A touch-sensitive X-Y position encoding overlay for computer input

Hlady, A. M.

Report (National Research Council of Canada. Radio and Electrical Engineering Division : ERB), 1968-12

ERB-801

UNCLASSIFIED

NATIONAL RESEARCH COUNCIL OF CANADA

RADIO AND ELECTRICAL ENGINEERING DIVISION

A TOUCH-SENSITIVE X-Y POSITION ENCODING OVERLAY FOR COMPUTER INPUT

- A. M. HLADY -

OTTAWA

DECEMBER 1968

ABSTRACT

Pointing remains man's most natural method of indicating selection. For this reason, a touch-sensitive $x-y$ overlay offers advantages over other input devices in many digital computer systems handling graphic information. An echo-ranging technique using high-frequency, elastic surface waves on a glass plate has been used as the basis for such a device.

CONTENTS

FIGURES

TABLES

LIST OF SYMBOLS

σ_{ij}	stress in the i direction acting in the j plane
λ	Lamé elastic constant
μ	Lamé elastic constant
u_i	displacement in the i direction
ρ	material density
ϕ	scalar elastic potential
ψ	vector elastic potential
c_L	longitudinal wave phase velocity
c_T	transverse wave phase velocity
c_R	Rayleigh wave phase velocity
k_L	longitudinal wave number
k_T	transverse wave number
k_R	Rayleigh wave number
ω	radian frequency
q	longitudinal wave depth decay factor
s	transverse wave depth decay factor
α	c_T/c_L
η_R	c_R/c_T
ν	Poisson's ratio
λ_R	Rayleigh wavelength
α_n	angle of longitudinal wave propagation in material n.

β_n	angle of transverse wave propagation in material n.
ℓ	radiator length dimension in direction of propagation.
w	radiator width dimension normal to direction of propagation.
k	electro-mechanical coupling coefficient
d	piezoelectric charge coefficient
g	piezoelectric voltage coefficient
ϵ_0	absolute dielectric constant of free space
E	Young's modulus of elasticity
Z	characteristic mechanical impedance
β	bearing angle from the radiator normal
K	relative dielectric constant

A TOUCH-SENSITIVE X–Y POSITION ENCODING OVERLAY FOR COMPUTER INPUT

– A.M. Hlady –

I Graphic Input Devices

The introduction of time sharing techniques for digital computers has widened the range of possible computer applications by providing an on-line computing capability to a number of users simultaneously. Successful application in fields without a numerical basis requires that the computer appear to the user as a processor of information in a form that is relevant to his particular field. The input–output equipment associated with the computer must, therefore, act as an information transformer, presenting digital information to the computer, and words, pictures, or sounds to the user. Furthermore, this transformation must take place in both directions if the system is to operate in an interactive, or conversational, mode.

More specifically, many fields require a machine capable of handling graphic information. In this case, the interface between the computer and the user consists of a graphic display unit and an associated graphic input device. At the present time, most computer controlled display units use a cathode-ray tube to present alphanumeric and pictorial information to the user.

The graphic input device in this type of system will, in general, be required to perform two basic functions; item selection, and position selection. The fundamental difference between these two functions is the direction of information flow around the man–machine loop.

Item selection is the selection by the user of a subset out of a set of items. The items must be displayed by the computer before the selection is made, implying information flow from the computer to the user and then back to the computer. An item could be a single point, a word, a line, etc.

Position selection is the selection of a particular location within a bounded region. The selection is conveyed to the computer and can then be displayed to provide visual feedback to the user. Entering drawings or handwriting into a computer involves the position selection function because a line can be considered as a continuous sequence of such selections.

One characteristic of all existing graphic input devices is that they are inherently capable of performing only one of these functions. The other has to be implemented indirectly with an additional cost in hardware and/or software. The first input device developed for use with a CRT display unit was the light pen [1], introduced almost a decade ago. It is currently the most widely used of the several input devices commercially available. Functionally, a light pen is a light sensing device, and physically, it consists of

a photodiode or a fiber optics bundle connected to a photomultiplier tube. It is activated by the light from the portion of a displayed image in its field of view. The light pen is inherently an item selecting device because information must flow from the computer to the user before the pen is activated. The item selected is the item being displayed at the time the pen signal occurs. The other function, position selection, is accomplished indirectly by using a tracking raster or pattern to locate the pen position.

The beam pen [2] is functionally similar to the light pen, but it eliminates the delays associated with phosphor excitation and photo-detector response by sensing the CRT beam current.

The position selecting devices that have been developed up to this time fall into two general classes. With one type, selections are indicated by remotely controlling the position of a cursor, or pointer, displayed on the screen. The controlling element is usually a lever or ball, commonly referred to as a joy-stick or track-ball, which is capable of being moved in two axes simultaneously. This is mechanically linked to a set of potentiometers and the resulting analog voltages are proportional to the selected position.

With the second type of position indicating device, selections are made with a hand held stylus, usually on an associated plate or tablet. The RAND Tablet [3] is probably the best known of this type of input device. It consists of a closely spaced grid of printed conductors which are energized by coded pulse trains. A stylus senses the pulses in the nearest conductors by capacitive coupling. The pulses are subsequently decoded to determine the stylus position with a resolution of approximately 0.01 inch.

Another approach uses a transparent, electrically conducting coating as the tablet surface [4]. A linear voltage gradient is generated across the surface in each axis alternately. The $x-y$ coordinates of the stylus location are proportional to the voltages measured at the point of contact.

Another device that has been reported [5], makes use of pulses of ultrasonic surface waves generated at the edges of a glass plate. A piezoelectric transducer in the stylus acts as a sensor. The location is determined by timing the propagation delay from the source to the stylus in each axis.

The Lincoln Wand [6] which offers a three dimensional input capability, also uses the propagation delay of ultrasonic waves. However, in this case, the waves propagate through the air. By timing the delays from four sequentially energized sources to a sensor in the stylus, the stylus coordinates with respect to x,y, and z axes can be calculated.

The Sylvania Data Tablet [7] has a transparent, electrically conducting layer on which an RF electric field is generated. Signals are capacitively coupled into the stylus, and by

using phase detection techniques, a resolution of 0.003 inch is obtained. Signals for the x and y axes are distinguished by using separate carrier frequencies.

A device using magnetic coupling has also been developed [8]. The stylus, which contains a pulsed source of magnetic flux, is used in conjunction with a tablet having layers of imbedded wires. The conductors are arranged in such a way that the voltage pulses induced in the conductors nearest the stylus produce a parallel binary output corresponding to the pen position.

Position selecting devices such as the ones just described can be used for item selection if a hardware or software comparison is done between the indicated position and the positions of the items being displayed. A decision algorithm may be required to distinguish the correct item from its neighbours.

Various factors besides resolution must be considered in choosing a graphic input device for a particular application. If the tablet is transparent, it can be used as an overlay on a computer-controlled display screen. The display can therefore be used to provide the user with direct visual feedback of the stylus location. This is a desirable feature from a human engineering point of view.

Devices that rely on electrostatic or magnetic coupling between the tablet and the stylus normally do not require the two to be in actual contact. Therefore, a non-conducting material such as paper or film can be placed on the surface without affecting the operation. This is very useful for entering maps or drawings into the graphic system because the figures can simply be traced out from a hard copy.

All of the devices just described require an active stylus; that is, the sensor, or in some cases, the radiator, is located within the stylus. This means that an electrical cable must connect the stylus to the system. In addition, the stylus must be large enough to accommodate these components and perhaps some associated circuitry such as a pre-amplifier. Because of the bulkiness, weight, and connecting cable, many active styli are difficult to use with any sort of dexterity. A passive stylus resembling, say, an ordinary pencil, would facilitate graphic input operations.

Furthermore, there are many applications of computer graphics in which it would be convenient and advantageous to be able to use a finger for item selection or coarse position selection input functions. For example, the words or phrases displayed for selection in an information retrieval system could be in a format suitable for this type of input technique. Ease of operation and speed are obvious advantages of selection by pointing. In addition, the fact that pointing is man's most natural method for indicating selection means that a touch-activated device would create a minimum amount of distraction for the user. This is important if, for example, the user is a young child communicating with a computer-assisted instruction system. For applications such as this,

the input device could also be used, in conjunction with the display unit, as a flexible, touch-activated keyboard.

A touch-sensitive device has been reported in the literature [9]. It consists of a number of wires brought to the front surface of a CRT faceplate. Each wire forms an arm of an ac bridge circuit, the bridge being unbalanced by body capacitance when a user touches one of the wires.

The graphic device described in this report was developed with two primary objectives in mind:

(*a*) It was to be capable of being used with a passive stylus, including a finger.

(*b*) It was to be used as a display overlay.

The first objective led to the use of a Rayleigh wave echo-ranging technique. This approach can be used with a transparent glass plate, which also fulfills the second requirement.

Rayleigh waves are elastic stress waves that propagate along the free surface of a solid. The theoretical study of elastic surface waves goes back to 1885 when their existence was predicted by Lord Rayleigh. The important characteristics of Rayleigh waves are developed in Appendix I. In summary they are:

(*a*) The phase velocity is independent of frequency, and is slightly less than the velocity of a shear wave in the same material.

(*b*) The majority of the wave energy is found within one wavelength of the surface.

(*c*) The particle displacements are elliptical.

Echo ranging with ultrasonic Rayleigh waves has already been successfully applied in the field of flaw detection for structural materials [10]. A graphic input device previously described [5] uses ultrasonic Rayleigh waves although not in an echo-ranging mode.

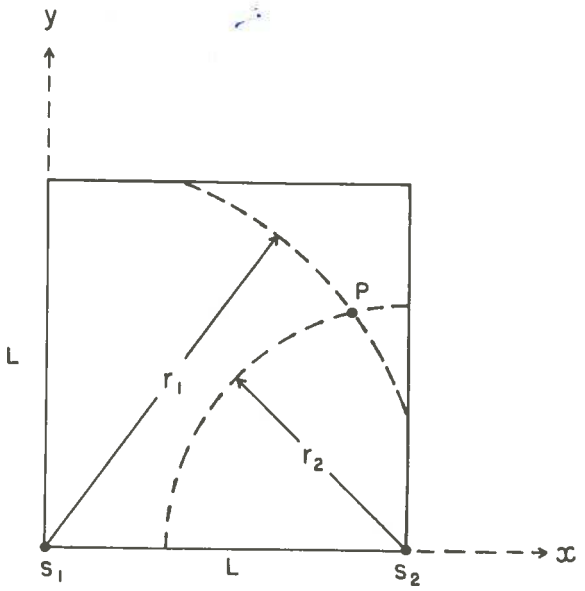
Basically, the system described in this report consists of a pulse-modulated Rayleigh wave radiator and a sensor to detect the waves reflected by any object making contact with the glass surface. The radiator and sensor are physically the same piezoelectric transducer which is electrically switched between the driving circuitry and the receiving circuitry. The distance from the radiator/sensor to the target is proportional to the time between the radiator pulse and the detected echo.

II System Design

All systems using echo ranging for target location have a similar set of system parameters. The basic difference is in the constraints that must be applied to these parameters. Considerable effort has gone into refining echo-ranging techniques for radar and sonar. Many of the resulting refinements, such as signal correlation techniques, could, in principle, be applied to the system under consideration. However, in most cases, the additional cost and complexity makes them impractical for this application.

The major parameters taken into consideration in the system design are the method of stylus location, the range or size of active surface, the carrier frequency, the bandwidth, and the sampling rate. These are by no means independent of each other, and any design represents a compromise or trade-off in performance characteristics.

The stylus location in two dimensions can be determined by measuring its distance from two fixed points, or its normal distance from two fixed lines. Considering a square plane surface in the first quadrant of an $x-y$ coordinate system, the first method will provide a unique solution for any location if the reference points are chosen at two adjacent corners, as indicated in Fig. 1.



$$x_P = \frac{L^2 + r_1^2 - r_2^2}{2L}$$

$$y_P = \sqrt{r_1^2 - x_P^2}$$

Figure 1 Target location by point reference

The radiators required to implement this method must have an effective beamwidth of 90°. A radiation characteristic of this type could be obtained by using a source with dimensions comparable to the wavelength. At frequencies in the order of 1–10 MHz, the wavelengths are in the millimeter region, and for a source of this size, the resultant

surface-wave energy density would be quite small. This is because of the practical limitations on driving-power densities for piezoelectric materials, and also because of the large beam spread.

