capsule system.

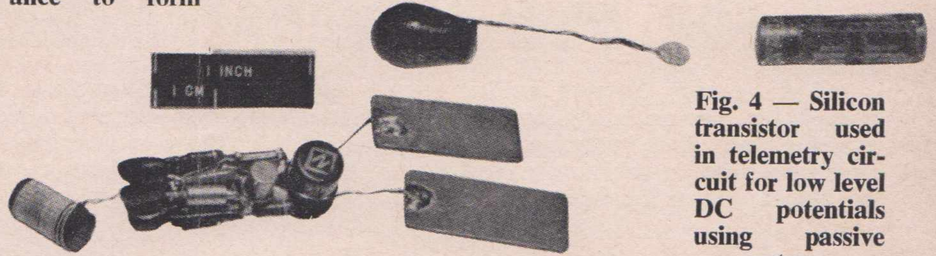


Fig. 3 — Examples of the use of capacitance and inductance to form passive temperature transmitters of high sensitivity.

Fig. 4 — Silicon transistor used in telemetry circuit for low level DC potentials using passive systems.

which energizes the circuit of the capsule, which in turn returns a portion of the energy to the outside equipment, together with the telemetered information. The capsule itself consists of an inductance and a capacitor, each of which may also be the transducer, depending on the required information. The variable inductance has been found useful for the measurement of pressure and a temperature-sensitive capacitor for temperature. In either case, bursts of energy with a frequency at, or near, the resonant frequency of the capsule are supplied from the outside antenna. Some of this energy is absorbed by the capsule during this transmitting time. When the transmitter is turned off, the capsule dissipates this energy at its own resonant frequency. Some fraction of this energy, modulated by measured parameters, is received by the same outside antenna. All that is required is to measure the frequency of this returned energy—this will be a direct measure of the internal physiological phenomena. To facilitate the transfer of energy both into and out of the capsule, the *Q* of its resonant circuit should be as high as practical. Even with optimum conditions, the ratio of the transmitter to received energy may be in the order of 100 db. Considerable effort has been expended to keep the capsule *Q* high and, at the same time, reduce the size to an even smaller volume.

Recently, a completely transistorized passive system based on earlier work has been designed (Fig. 1). This passive telemetering system is divided into three essentially separate sections; the transmitter; the capsule; and the receiver. All three are loosely tied together by the antenna system and the timing system. Fig. 2 shows the overall system operation.

The whole system cycle is divided into three almost-equal time intervals: transmit time; delay time; and receive time. Their repetition rate and duration are controlled by the master blocking oscillator and one-shot multivibrators No's. *1*, *2*, and *3*. The transmitter is activated by multivibrator *1* and energizes the capsule during its *on* time. Thereafter, multivibrator *2* determines the length of time before an uncontaminated signal is received from the capsule, and multivibrator *3* controls the time at which the receiver output is measured to determine the capsule's natural resonant frequency.

The transmitter is a self-oscillating power transistor delivering about 300 volts peak-to-peak across the antenna. The length of time the transmitter is energized, as determined by multivibrator *1*, controls the *on-off* keyer in the emitter of an oscillating transmitter.

The capsule, after absorbing some of the energy from the transmitter, returns a small portion of it to the antenna at a frequency determined by the information to be telemetered.

This combination of signals is fed into the receiver, which amplifies and limits over the 100-db range between the *transmit* and the *receive* level. A sample of the combined signal is taken at the last limiter during the receiving time. This sample is rectified and displayed on a front panel meter indicating the amount of energy the capsule is returning to the receiver.

The output of the limiter is also applied to a conventional frequency discriminator. The output of the discriminator is amplified and passed through an emitter follower to reduce the driving impedance. At this point it is sampled during the receiving interval. To make the output signal continuous instead of pulsed, the pulse amplitude is stored during the *transmit* and *delay* times and corrected to the new value when the next *receive* time interval occurs.

This results in the system measuring the natural resonant frequency of the passive capsule, being independent of both transmitter frequency and, within operating limits, distance from the antenna.

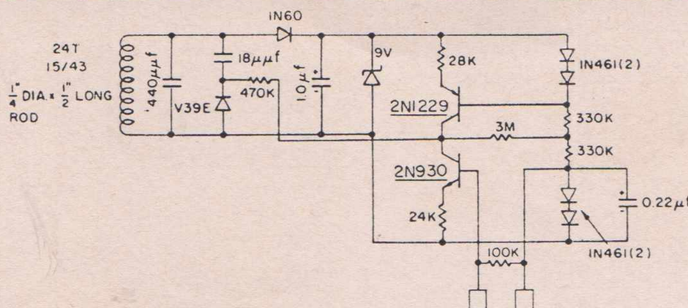


Fig. 5—Passive biological potential capsule.

