

THE BELL SYSTEM TECHNICAL JOURNAL

Volume 47

February 1968

Number 2

Copyright © 1968, American Telephone and Telegraph Company

Optical Heterodyne Experiments with Enclosed Transmission Paths

By O. E. DeLANGE and A. F. DIETRICH

(Manuscript received September 13, 1967)

Experiments to determine the feasibility of using optical heterodyne receivers at the end of long and complicated transmission paths showed that nearly ideal performance is possible. The paths were enclosed, thus avoiding atmospheric turbulence. We set up two receivers, one coherent and the other noncoherent, and compared their signal-to-noise performances and stability. The degree of beam distortion resulting from transmission over a path was determined by comparing the path losses as measured by the two receivers. Heterodyne detection was nearly 80 per cent efficient for a four-mile path involving 63 reflections and transmission through a considerable number of other optical components. Stable heterodyne operation was observed.

I. INTRODUCTION

It is well known that efficient operation of a light heterodyne receiver requires a high degree of phase coherence between the received signal beam and the local oscillator beam.¹ For example, if the angle between the received beam and the local beam is in error by as little as one second of arc there can be considerable loss of signal. Also it is obvious that if the beam were translated by a beam width the detection efficiency would be reduced to zero. These requirements are so critical that many declared successful operation of a heterodyne detector at the end of a long transmission line to be very doubtful.

Experiments on the transmission of light beams through the open

atmosphere have shown that atmospheric turbulence might render the beam unsuitable for heterodyne detection.² However, the medium which we visualize for long-distance optical transmission consists of a series of lenses or mirrors for guiding the beam through an enclosure which isolates it from its environment. It is quite likely that successful guided transmission will require some type of beam position servo control in some of the guiding elements. If we provide this control, the position of the received beam can be stabilized so that an efficient heterodyning adjustment, once made, will be maintained. Stabilizing the position of the received beam automatically stabilizes its angle of arrival. During some previous experiments³ we observed that single-mode transmission appeared to be preserved even after many reflections. Because of these favorable considerations, we decided to set up an experimental heterodyne receiver at the end of the longest transmission path available and observe its operation.

II. THE EXPERIMENTS

2.1 *Transmission Path*

The path consisted of a horizontal aluminum tube six inches in diameter and 100 meters long. We put eight concave mirrors, one inch in diameter, at one end of this pipe and seven at the opposite end to transmit a laser beam for eight round trips through the line. The total distance was one mile through enclosed atmosphere.

2.2 *Experimental Arrangement*

Fig. 1 shows the experimental arrangement in block form. The laser, which operated at 0.63μ , was in a cavity only 20 cm long. This means it could oscillate at only one frequency as long as it was kept tuned near the center of the doppler line. Temperature stabilization of the cavity was sufficient to assure single-frequency operation without automatic frequency control. Power output was approximately one milliwatt.

The laser output was amplitude-modulated a few percent by applying an 84 KHz sine wave to a modulator which consisted of a suitably-oriented rod of KDP followed by an analyzer. The combination of a lens and a concave mirror constituted a telescope for adjusting the beam diameter and radius of curvature properly for launching it into the pipe line. The flat mirror between the lens and the curved mirror served only to fold the path back on itself. See Fig. 1. The tilt

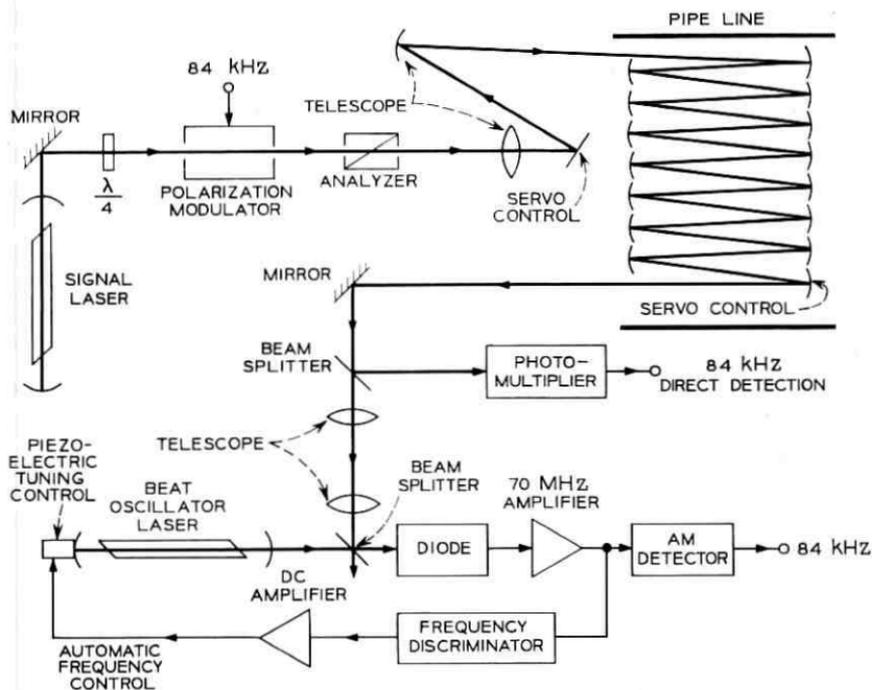


Fig. 1 — Experimental arrangement, one-mile path.