Unfortunately, trying to shape the surface-wave beam by focusing results in a significant power loss due to partial conversion into spatial waves. An alternative source could consist of an array of larger, narrow-beam radiators combined to have an equivalent beamwidth of 90°.

Additional computation is required to convert the measurements obtained with this method into $x-y$ coordinates. Because most graphic systems are based on $x-y$ coordinates, this represents a disadvantage if computer time is at a premium.

Another consideration is that only a selected portion of the entire surface can be used in order to avoid the problem of 'edge clutter'; that is, edge reflections obscuring valid signals.

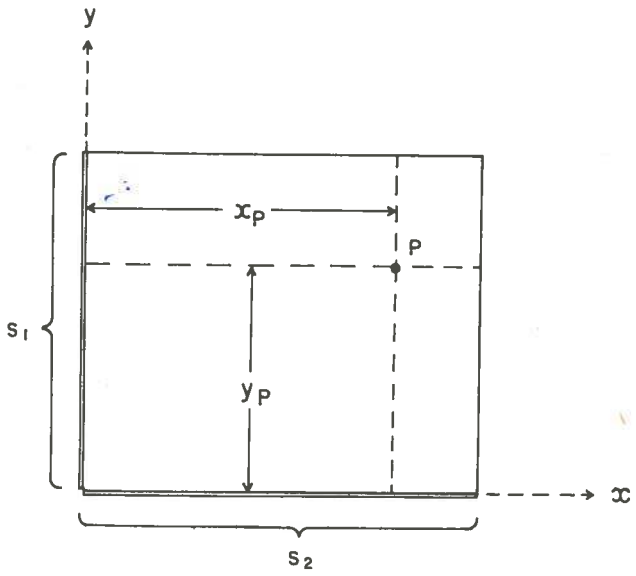


Figure 2 Target location by line reference

The line reference method avoids many of these problems. To implement the method, shown in Fig. 2, each source must provide a wavefront that is uniform and linear for the entire width of the surface. However, the $x-y$ coordinates are obtained directly with this method. In addition, the edge clutter problem does not occur. The line reference method of target location was selected.

Because of the practical problems involved in obtaining the desired radiator characteristics, it was found necessary to use two arrays per axis, one at each side of the surface. A detailed account of the radiator development is given in the next section.

The choice of plate material was limited by the transparency requirement. Fused quartz has the lowest attenuation coefficient for ultrasonic waves, but a sheet of the size required here would be totally impractical. Table I shows the surface-wave attenuation of several kinds of commonly available sheet glass. The results were corrected for beam spreading so that the attenuation listed is due only to scattering and absorption. Plate glass was selected.

TABLE I

Surface Wave Characteristics at 8 MHz

	Velocity (meters/sec.)	Attenuation Coefficient (nepers/cm.)
Plate glass	3170	0.159
Armor plate tempered glass	3140	0.168
Double diamond window glass	3120	0.163

All of the glass tested was found to have several surface flaws per square foot that were large enough to produce noticeable echoes. These flaws consisted of pits and bubbles created during the manufacturing process and scratches caused by subsequent handling. Fortunately, many of the flaws were shallow enough to be eliminated by local hand grinding and polishing. Grinding with a fine compound (9.5 micron) and polishing with cerium oxide produced satisfactory results.

There are several factors involved in the choice of carrier frequency. For a given fractional bandwidth, the system resolution is directly proportional to the frequency. Secondly, surface-wave attenuation is also almost directly proportional to frequency in the lower megahertz range. In addition, for a given radiator size, the beamwidth is approximately inversely proportional to frequency. A high frequency requires thinner piezoelectric transducers, which implies more fragile transducers and a higher shunt capacitance. With solid-state sources, a high frequency usually means more expensive drive power. By taking these factors into consideration, a frequency of 8 MHz was chosen for the initial model. At this frequency, the Rayleigh wavelength, λ_R, on glass is 0.4 mm.

Resolution, in this device, refers to the minimum stylus movement that the system can resolve. Assuming a stable timing source and no drive pulse jitter, the maximum uncertainty in the propagation delay time of the pulse modulated signal is equal to the rise-time of the signal envelope. This, then, is the minimum resolvable change in delay time.

The Rayleigh wave radiators used in the initial model were limited to a minimum Q of about five in order to maintain a good gain–bandwidth product for the electro-mechanical transformation. This is a constraint on the over-all system bandwidth and the electronic circuitry was designed to have an electrical bandwidth matching that of the transducers.

For a pulse modulated signal, the envelope risetime can be expressed as:

$$\text{Risetime (10–90\%)} = 0.7/-3 \text{ db bandwidth}$$

For a carrier frequency of 8.0 MHz and a Rayleigh wave velocity of 3170 m/sec on glass, the system has an inherent resolution greater than 0.03 inch. As outlined in Section VI, the actual usable resolution may be lower than this because of insufficient or irregular sensitivity across the surface.

A usable surface size of 10 X 10 inches was chosen as being adequate for many graphic input applications. Because a space of 2 inches was required between the radiators and the usable portion of the surface, the maximum range became approximately 12 inches. This represents a maximum two-way propagation time of about 200 μsec. Using four separate scans for each sample, this means that the maximum possible sampling rate is 1250 Hz.

Assuming stylus speeds of up to 18 inches/sec for position selection, the sampling rate compatible with 0.03 inch resolution is about 600 Hz. This is only half of the maximum available rate. A considerably lower sampling rate would be adequate for performing item selection.

Several factors affect the positional accuracy of the system. Inaccurate results could be caused by variations in the Rayleigh wave velocity because of changes in the plate temperature or because of material inhomogeneity. Special types of glass have been developed for ultrasonic delay line applications that have temperature coefficients of delay as low as 1 ppm/°C. In this case, ordinary plate glass with a coefficient of about 60 ppm/°C was considered to be adequate. A temperature change of 70°F would be required to cause an error equivalent to the inherent resolution of 0.03 inch.

Another factor is the accuracy of the timing circuits which digitize the propagation delay times. The critical parameter here is the stability of the counter oscillator and the count starting delay. The contribution from the timing circuits remains below 0.5% with conventional circuitry.

For the array configuration being used, the maximum error due to the nonlinearity of the wavefront is about 0.3% at the closest range, and is considerably less than that over most of the surface.

Taking all these factors into consideration, the over-all positional accuracy is held to better than 1% with conventional techniques, and could be improved, if required, by careful design.

III Rayleigh Wave Radiators

Radiator Design

In general, surface waves can be generated by any periodic surface perturbations with the desired frequency. At ultrasonic frequencies in the lower megahertz range, one of the most efficient and convenient methods of producing surface waves is by mode conversion of a longitudinal wave.

Consider an interface between two solid materials with a longitudinal wave, A, incident on the boundary at an angle α_1 to the normal. The normal and tangential components of the stress and displacement must be continuous across the interface. These boundary conditions are satisfied in the general case if both longitudinal and transverse reflected waves, A_1 and B_1, and longitudinal and transverse transmitted waves, A_2 and B_2, are produced, as shown in Fig. 3. This is the phenomenon of mode conversion.

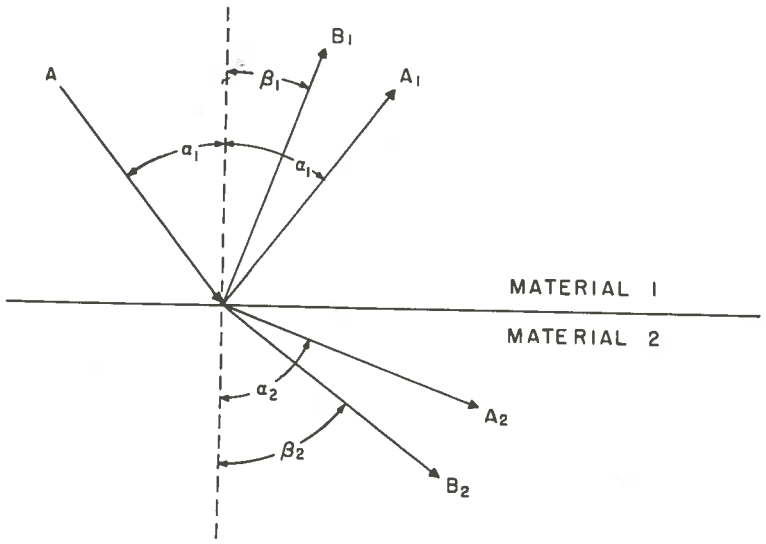


Figure 3 Mode conversion

By writing the equations for the stress and displacements for these waves, and applying the boundary conditions, it can be shown that elastic stress waves obey Snell's law of refraction. That is,

$$\frac{\sin\alpha_1}{c_{L_1}} = \frac{\sin\alpha_2}{c_{L_2}} = \frac{\sin\beta_2}{c_{T_2}}$$

where c_{L_1} is the longitudinal phase velocity in material 1, and c_{T_2} is the transverse phase velocity in material 2.

Consider the case where materials 1 and 2 are chosen such that

$$c_{L_1} < c_{T_2} < c_{L_2} \ .$$

Then since

$$\sin\beta_2 = \frac{c_{T_2}}{c_{L_1}} \sin\alpha_1 \ ,$$

the angle of incidence, α_1, can be chosen so that $\sin\beta_2 > 1$, which also implies that $\sin\alpha_2 > 1$. A physical interpretation of this is that total internal reflection occurs. However, in order that the boundary conditions remain satisfied at the interface, inhomogeneous longitudinal and transverse waves are formed on the surface of material 2. An inhomogeneous plane wave is one whose amplitude decays with distance in a direction perpendicular to its direction of propagation.

Rayleigh waves can be represented as a combination of two inhomogeneous waves, one longitudinal, and the other transverse. Although Rayleigh waves exist only on a free surface, the surface waves generated at this interface have similar characteristics. However, as they propagate along the interface they excite spatial waves in material 1 which carry away a portion of their energy.

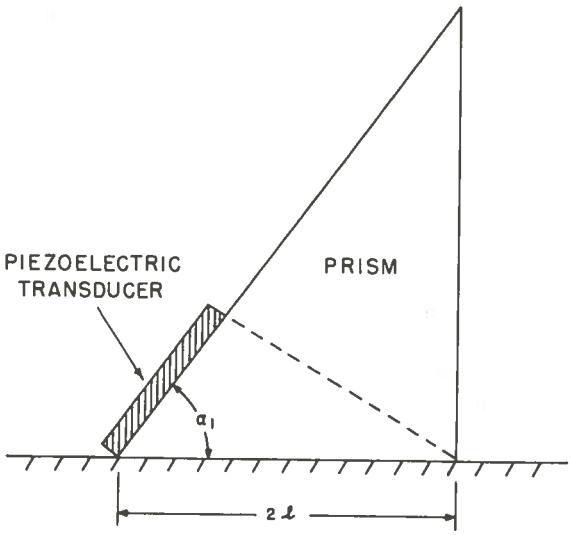


Figure 4 Rayleigh wave radiator

In a practical implementation of mode conversion for Rayleigh wave generation, a prism with a thickness-mode piezoelectric transducer (Fig. 4) is used as the source of longitudinal waves. The perturbations generated at the interface between the prism and the lower material propagate as Rayleigh waves on the free portion of the surface.

By approximating the waves in the interface region with a system of normal and tangential stresses on a free surface, Viktorov [15] has obtained expressions for the normal and tangential surface displacements in front of the radiator. The portions of both expressions that represent a Rayleigh wave are of the form:

$$R = C\frac{\sin\left[\left(k_{L_1}\sin\alpha_1 - k_{R_2}\right)\ell\right]}{\left(k_{L_1}\sin\alpha_1 - k_{R_2}\right)}$$

where C is a complex constant which depends on the elastic properties of material 2 and the amplitude of the stresses at the interface;

k_{L_1} is the longitudinal wave number in the prism;

k_{R_2} is the Rayleigh wave number in material 2;

2ℓ is the length of the interface illuminated by the incident beam;

α_1 is the prism angle.