Pressure and temperature sensors

It became evident early in these RCA Laboratories investigations that measurements of temperature and pressure might be made using either the inductance or the capacity as the transducer. In the case of pressure some nominal mechanical power is available to move the variable reluctance transducer and cause a chance of inductance in the tuned circuit. For temperature conversion, a ceramic condenser is available whose capacity is a measure of its temperature. This, when combined with a suitable inductance, in a resonant circuit, forms a passive temperature trans-

mitter of high sensitivity. An example of each of these types is shown in Fig. 3. These temperature sensors are now in use in the study of the ovarian function.[12,13]

Voltage sensor

In designing a potential-measuring capsule with sufficient deviation in the returned signal to make measurements in the 1-mv range, some sort of electrical gain must be acquired inside the capsule. To do this, some of the energy transmitted into the capsule must be converted into a source of direct current. In addition, the voltage must be regulated so as to keep the amplifier stable with respect to the coupling between capsule and transmitter. Also, the amplifier must not consume very much power because every drain on the tuned circuit is reflected in a reduction in circuit *Q*, accompanied by a loss in operating distance of the capsule from the transmitter.

Recently, junction transistors became available which operate at extremely low collector currents and still exhibit high current gains. These transistors are of necessity made of silicon to keep leakage currents low compared to the signal and bias currents. Fig. 4 shows such a transistor incorporated into a circuit capable of telemetering low-level DC potentials using the passive system. Fig. 5 is a schematic diagram of the potential transducer.

Operation of the circuit is as follows. During transmit time, energy is coupled from the transmitter into the inductance. This inductance is roughly resonated with the 440-pf condenser and modulated in its resonance by the 18-pf and the V39E varactor diode. A small part of absorbed energy is rectified by the 1N60 diode and regulated and filtered by the 1.0μf capacitor and 9-volt zener diode. Transistors 2N930 and 2N1229 form the DC voltage amplifier, the 2N930 being the active amplifier and the 2N1229 forming a synthetic load resistance. Proper biasing is afforded by the 1N461 diodes and the emitter resistors. The amplifier provides a gain of about 200. The output of the amplifier modulates the V39E varactor diode. This assortment of semiconductors, resistors, capacitors, and one inductance with no serious attempt at miniaturization fits into a volume of about 0.3 cubic inch. It is obvious that integrated circuit techniques could reduce this volume by a large factor.

A simpler circuit design is possible if advantage is taken of the time sequence of the transmitter-receiver. Fig. 6 gives one such capsule design. In this system, energy is absorbed during transmit time and charges the storage capacitor *C* up to the full zener voltage. Since there is a time delay between transmit and receive time, the potential across *C* will decay to a value determined by the current load of the field effect transistor.

The potential on the capacitor is also the bias voltage on the varactor diode so that the input to the field effect transistor effectively modulates the bias on the varactor and, therefore, the resonant frequency of the ringing circuit. This design reduces the number of active and inactive elements in the capsule, thus enabling an additional reduction in size.

The sensitivity of this sensor is determined by the available gain in the field effect transistor at the low voltage (= approx. 5 volts) available across the capacitor. Developments in this type of transistor are expected to improve this parameter. Meanwhile, it is also practicable by microcircuitry techniques to provide a field effect transistor directly coupled to a junction transistor in the same assembly at very little expense in power. A typical input-output curve is seen in Fig. 7.

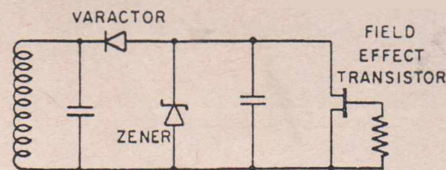


Fig. 6—Simplified passive biological potential capsule.

New technology

As can be seen, the newly emerging microcircuit-microwatt technology is beginning to give the designer of implantable medical sensors the ability to use some electronic gain inside the body. In the past, only transducers sensitive enough to produce proper system deviation directly could be used. Since most electronic transducers convert information to electrical potential, this type of capsule appears to have the widest range of application. Some examples include: direct monitoring of the electrocardiogram; monotoring other internal functional potentials; measurement of internal pH using glass electrodes; in addition to an alternative way of measuring temperature (thermistors) and pressure (strain gauges).