This can be rewritten as a $\frac{\sin F}{F}$ expression, which has a maximum at $F = 0$;

i.e., at $\sin\alpha_1 = \frac{k_{R_2}}{k_{L_1}} = \frac{c_{L_1}}{c_{R_2}}$

where c_{R_2} is the Rayleigh wave phase velocity in material 2.

This means that for a given excitation, the largest Rayleigh waves will be radiated when the prism angle, α_1, satisfies the above equation. At this angle, the spatial period of the surface perturbations corresponds to the wavelength of the resultant Rayleigh waves, λ_R. For this optimum angle to be real, the prism material must be chosen so that

$$c_{R_2} > c_{L_1} .$$

The expression for R indicates that its maximum value is linearly dependent on the dimension ℓ, the interface length dimension in the direction of propagation. This comes about from the approximations used in the derivation of the expression. In reality, surface perturbations generated at the prism interface are attenuated by propagation along the interface. This attenuation is largely due to energy being carried away by spatial waves that are excited in the prism. The result is that perturbations occurring near the front of the prism interface contribute the majority of the radiated energy.

There are several reasons for choosing the prism type of Rayleigh wave radiator for this application in preference to other types described in the literature. Most other radiators depend upon surface disturbances created by a stress wave impinging on the surface at normal incidence. For example, one proposed method [15] consists of a metal plate with a thickness mode vibrator bonded to its upper surface. Its lower

surface is shaped to have a corrugated profile with alternate projections and slots each of width $\lambda_R/2$. When this assembly is placed on a flat surface, it produces normal perturbations with a spatial period λ_R. This method is reported to be more efficient than the prism method. In addition, there is no limitation on the materials that would correspond to the velocity restriction for the prism. However, in spite of these advantages, there are certain characteristics inherent in surface wave radiators using normal incidence that make them less desirable for use in a surface echo-ranging application.

One of these characteristics is the level of spurious radiation in other modes. Normal-incidence techniques involve the transmission of a relatively large portion of the incident energy into the lower material in the form of longitudinal and/or transverse spatial waves. These spatial waves propagate through the lower material and may experience multiple reflections within that material. They can cause problems as spurious signals in an echo-ranging system.

On the other hand, it has been shown that the prism method involves total internal reflection of the incident energy within the prism. This implies that, in theory no energy is transmitted into the lower material as spatial waves, which is in contrast to the effect using normal-incidence techniques. It would seem, however, that the reflected energy remaining in the prism could also cause spurious signals. In practice, one usually has more design freedom with the prism than with the material on which the surface waves are to be generated. By shaping the prism to trap the reflected energy and by coating its free surfaces with an absorbent material, the spurious signal level can be kept down.

Another characteristic associated with spurious signals is the directionality of the radiator. Because of their symmetry, normal-incidence radiators produce Rayleigh waves of equal amplitude in both directions along their axis of propagation. Prism-type radiators are nearly unidirectional; the amplitude of waves propagating in the forward direction from the prism being $2k_R\ell$ times larger than the waves propagating in the opposite direction [15]. This is approximately 50 db for the radiators being used in this application.

A third characteristic of normal-incidence radiators is apparent in applications in which the carrier frequency is pulse modulated, as in echo-ranging systems. This is the risetime of the wave packet, which affects the system resolution. In a normal-incidence radiator, surface disturbances are created at all points of the interface simultaneously, and begin propagating along the interface away from their point of origin at a velocity of c_R. Therefore, for an interface length of 2ℓ, a time of $2\ell/c_R$ is required before all the disturbances created at a given instant arrive at the free edge of the interface. The effect is to increase the risetime of the Rayleigh wave packet by $2\ell/c_R$.

For the case of prism type radiators, the waves created by a pulsed excitation of the piezoelectric transducer do not create disturbances over the entire interface simultaneously. This means that the envelope of the Rayleigh wave packet retains the shape of the excitation voltage envelope.

Among the various piezoelectric materials available at 8 MHz, piezo-ceramics are good sources of ultrasonic energy because of their relatively high electromechanical coupling coefficient, k. The square of this coefficient represents the ratio of electrical energy converted into mechanical energy to the total electrical energy input. It also represents the converse ratio, in which mechanical energy is being converted into electrical. The coefficient, k, therefore, provides an indication of the relative merit of a piezoelectric material as a radiator and a sensor.

In order to compare the relative merits of a material for radiator or sensor applications separately, the piezoelectric charge and voltage coefficients, d and g, are often specified. The d coefficient is a measure of the strain developed in a transducer for a given field applied across it, while g indicates the open-circuit field developed for a given stress. These two coefficients are related to each other, and to the coupling coefficient by the following expressions:

$$d = K\epsilon_0 g \qquad\qquad k^2 = gdE$$

where k is the electromechanical coupling coefficient;
E is the modulus of elasticity (newtons/meter^2);
K is the relative dielectric constant of the material;
ϵ_0 is the dielectric constant of free space (8.85×10^{-12} farads/m).

The piezoelectric material used in this analysis is Glennite HST-41, a lead zirconate – lead titanate ceramic. Its relevant characteristics for thickness mode vibration are listed in Table II. The corresponding figures for X-cut quartz are listed for comparison.

TABLE II

Characteristics of Piezoelectric Materials

Coefficient	HST-41	Quartz
Electromechanical coupling coefficient, k_{33}	0.66	0.10
Charge coefficient, d_{33} (m/V $\times 10^{-12}$)	325	2.3
Voltage coefficient, g_{33} (Vm/newton $\times 10^{-3}$)	22	58
Free dielectric constant, K_3	1800	4.5

The bandwidth and mechanical output power of a piezoelectric transducer are related to the mechanical impedance of the materials to which it is coupled. For the materials and frequencies being considered here, the attenuation coefficient is small with respect to the propagation constant. For this case, the characteristic mechanical impedance of a material is real and, for longitudinal waves, is given by

$$Z = \rho c_L$$

where ρ is the density of the material.

The characteristic mechanical impedance of HST-41 ceramic is approximately 30×10^6 kg/m^2 sec.

It has been shown [17] that the largest mechanical gain–bandwidth product for a piezo-ceramic transducer is obtained with air backing and a load impedance matching the source impedance.

One of the commonly available materials that meets the velocity requirement for a Rayleigh wave radiator prism to be used on a glass surface is an acrylic resin such as Plexiglas or Lucite. Unfortunately, these materials have a characteristic impedance of about 3.2×10^6 kg/m^2 sec, which does not match the piezo-ceramic source impedance very well.

There are many analogies between stress wave propagation and electromagnetic wave propagation in transmission lines. By considering the stress and displacement boundary conditions at the interface between two materials, it can be shown that, for normal incidence by a longitudinal wave, the reflection coefficient is given by:

$$V = \frac{\rho_2 c_{L_2} - \rho_1 c_{L_1}}{\rho_2 c_{L_2} + \rho_1 c_{L_1}} = \frac{Z_2 - Z_1}{Z_2 + Z_1}$$

It can also be shown that the reflection coefficient will be reduced to zero if materials 1 and 2 are separated by a quarter-wavelength layer of material 3 with a characteristic impedance $Z_3 = \sqrt{Z_1 Z_2}$.

Attempts were made at using a quarter-wavelength transformer to obtain a closer impedance match between the transducer and the prism. The ideal matching material would, in this case, have an impedance of 9.8×10^6. Quarter-wavelength plates of magnesium having $Z = 10.1 \times 10^6$ were tried as matching transformers. However, this technique was abandoned because of the coupling losses introduced by the extra bonding layer, and the increased difficulty in achieving consistent results from one radiator to another.

On the basis of these tests and some experimentation with various backing materials, it was decided to sacrifice bandwidth for sensitivity by using air-backed transducers bonded directly to the prisms. The bonding was done using an epoxy adhesive which has a characteristic impedance similar to that of the prism. Close control of the bonding layer thickness was not required because the similarity of the impedances meant that the reflection coefficient was not affected significantly by the bond thickness.

The radiator bandwidth achieved by this method is approximately 1.7 MHz, or 20% of the nominal 8 MHz resonant frequency. Figure 5 is a graph of the parallel resistive and reactive components of the electrical input impedance of a test array

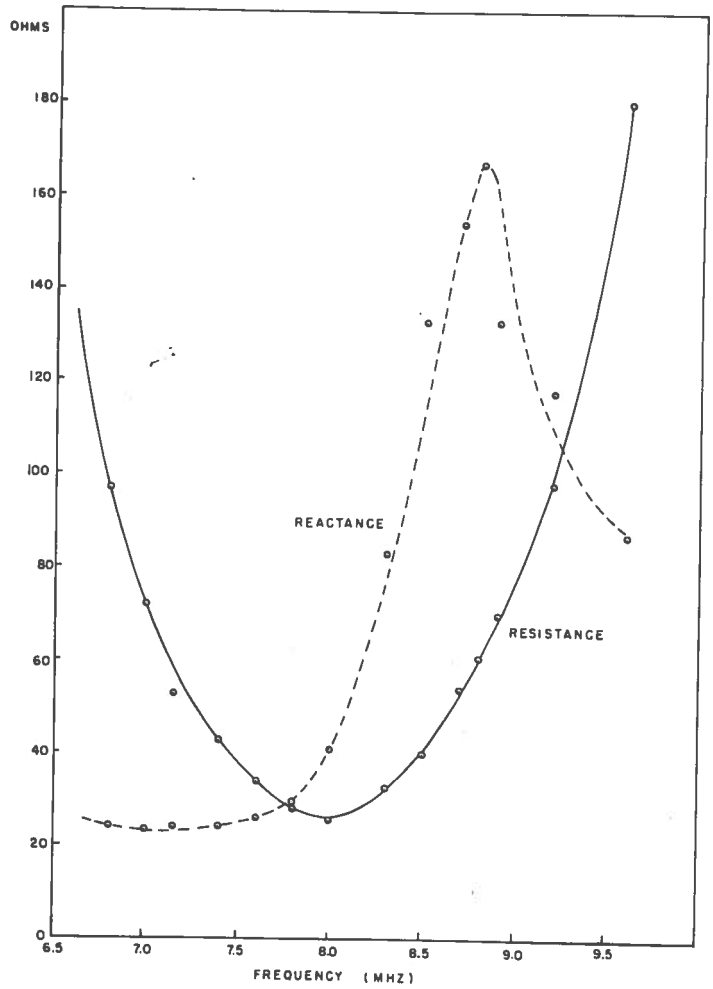


Figure 5 Transducer impedance components

versus frequency. The array consisted of four $\frac{1}{4}$ X $\frac{1}{2}$-inch transducers bonded to a Plexiglas block and electrically connected in series. The minor departures of the measured values from a smooth curve are primarily due to the small differences among the four transducers.

The minimum usable range of the radiators is determined by the time required for the initial clutter to drop below the signal threshold level. The clutter is caused by the longitudinal waves from the drive pulse experiencing multiple internal reflections within the prism. For these particular radiator assemblies, the energy is absorbed in about 35 μsec, corresponding to a range of 2.2 inches.

The coupling at the interface between the prism and the glass was found to be much more critical than the coupling from the transducers to the prism. It has already been noted that the surface wave attenuation in the interface is considerably higher than on a free surface. From tests performed with various coupling materials, indications are that the materials and layer thicknesses that provide good mechanical coupling from the prism to the glass are also the ones that make surface wave propagation in the interface difficult.

Liquid couplants such as oil or glycerine provide good Rayleigh wave outputs, but they are inconvenient for a fixed installation. However, soft rubber is a solid that is characteristically similar to liquids for stress wave propagation. The shear modulus, μ, of liquids is zero and shear waves cannot propagate through them. The shear modulus of rubber is also small with respect to its Lamé elastic constant λ. This is illustrated by μ/λ ratios of 0.72, 1.0, and 7.0×10^{-4} for steel, glass, and rubber, respectively, using values of μ and λ from Kolsky [12]. A silicone rubber adhesive was finally selected for bonding the prisms to the glass surface, with results similar to those obtained with liquid couplants.

Array Design

The requirement for this application was to generate a Rayleigh wave with a linear wavefront approximately 10 inches in width. It is not practical to fabricate a single 8-MHz transducer of this length, so a linear array of smaller radiators had to be constructed. To combine separate radiator elements into an array, it is necessary to know the shape of the beam radiated by an individual element. The elements can then be combined in such a way as to minimize the interference effects that are created by overlapping beams.

A rigorous analysis of the directivity of a prism type of Rayleigh wave radiator is complicated by the fact that the mode conversion to surface waves occurs in the near-field region of the longitudinal wave source. Another difficulty is the complexity of the surface wave propagation in the interface between the prism and the glass.

Viktorov and Zubova have derived an approximate expression for the directivity function [16] which was obtained by replacing the actual interface region with a system of elementary radiators distributed over the region. A factor $\exp(-mx)$ is included to

account for the surface wave attenuation in the interface region. The resulting expression is:

$$D(\beta) = \frac{m}{k_R}(1 + \coth m\ell)\ \frac{\sin(k_R w \sin\beta)}{k_R w \sin\beta}$$

$$\left(\sin^2\left[k_R \ell(1-\cos\beta)\right]\cosh^2\frac{m\ell}{\cos\beta} + \cos^2\left[k_R\ell(1-\cos\beta)\right]\right.$$

$$\left.\sinh^2\frac{m\ell}{\cos\beta}\right)^{1/2}\left[(1-\cos\beta)^2 + \left(\frac{m}{k_R\cos\beta}\right)^2\right]^{-1/2}\exp\frac{-m\ell}{|\cos\beta|}$$

where β is the angle from the normal to the radiator;

2ℓ is the length of the interface in the direction $\beta = 0$ illuminated by longitudinal waves;

$2w$ is the width of the interface perpendicular to the direction $\beta = 0$;

k_R is the Rayleigh wave number;

m is an experimentally determined coefficient.

Directivity measurements were made on prism type radiators to compare experimental results with the analytical ones. The measurement apparatus consisted of a probe mounted in a holder capable of traveling across the test surface along a line parallel to the radiator length. The probe, as shown in Fig. 6, was constructed to have a wide bandwidth for use in a pulsed measurement system. A small amplitude correction factor has been computed for each measurement obtained from the probe so that the results will correspond to the values that would be measured along an arc of constant radius.

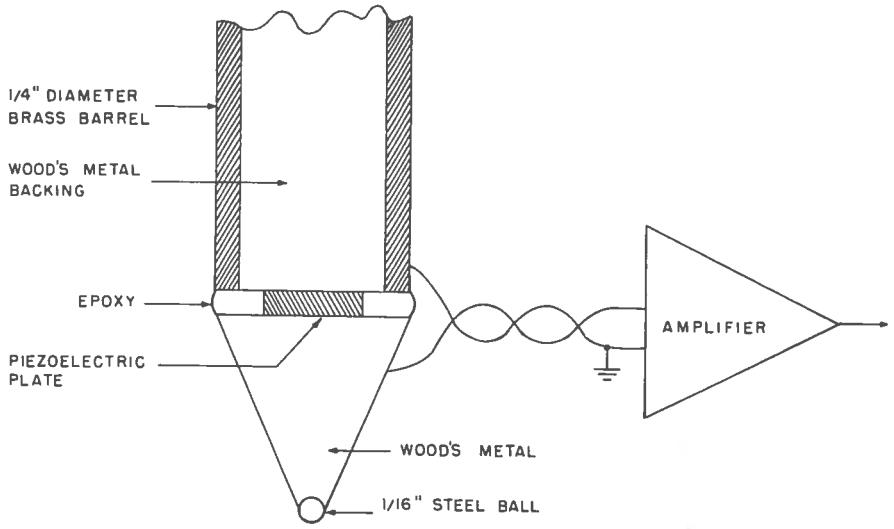


Figure 6 Rayleigh wave probe

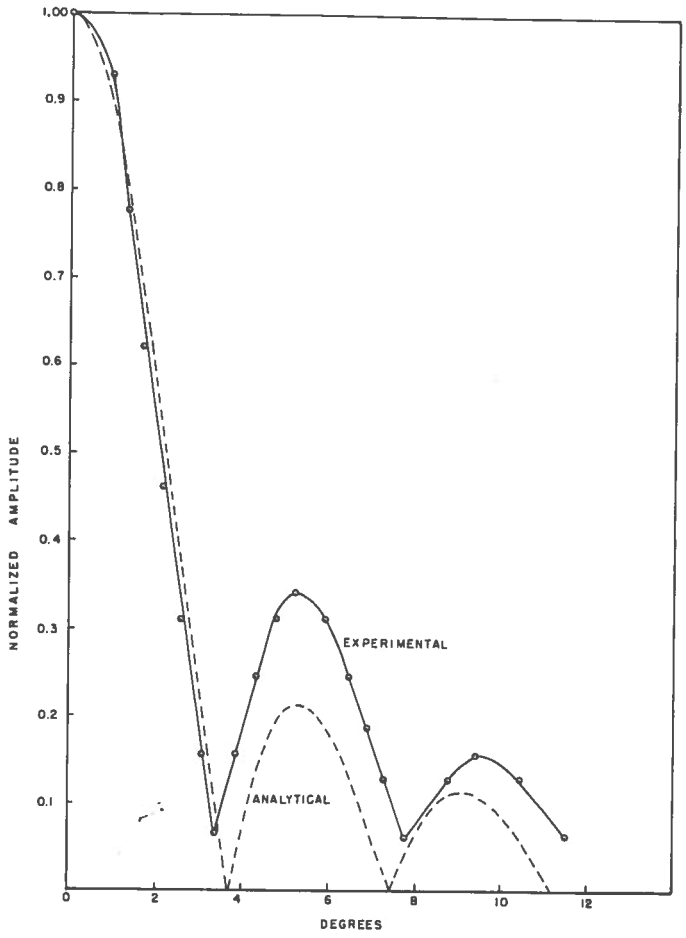


Figure 7 Directivity pattern for Rayleigh wave radiator

Figure 7 shows a comparison of the corrected and normalized experimental results with those computed from the directivity expression listed above. The beamwidths compare very well, but the sidelobes in the experimental results are considerably larger than the analytical expression predicts. From the theory, one would expect sidelobes 13.4 db below the main beam. However, the experimental results from many trials indicated sidelobes approximately 9.5 db below the main beam. In order to produce a satisfactory array pattern, an attempt was made to reduce the sidelobe level of the radiators.

In analyzing the radiation of spatial waves, it is well known that, for a square radiating surface, the directivity pattern observed in a plane parallel to one of the radiator edges can be described by a $\frac{\sin F}{F}$ function. However, in an observation plane through a diagonal of the square, the directivity pattern is described by $\frac{\sin^2 F}{F^2}$. This pattern, relative to the first, has approximately the same -3 db beamwidth, but the sidelobes are 13 db lower and in phase with the main beam.

Because a prism type radiator essentially transforms the spatial radiation in the prism into radiation along a surface plane, it was felt that the orientation of the longitudinal wave radiator relative to the transformation plane would affect the directivity pattern of the resulting Rayleigh waves. Consequently, a radiator was made using a square piezoelectric plate mounted with its diagonal parallel to the interface plane. Directivity measurements made on this radiator showed a pattern similar to the previous radiators, with no marked change in the sidelobe level. Similar results were obtained when the experiment was repeated using a source cut to the shape of an isosceles triangle with its base parallel to the interface plane. The failure of these tests to agree with the results anticipated by the analogy with spatial wave radiators was attributed to the high attenuation of surface waves in the prism–glass interface. This attenuation tends to distort any pattern created at the interface by the spatial wave source.

For a practical overlay size, the region of interest corresponds to the far-field region of the individual array elements, but to the near-field region of the over-all array. In the near field which, for practical purposes, can be considered to extend to a range of about $w^2/1.5\lambda_R$, the path lengths to a given point from various parts of a radiator can differ by several wavelengths, causing interference effects.

Linear array configurations were evaluated by computing the response along paths at various ranges using the element width, spacing, and relative phase as parameters. On the basis of results obtained, a radiator width ($2w$) of 0.465 inch and a center-to-center spacing of 0.565 inch were used. Eighteen radiators were required to make up each 10-inch array.

The extent of the near field for individual radiators 0.465 inch wide corresponds very closely to the minimum usable range of 2.2 inches determined by the drive pulse clutter. For a 10-inch radiator, however, the near-field region extends many feet, and as a result, the radiation pattern within the region of interest was characterized by nulls and peaks.

When the arrays were assembled and tested, it was found that the actual radiation pattern was more irregular than the computations indicated. Factors contributing to this discrepancy were the variation in the spacing, orientation, and bond characteristics due to assembly tolerances. The gaps in the pattern were sufficiently large and numerous that it was necessary to add a second set of arrays on the opposite sides of the plate. This set is offset with respect to the first so that beams from opposite sides are effectively interleaved. The arrays are energized sequentially to avoid mutual interference.

The radiator elements in each array are electrically connected in series to facilitate impedance matching between the array and the electrical driving circuit. For a series connection, the electrical input impedance of each array at resonance is about $150 - j\,200$ ohms.