Each of these devices is presently being used in an experimental manner in an effort to learn more about the functions of the human body. It may not be too long before these devices will be used more widely as new tools for medical diagnosis.

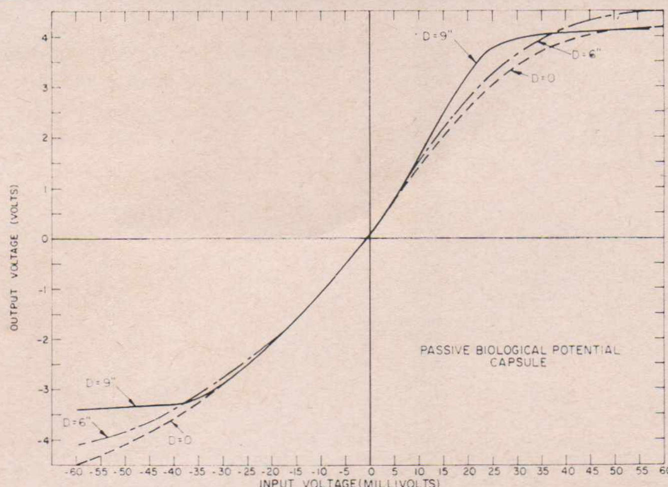


Fig. 7 — Input vs. output at various distances from antenna.

Bibliography

1. Warniek, A., and Drake, E. H., M.D., "A New Intracardiac Pressure Measuring System," *1958 IRE Convention Record,* Pg. 68.

2. Sikorski, M. E., "Sensitive Tunnel-Diode Pressure Transducer," TA 7.5, International Solid-State Conference 1962.

3. Farrar, J. T., Zworykin, V. K., and Baum, J., (1957), "Pressure-Sensitive Telemetering Capsule for Study of Gastrointestinal Motility," *Science,* 126, 975-976.

4. Mackay, R. S., and Jacobson, B. (1957) "Endoradiosonde," London 179, *Nature,* 1239-1240.

5. Jacobson, B., and Mackay, R. S. (1957), "A pH-Endoradiosonde," *Lancet,* 272, 1224-1225.

6. Von Ardenne, M., and Sprung H. B., "Uber Versuche mit Linem Verschluckbaren Intestinakender," *Die Naturwissenschaften,* 45, (7), 154-155.

7. Nöller, H. G., "Die Endoradiosondentechnik und Ihre Bedeutung für die Inner Medizin," *Verhandl, deut. Ges. inn. Med.* 65, 727, 730.

8. Jacobson, B., "Endoradiosonde Techniques — A Survey," *Med. Electron. Biol. Engng.* Vol. 1, pp. 165-180. Pergamon Press, 1963.

9. Haynes, H. E., and Witchey, A. L., "The Pill That Talks," RCA.

10. Zworykin, V. K., Farrar, J. T., Hatke, F. L., etc., "The Measurement of Internal Physiological Phenomena Using Passive-Type Telemetering Capsules," *1961 IRE Convention Record,* Part 9.

11. Farrar, J. T., Zworykin, V. K., etc., "Telemetering of Intraenteric Pressure in Man by an Externally Energized Wireless Capsule," *Science,* Vol. 131, No. 3416, P. 1814, June 17, 1960.

12. Dordick, H. S., Balin, H., Israel, S. L., and Hatke, F. L., "The Determination of Ovarian Function by Means of Permanently Implanted Endoradiosondes," *IFME 4th Conference Digest, July* 1961.

13. Balin, H., Busser, J., Hatke, F., Fromm, E., Wan, L., Israel, S. L., "Radio Telemetry System for the Study of Ovarian Physiology," *Obstetrics and Gynecology,* Vol. 24, No. 2, August 1964.

Radiotronics is published quarterly by the Wireless Press for Amalgamated Wireless Valve Co. Pty. Ltd.

This publication is available at a cost of 50c per copy from the Sales Dept., Amalgamated Wireless Valve Co. Pty. Ltd., Private Mail Bag, Ermington, N.S.W.

Copyright. All rights reserved. This magazine, or any part thereof, may not be reproduced in any form without the prior permission of the publishers.

Devices and arrangements shown or described herein may embody patents. Information is furnished without responsibility for its use and without prejudice to patent rights.

Information published herein concerning new releases is intended for information only, and present or future Australian availability is not implied.