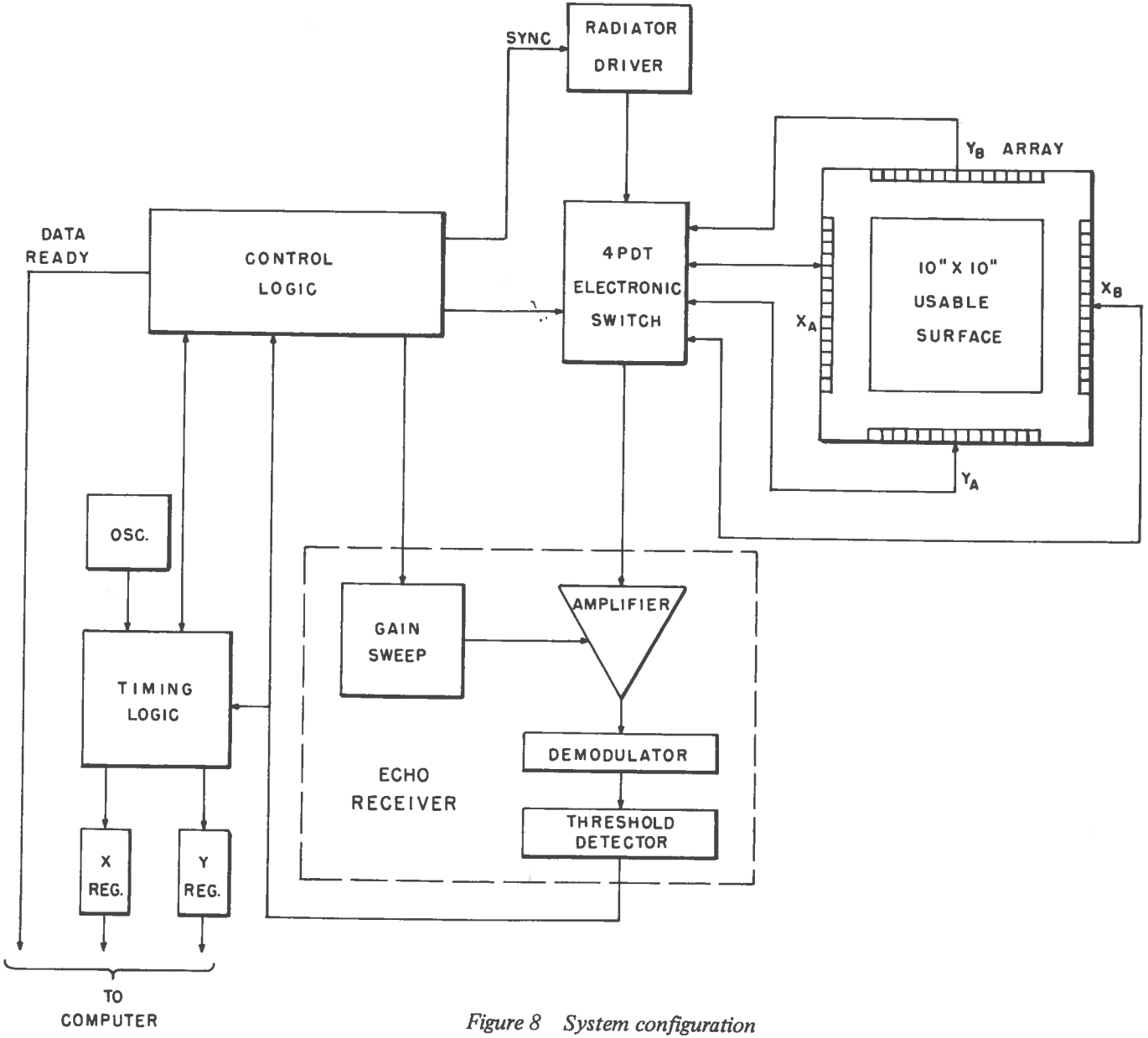
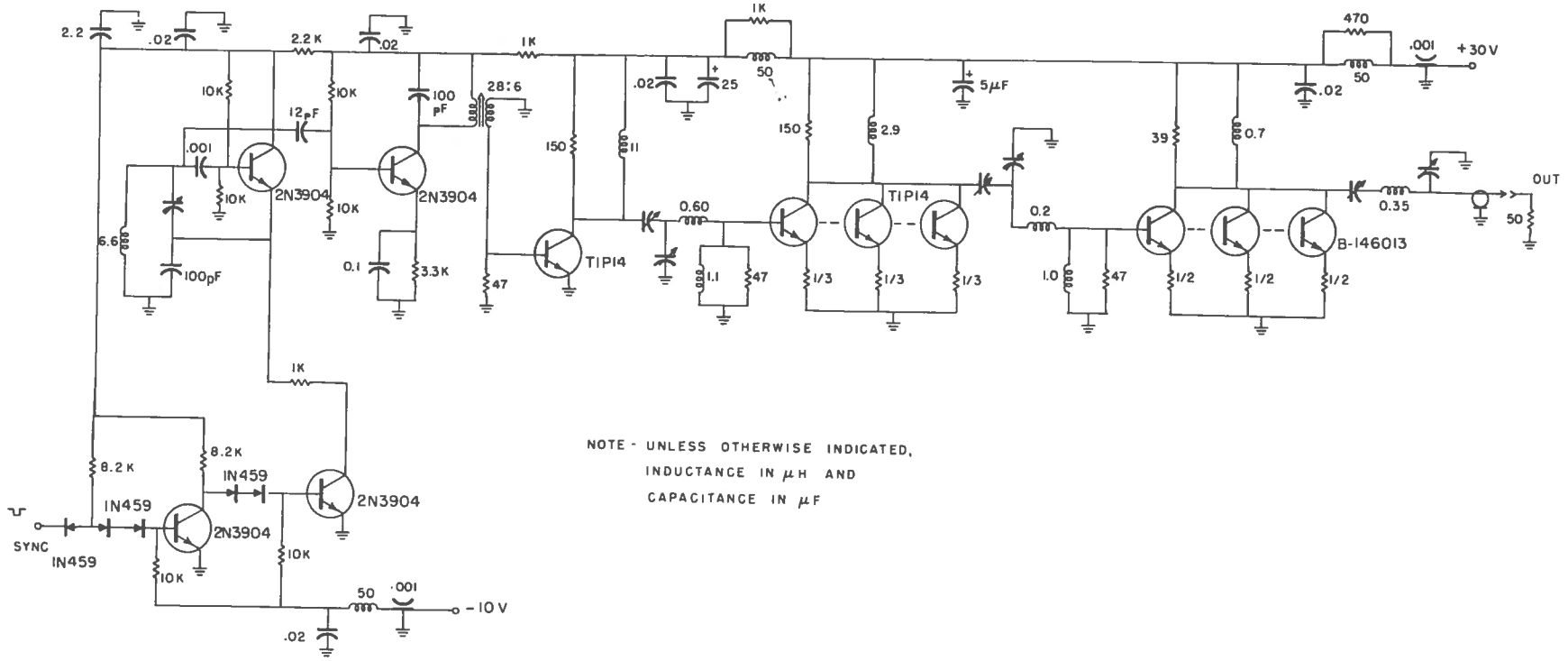


Figure 8 System configuration



NOTE - UNLESS OTHERWISE INDICATED,
INDUCTANCE IN μH AND
CAPACITANCE IN μF

Figure 9 Radiator driver

IV Echo-Ranging Circuitry

The signal processing circuitry consists of a radiator driver, an echo receiver, and and electronic switch for signal routing. The block diagram in Fig. 8 shows how these are connected into the system.

The send–receive cycle for each array, which will be referred to as a scan, is initiated by a sync pulse from the control logic. A driver pulse is produced and routed through the switch to the correct radiator array. During this pulse, the receiver is disconnected from the signal path and the input is short-circuited. After a short transition delay, the receiver is connected to the array, and the driver is disconnected. This state is maintained until an echo larger than a predetermined threshold is sensed, or until the elapsed time corresponds to the maximum allowable range. The receiver is then disconnected, and its input short-circuited again.

The Radiator Driver

The driver consists of a gated oscillator followed by four stages of power amplification, as shown in the schematic of Fig. 9. The required driver pulse length is approximately 3 μsec with a duty cycle of less than 0.1%. Because of the low duty cycle and a pulse length shorter than the thermal time constants, the pulse power output was limited by the current and voltage ratings of the transistors rather than by dissipation. No attempt was made to optimize the collector efficiency. Resistive damping was added to the RF chokes to decrease the ringing tendency after a pulse. Emitter resistors were used with the parallel transistors of the final stages to obtain better load sharing.

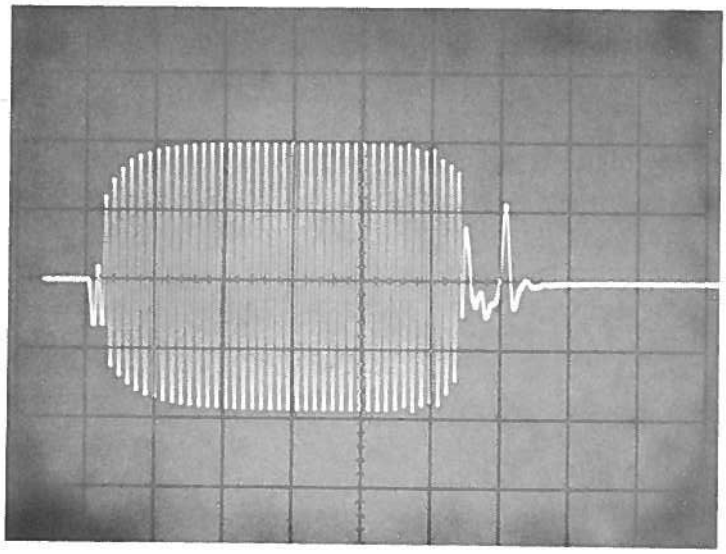


Figure 10 Output of driver into resistive load, 50 V/cm and 1 μsec/cm

The circuit was designed to work into a load resistance of 50 ohms. A series-tuned L network matches the 5-ohm transistor output impedance to the load, and T networks are used for interstage matching. In each case, the type of network was chosen to use practical component values. The driver output across a 50-ohm resistive load for a 5-μsec pulse is shown in Fig. 10. The peak pulse power is approximately 90 watts.

Each transducer array is connected to the driver by a 50-ohm coaxial cable. The parallel impedance components of each array at 8 MHz are approximately 150 ohms resistive and 200 ohms capacitive. This is transformed to 50 ohms resistive with an L matching network, the transducer capacitance forming a part of the network.

The Electronic Switch

This switch, a four-pole double-throw type, serves a dual purpose in the system; it permits the four separate arrays to be multiplexed into a single driver and receiver, and isolates the receiver from the circuit during the driver pulse. The proper switching sequence is selected by the control logic. The circuit, shown in Fig. 11 is designed to minimize electrical switching transients at the radiators in order to prevent the generation of spurious waves. Each switch element basically consists of a pair of diodes that are forward or reverse biased by a bias control transistor. A forward bias current of several milliamperes establishes sufficient diode conduction to allow the passage of at least 100 watts of RF pulse power.

At 8 MHz, the switch provides 40-db isolation in the OFF state, and approximately 0.5-db insertion loss in the ON state. Switching between states requires several microseconds, which is adequate for this application.

The Echo Receiver

The receiver consists of four stages of RF amplification followed by a demodulator and a threshold detector, as shown in Fig. 12. The input stage uses a 2N930 low-noise transistor to minimize the noise figure. This is followed by three integrated-circuit amplifier stages with single-tuned transformers for interstage coupling. Each integrated circuit is a differential input, single-ended output amplifier with a voltage gain of approximately 24 db.

The complete receiver has a maximum stable RF voltage gain of 110 db at 8.0 MHz. This is with a bandwidth of 0.8 MHz and a peak output of 1 volt. The noise figure measured under these conditions is 2.0 db.

The receiver gain is electronically swept during each scan in order to compensate partially for the surface wave attenuation with increasing range. Gain control is provided by a voltage sweep circuit connected to the integrated circuit amplifiers.

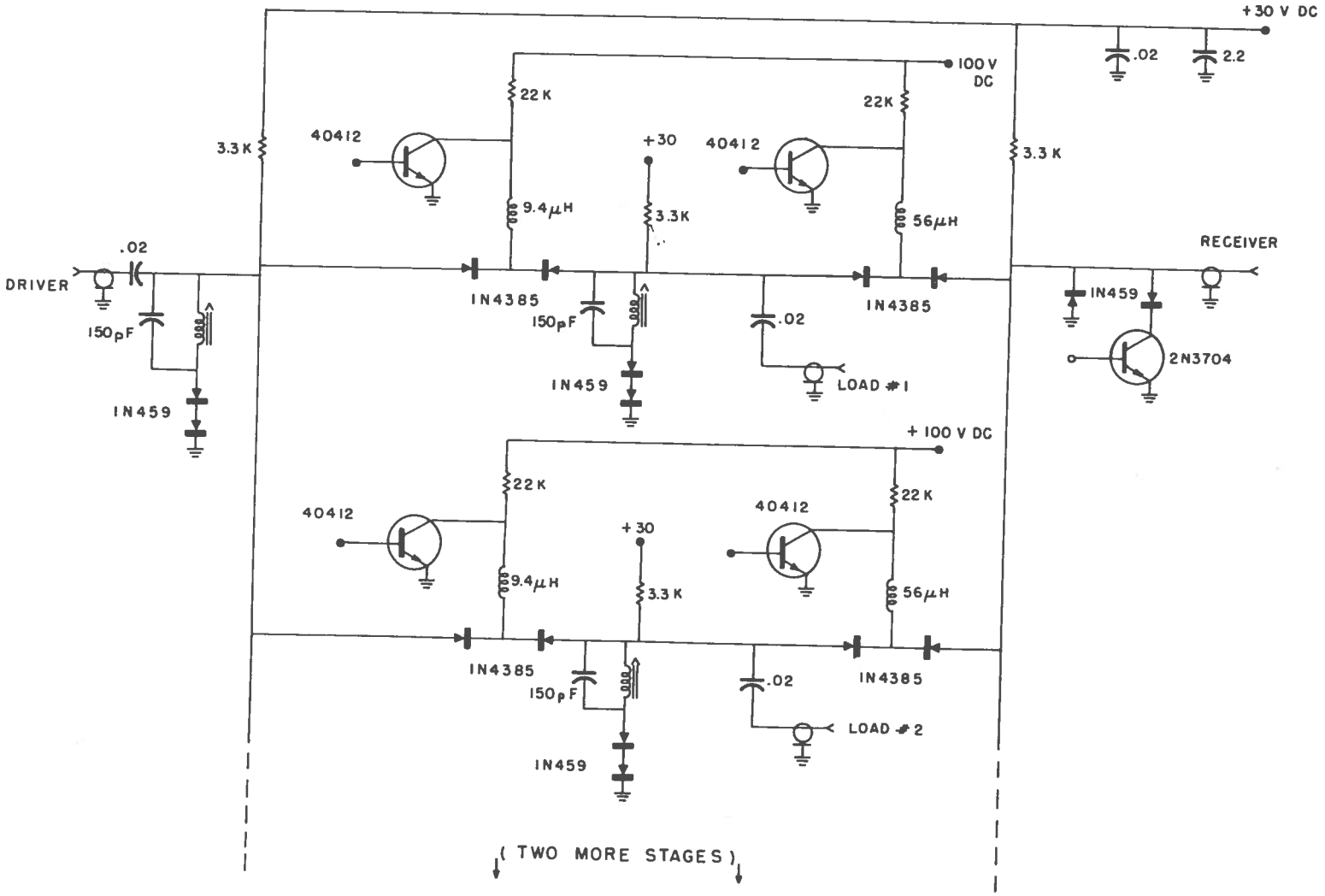


Figure 11 Electronic switch

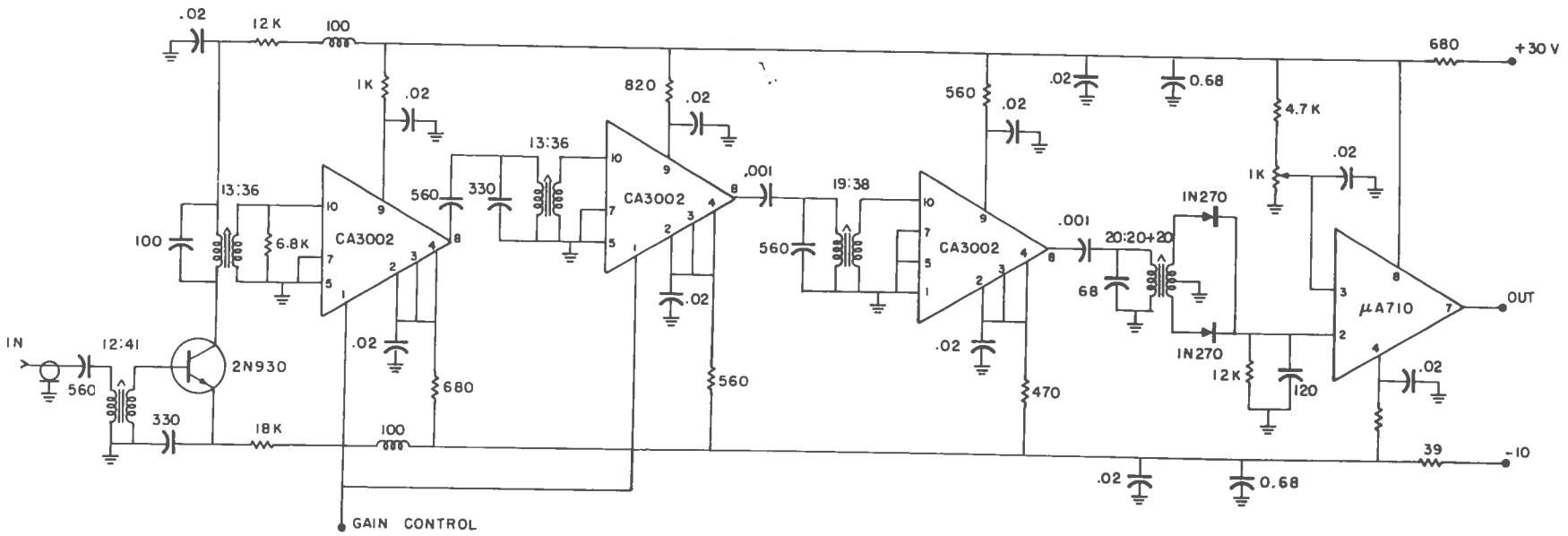


Figure 12 Echo receiver

The threshold detector is a high-speed differential comparator in integrated circuit form. It has a response time of 40 nanoseconds, and the output voltage is compatible with the digital integrated circuits used in the system. The threshold is set high enough to discriminate against echoes caused by fingerprints and marks on the glass.

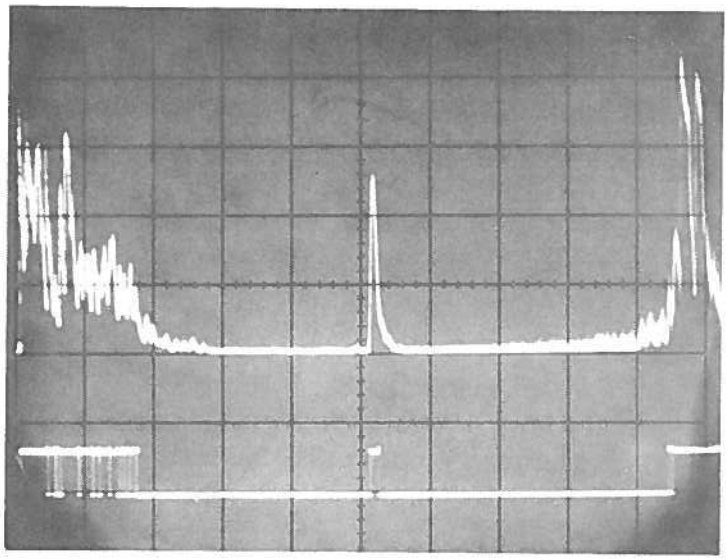


Figure 13 Demodulator and threshold detector outputs, 25 μsec/cm, upper 0.5 V/cm, lower 5.0 V/cm

Figure 13 shows the demodulator output and the threshold detector output for a single complete scan. The receiver gain is being swept through approximately 30 db. Initial clutter from within the prism is visible on the left. At the center is an echo caused by a finger touching the glass. The echo on the right is from the opposite side of the plate.

V Timing and Control Circuitry

Timing Logic

The coordinate grid on which the system is based is considered to have x and y axes coincident with the edges of the usable surface, the origin being in the lower left corner. This puts the entire usable aperture in the first quadrant, making all the coordinates positive. If required, coordinate scaling and axes translation could be performed by the computer software.

The four transducer arrays and the corresponding scans will be designated as x_A, x_B, y_A, y_B. For each scan, a free-running binary counter is started at a reference time and stopped by the first echo signal above a fixed threshold. For the A scans, originating

at the lower and left edges, the counter contents represent the target coordinates for a counting-up process starting from zero at the appropriate axis. For the B scans, on the other hand, increasing elapsed time corresponds to decreasing coordinate values. By choosing reference lines coincident with the upper and right boundaries of the aperture, the counter contents represent the target coordinates for a counting-down process starting from the maximum value at the reference lines. Ideally, the difference in target coordinates as determined by the A scan and the B scan is equal to the length of the target in that axis.

The timing circuitry contains a single up–down counter from which data can be transferred into either an x or y buffer register.

One problem encountered in using the double scan per axis approach is caused by the wave packet generated at one array striking the opposite array. This produces a voltage pulse approximately 80 db above the echo threshold level and the time of its occurrence corresponds to an echo originating at the exact midline between the two arrays. Because this opposite array is separated from the receiver by a switch having only 40-db isolation, the signal appeared as a strong echo. This problem is not apparent in Fig. 13 because the opposite array was disconnected at the time. After unsuccessful attempts to provide adequate isolation, the problem was circumvented by closing the receiver range gate during that time so that the timing circuitry was unaffected by the signal. Because of the overlapping effect of scanning from opposite sides, the gap in coverage can be limited to about $\frac{1}{8}$ inch.

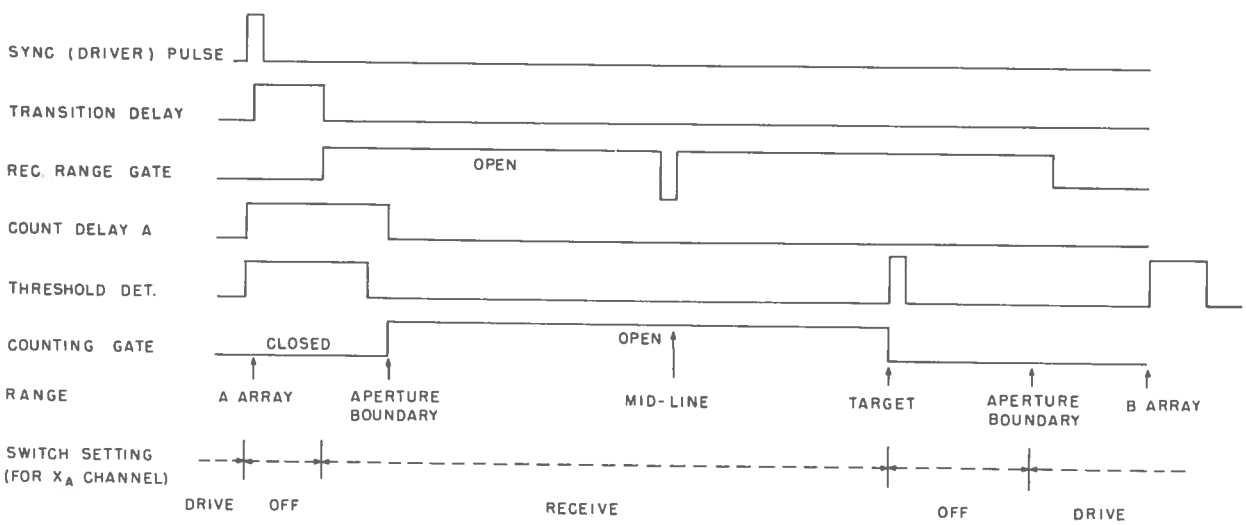


Figure 14 x_A scan sequence

The spacing between the arrays and the reference lines makes it necessary to introduce a delay between the driver pulse and the start of counting. Figure 14 shows the sequence of operations that occurs for a single scan.

The timing circuitry is calibrated by first adjusting the A count delay so that counting begins at the desired axis location. The counting rate is then adjusted to provide a full count at the desired location for the opposite boundary. For simplicity, this boundary value is constrained to be a power of two. Finally, the B count delay is adjusted so that B scan counting begins at the same boundary.

The size and position of the coordinate grid with respect to the arrays can be modified by recalibrating with the procedure outlined above. This would allow coordinate registration with the grid on a display screen that may be behind the plate.

Control Logic

The scanning sequence necessary to obtain the stylus location, referred to as a sample, is determined by the control logic. In normal operation, each sample can consist of two, three, or four separate scans. This is because the B scan for either axis is skipped if an echo occurs during the A scan for that axis. Switches, included in the developmental model for testing and evaluation purposes, enable the user to select an operating sequence consisting of either the A or B scans separately, or normal operation.

The x and y buffer registers are up-dated with the results of each sample. In addition, a DATA READY signal is generated if the register contents are to be transferred to the computer.

The control logic provides two modes of operation: a continuous mode, and a discrete mode. These modes, manually selected by a switch setting, are designed to provide efficient data generation for position selection and item selection respectively.

A blank sample, which means any sample for which an echo is not received on both x and y, does not initiate a data transfer; i.e., a DATA READY signal is not generated. For noise rejection purposes, the first non-blank sample does not produce a DATA READY signal either. The second consecutive non-blank sample generates a DATA READY signal in both operating modes. Subsequent samples provide DATA READY signals in the continuous mode only.

The data generated in the continuous mode represents, in real time, the path described by the stylus on the surface. This is necessary for position selection functions such as entering drawings into the system.

In the discrete mode, on the other hand, the computer is informed of the initial contact point only. A single data transfer occurs for each stylus contact, regardless of the

number of samples for which the contact is maintained. This mode is designed to eliminate a considerable amount of superfluous data if the device is being used for item selection purposes.

The flow diagram in Fig. 15 shows the alternative operating sequences for one sample. The timing and control circuits were constructed using digital integrated circuits of the transistor-transistor logic (TTL) type. Approximately 125 gates and 35 flip flops were required.

VI System Performance

The device has been interfaced with a Digital Equipment Corporation PDP-8 computer for testing purposes. After the appropriate input/output selection by the computer program, a DATA READY signal from the input device can produce a computer interrupt. An interrupt service routine then performs a data transfer from the x and y buffer registers into the computer. This is useful in that the sequence of locations representing the path of a moving stylus can be stored in real time at several hundred samples per second, and printed out later for individual examination.

It was difficult to conduct quantitative performance tests because of the non-uniformity of the radiation pattern over the surface. It was found that a contact area approximately $\frac{1}{4}$ inch in diameter is necessary to ensure reasonable operation anywhere within the working surface aperture. The contact area must be as large as that to bridge the regions of low sensitivity that exist in spite of the double scan per axis arrangement. This means that even though the inherent positional resolution, as determined by the bandwidth, is approximately 0.03 inch, the working resolution is about an order of magnitude below that. This restricts the use of the present experimental model to item selection with a finger or a stylus of similar size.

The echo strength at the receiver is determined by the size of the surface contact area on a macroscopic scale, and also on the degree of actual contact within that area on a microscopic scale. Therefore, for a given size, a material that is soft enough to conform to the surface is more practical than a hard material for the stylus tip. In addition, when using a finger, the natural moisture aids in establishing a good contact. As an example, the echo in Fig. 13 is from a finger touching a region of good sensitivity.

Using the device with a finger, however, imposes an upper limit on the usable sensitivity because of the requirement that echoes from fingerprints remain below the signal threshold. This restriction does not significantly affect the operation with finger-sized styli. However, to obtain sufficient echo power from styli having a small

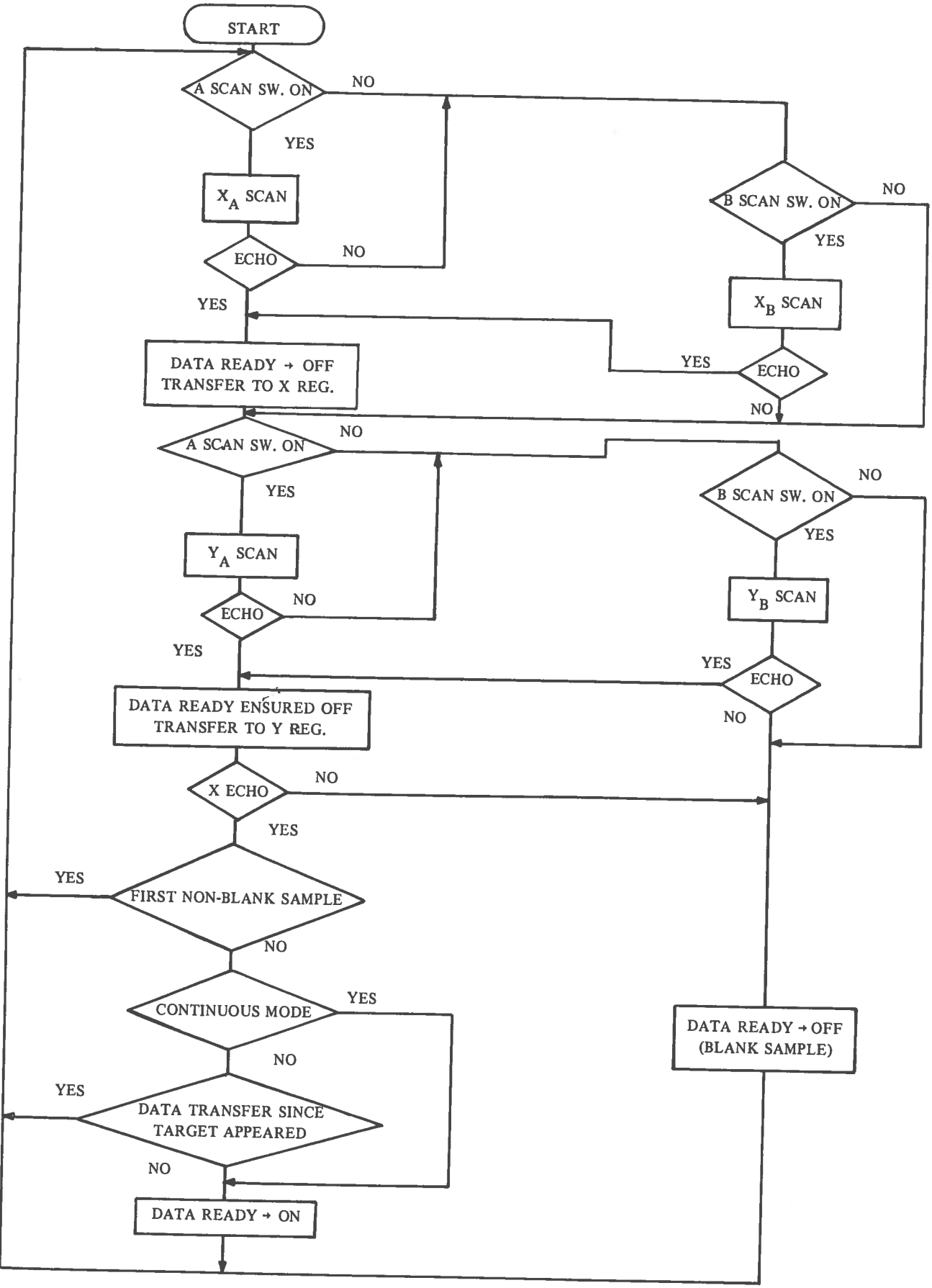


Figure 15 Sample sequence

over-all contact area, the tip must be designed to achieve good coupling with the glass surface. One technique that has proved successful is to use a fine felt-tip pen filled with a clear volatile liquid such as alcohol. Because the device is sensitive to scratches and dirt on the glass, a reasonable amount of care must be used in order to keep the surface clean.

Although the present device has served as an experimental model for the evaluation of $x-y$ position encoding by surface wave echo ranging, its usefulness in practical applications is limited by the poor radiation pattern of the arrays. It is felt that a more sophisticated approach to the array design and assembly could achieve a relatively uniform radiation pattern. In addition, the usable dynamic range for stylus echoes may be increased by lowering the carrier frequency. These modifications would permit operation with styli small enough to utilize the inherent resolution of the system, and would improve the performance with all types of styli. Efforts are being made to implement these changes so that the full potential of surface wave echo ranging can be demonstrated more effectively.

References

1. Gurley, B.M. and Woodward, C.E. Light pen links computer to operator. Electronics, 32 (47): 85; 1959.
2. Haring, D.R. The beam pen: a novel high speed input/output device for cathode ray tube display systems. Proc. FJCC, 27: 847; 1965.
3. David, M.R. and Ellis, T.O. The RAND tablet: a man–machine graphical communication device. Proc. FJCC, 26: 325; 1964.
4. Hargreaves, B., Joyce, J.D., Cole, G.L., et al. Image processing hardware for a man–machine graphical communication system. Proc. FJCC, 26: 363; 1964.
5. Woo, P.W. A proposal for input of hand drawn information to a digital system. IEEE Trans. on Electronic Computers, EC-13: 609; 1964.
6. Roberts, L.G. The Lincoln wand. Proc. FJCC, 28: 223; 1966.
7. Teixeira, J.F. and Sallen, R.P. The Sylvania data tablet: a new approach to graphic data input. Proc. SJCC, 32: 315; 1968.
8. Lewin, M.H. A magnetic device for computer graphical input. Proc. FJCC, 27: 831; 1965.
9. Johnson, E.A. Touch display – a novel input/output device for computers. Electronics Letters, 1 (8): 219; 1965.
10. Minton, W.C. Inspection of metals with ultrasonic surface waves. Nondestructive Testing, 12 (4): 13; 1954.
11. Cook, E.G. and Van Valkenburg, H.E. Surface waves at ultrasonic frequencies. ASTM Bulletin, No. 198: 81; 1954.
12. Kolsky, H. Stress waves in solids. Oxford Press, New York, 1953.
13. Brekhovskikh, L.M. Waves in layered media. Academic Press, New York, 1960.
14. Viktorov, I.A. Rayleigh waves in the ultrasonic range. Soviet Physics – Acoustics, 8 (2): 118; 1962.

15. Viktorov, I.A. Investigation of methods for exciting Rayleigh waves. Soviet Physics – Acoustics, 7 (3): 236; 1962.

16. Viktorov, I.A. and Zubova, O.M. Directivity diagrams of radiators of Lamb and Rayleigh waves. Soviet Physics – Acoustics, 9 (2): 138; 1963.

17. Kossoff, G. The effects of backing and matching on the performance of piezoelectric ceramic transducers. IEEE Trans. on Sonics and Ultrasonics, SU-13 (1): 20; 1966.

APPENDIX I

Characteristics of Rayleigh Waves

An unbounded solid medium can support two kinds of elastic stress waves; longitudinal and transverse, or shear, waves. These are spatial waves that propagate within the material. A solid which has a bounding surface can support a third type of wave which propagates along the surface. In the case of a free surface, i.e., one bounded by a vacuum, these surface waves are generally known as Rayleigh waves. The basic characteristics of Rayleigh waves are derived in this appendix.

Consider a perfectly elastic, homogeneous, isotropic solid material with a free surface in the z plane at $z = 0$, the solid occupying the half-space for $z > 0$.

For such a solid, stresses and the corresponding strains are related by a generalized expression of Hooke's Law:

$$\sigma_{xx} = \lambda\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right) + 2\mu\frac{\partial u_x}{\partial x}$$

$$\sigma_{xz} = \sigma_{zx} = \mu\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right)$$

and similar expressions for σ_{yy}, σ_{zz}, σ_{xy} and σ_{yz}.

σ_{xx} is the stress in the x direction acting in the x plane; i.e., a longitudinal stress.

σ_{xz} is the stress in the x direction acting in the z plane; i.e., a transverse stress.

u_x is the x component of the particle displacement.

λ and μ are elastic constants of the solid known as Lamé constants. (μ is also known as the shear modulus, or the modulus of rigidity).

Applying Newton's second law for dynamic equilibrium in an infinitesimal cube $\partial x\ \partial y\ \partial z$, and simplifying, we obtain for the x direction:

$$\rho\frac{\partial^2 u_x}{\partial t^2} = \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{xz}}{\partial z}$$

and similar equations for the y and z directions, where ρ is the density of the material.

Using the previous expressions for the stresses:

$$\rho \frac{\partial^2 u_x}{\partial t^2} = (\lambda + \mu) \frac{\partial}{\partial x} \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} \right) + \mu \nabla^2 u_x$$

for the x direction.

We can define a scalar potential ϕ and a vector potential Ψ such that the particle displacement can be expressed as:

$$\mathbf{U} = \text{grad } \phi + \text{curl } \Psi \ .$$

If we consider waves propagating in the x direction,

$$\frac{\partial \mathbf{U}}{\partial y} = \frac{\partial \phi}{\partial y} = \frac{\partial \Psi}{\partial y} = 0.$$

Ψ can then be chosen such that $\psi_x = \psi_y = 0$. Henceforth, ψ_y will be denoted as ψ. For these conditions:

$$u_x = \frac{\partial \phi}{\partial x} - \frac{\partial \psi}{\partial z}$$

$$u_y = 0$$

$$u_z = \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial x}$$

Substituting these expressions into the previous ones for the x and z directions, we get:

$$\rho \frac{\partial}{\partial x} \left(\frac{\partial^2 \phi}{\partial t^2} \right) - \rho \frac{\partial}{\partial z} \left(\frac{\partial^2 \psi}{\partial t^2} \right) = (\lambda + 2\mu) \frac{\partial}{\partial x} (\nabla^2 \phi) - \mu \frac{\partial}{\partial z} (\nabla^2 \psi)$$

$$\rho \frac{\partial}{\partial z} \left(\frac{\partial^2 \phi}{\partial t^2} \right) + \rho \frac{\partial}{\partial x} \left(\frac{\partial^2 \psi}{\partial t^2} \right) = (\lambda + 2\mu) \frac{\partial}{\partial z} (\nabla^2 \phi) + \mu \frac{\partial}{\partial x} (\nabla^2 \psi)$$

These two equations are satisfied if:

$$\frac{\partial^2 \phi}{\partial t^2} = \left(\frac{\lambda + 2\mu}{\rho} \right) \nabla^2 \phi$$

and

$$\frac{\partial^2 \psi}{\partial t^2} = \left(\frac{\mu}{\rho} \right) \nabla^2 \psi$$

These wave equations represent the longitudinal and transverse components, respectively, of any stress wave propagating in the solid. The phase velocities of these components are:

$$c_L = \sqrt{\frac{\lambda + 2\mu}{\rho}} \quad \text{and} \quad c_T = \sqrt{\mu/\rho}$$

The corresponding wave numbers are:

$$k_L = \omega/c_L \quad \text{and} \quad k_T = \omega/c_T$$

Considering sinusoidal variation with time, the wave equations become:

$$\nabla^2 \phi = -k_L^2 \phi$$

$$\nabla^2 \psi = -k_T^2 \psi$$

Assume solutions of the form:

$$\phi = F(z) \exp jk_R x$$

$$\psi = G(z) \exp jk_R x$$

in which the time factor $\exp j\omega t$ is implicit.

Substituting back into the wave equations:

$$\frac{d^2F(z)}{dz^2} - (k_R^2 - k_L^2)\, F(z) = 0$$

$$\frac{d^2G(z)}{dz^2} - (k_R^2 - k_T^2)\, G(z) = 0$$

The general solutions for these equations are:

$$\phi = \left[A_1 \exp\left(\sqrt{k_R^2 - k_L^2}\; z\right) + A_2 \exp\left(-\sqrt{k_R^2 - k_L^2}\; z\right)\right] \exp\left(jk_R x\right)$$

$$\psi = \left[B_1 \exp\left(\sqrt{k_R^2 - k_T^2}\; z\right) + B_2 \exp\left(-\sqrt{k_R^2 - k_T^2}\; z\right)\right] \exp\left(jk_R x\right)$$

For a solution to represent a Rayleigh wave, it has to decay with depth into the material; i.e, with increasing z. Therefore, for the case under consideration, we can set A_1 and B_1 to zero. Now we apply the boundary conditions that the stress components normal to the surface are zero at the surface,

i.e.,

$$\sigma_{zz} = \sigma_{zx} = 0 \qquad \text{at } z = 0$$

In terms of ϕ and ψ, the stresses are:

$$\sigma_{zz} = (\lambda + 2\mu)\frac{\partial^2\phi}{\partial z^2} + \lambda\frac{\partial^2\phi}{\partial x^2} + 2\mu\frac{\partial^2\psi}{\partial x \partial z} = 0$$

$$\sigma_{zx} = \mu\left(2\frac{\partial^2\phi}{\partial x \partial z} - \frac{\partial^2\psi}{\partial z^2} + \frac{\partial^2\psi}{\partial x^2}\right) = 0$$

Letting $q^2 = (k_R^2 - k_L^2)$ and $s^2 = (k_R^2 - k_T^2)$,

$$\phi = A \exp(-qz) \exp(jk_R x)$$

$$\psi = B \exp(-sz) \exp(jk_R x)$$

By substituting these expressions into the previous ones, and then combining, squaring both sides, and dividing by $\mu^2 k_R^8$ we obtain:

$$\eta^6 - 8\eta^4 + 8(3 - 2\alpha^2)\eta^2 - 16(1 - \alpha^2) = 0$$

where $\alpha = c_T/c_L$ and $\eta = c_R/c_T$.

This equation, a cubic in η^2, is called the Rayleigh equation. The Rayleigh wave corresponds to a real root, η_R, of the equation which lies between zero and one.

The phase velocity of a Rayleigh wave can then be expressed as

$$c_R = \eta_R c_T = \eta_R \sqrt{\mu/\rho}$$

This shows that the phase velocity depends only on the properties of the material in which it is propagating and, therefore, does not exhibit dispersion.

The elastic properties of a material can be expressed in terms of other constants which depend upon λ and μ. One of these is Poisson's ratio, ν, which can be expressed as:

$$\nu = \frac{\lambda}{2(\lambda + \mu)}$$

so that

$$\alpha^2 = \frac{\mu}{\lambda + 2\mu} = \frac{1 - 2\nu}{2 - 2\nu}$$

An approximate solution of Rayleigh's equation for η_R in terms of ν is given in reference 11 as:

$$\eta_R = \frac{0.87 + 1.12\nu}{1 + \nu}$$

Using plate glass as an example, the relevant constants are [12]:

$$\nu = 0.25$$
$$\mu = 2.8 \times 10^{11} \text{ dynes/cm}^2$$
$$\lambda = 2.8 \times 10^{11}$$
$$\rho = 2.51 \text{ grams/cm}^3$$

Then, $\eta_R = 0.92$,

and $c_R = 0.92\, c_T = 3070$ m/sec

The particle displacements for a Rayleigh wave are found by substituting the exponential solutions for ϕ and ψ into the expressions for u_x and u_z, and taking the real parts of the results.

Then,

$$u_x = \acute{A} k_R \left[\exp(-qz) - \frac{2qs}{s^2 + k_R^2} \exp(-sz) \right] \sin k_R x$$

$$u_z = Aq \left[\exp(-qz) - \frac{2k_R^2}{s^2 + k_R^2} \exp(-sz) \right] \cos k_R x$$

The sinusoidal variation with time is implicit in these equations.

The particle displacements are described by an elliptical orbit with the major axis perpendicular to the surface (u_z), and the minor axis parallel to the direction of propagation (u_x). The eccentricity of the ellipses, (u_x/u_z), depends on the distance from the surface and on the elastic constants of the material. For glass, the eccentricity for particles at the surface is 0.68.

It has already been shown that the decay of Rayleigh waves with depth into the material is an exponential function. The exponential decay factors, q, for the longitudinal component, and s, for the transverse component, can be expressed in terms of the Rayleigh wavelength λ_R:

$$q = \sqrt{k_R^2 - k_L^2} = \frac{2\pi c_R}{\lambda_R} \sqrt{\frac{1}{c_R^2} - \frac{1}{c_L^2}} = \frac{2\pi}{\lambda_R} \sqrt{1 - \eta_R^2 \alpha^2}$$

$$s = \frac{2\pi}{\lambda_R} \sqrt{1 - \eta_R^2}$$

Using glass, with $\nu = 0.25$, as an example:

$$q = \frac{5.36}{\lambda_R} \quad \text{and} \quad s = \frac{2.47}{\lambda_R}$$

Substituting these values into the expressions for particle displacement, it is found that, at a depth of λ_R, u_x has decayed to 0.107 and u_z to 0.195 of the surface value. Considering the wave energy to be proportional to the sum of the squares of the displacements, the energy at a depth of λ_R is only 3% of the energy at the surface. An important implication of this result is that any solid plate having a thickness that is large with respect to the Rayleigh wavelength can be considered to be a close approximation to the solid half-space assumed in the analysis.

The amplitude of elastic waves decreases with distance from their source through three mechanisms – beam divergence, scattering, and absorption. Because a Rayleigh wave is essentially a two dimensional phenomenon, the decrease in amplitude due to beam divergence is proportional to $1/\sqrt{r}$, where r is the distance from the source. This is in contrast to spatial waves which follow a $1/r$ relationship. The Rayleigh wave attenuation due to scattering and absorption is related to that of longitudinal and transverse spatial waves [14], and the attenuation factor is approximately proportional to the frequency in the ultrasonic range. Rayleigh waves are able to propagate along curved surfaces in addition to plane surfaces. The phase velocity is slightly dependent upon the radius of curvature of the surface with respect to the wavelength, being greater than c_R for a convex surface, and less for a concave surface [14].

The analysis presented here has shown that a Rayleigh wave can be represented as a combination of two inhomogeneous waves, one longitudinal, and one transverse. This combination acts in such a way as to cancel the stresses on the surface.

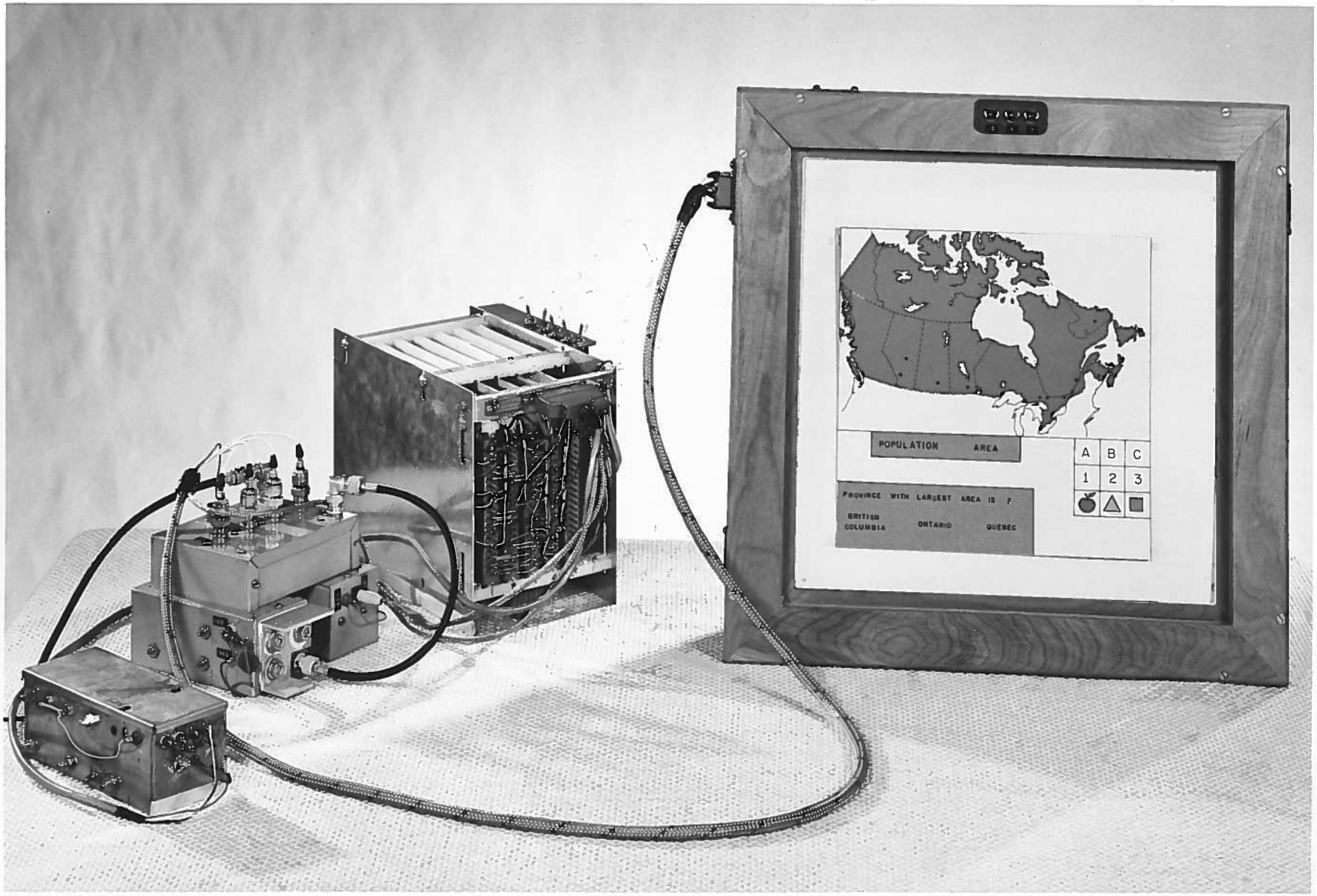


Plate I Touch-sensitive x–y position encoder