of the flat mirror was controlled by voltages applied to two piezoelectric cylinders. This was part of a servo system used to assure that the beam was properly directed to arrive at the center of the mirror at the far end of the line.

After eight round trips through the line the emerging beam, whose position was also servo-controlled, was directed along two separate paths by a beam splitter. One path included a photomultiplier functioning as a direct detector for recovering the 84 KHz modulation. The second included a telescope for reducing the beam diameter and converting it to a nearly plane wave. This wave was combined with the beam from the local laser on a second beam splitter. One output from this beam splitter was focused onto a photodiode detector. A 70 MHz intermediate frequency was obtained from this detector as a result of the beating of the two signals. The amplified output of the 70 MHz receiver was applied to a diode detector from which the 84 KHz modulation was recovered. A 70 MHz frequency discriminator furnished an error voltage for automatic frequency control of the

local oscillator laser. The amplified error voltage was fed back to a piezoelectric mirror mount in such a way as to maintain the frequency difference between the two lasers at 70 MHz.

The beam emerging from the line was centered between four sensors as shown on Fig. 2. The error voltage from these sensors was amplified and applied to piezoelectric cylinders which controlled the tilt of the last mirror at the far end of the line and thus kept the received beam centered. The remaining mirrors in the line were not servo controlled but had manual electrical tilt adjustments.

The direct detector consisted of an RCA 7326 photo-multiplier tube with a circuit in its anode tuned to pass 84 KHz. Because of the gain in the multiplier section of this tube it was possible to make the signal and shot noise power so much greater than thermal noise that the thermal noise could be neglected. The cathode of this tube is large enough to make signal output largely independent of beam position for moderate amounts of beam translation. The heterodyne detector was a Texas Instrument LSX 900 photodiode. A transformer tuned to a center frequency of 70 MHz coupled this diode to the input of a low-noise preamplifier. The transformer and amplifier had

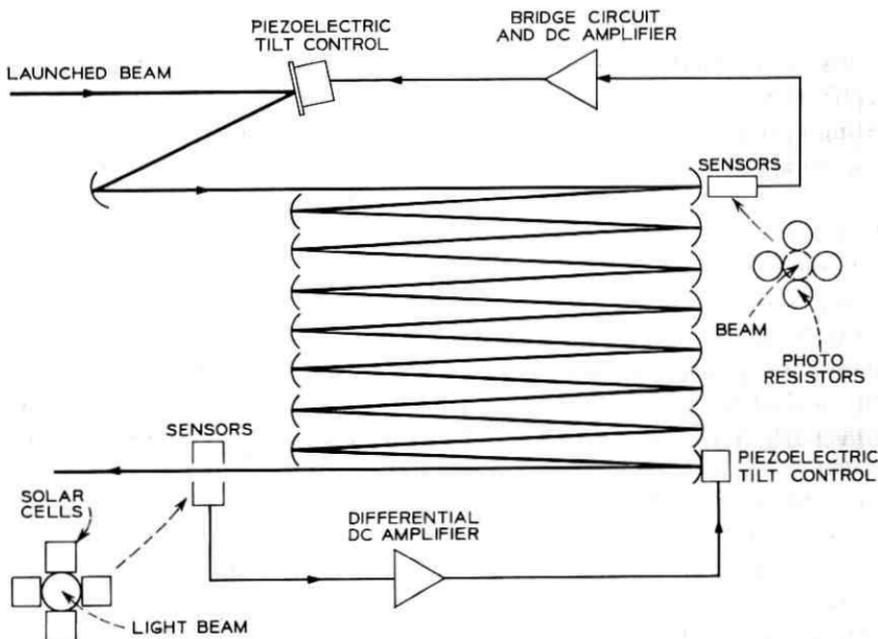


Fig. 2 — Servo-controlled beam positioners.

a bandwidth of approximately 20 MHz. The active area of the diode was only 0.01 inches in diameter which meant that the beam had to be very accurately centered on it for maximum response.

2.3 *Determining Heterodyne Efficiency*

The basic plan for the experiment was to measure the loss introduced by the line as indicated by the heterodyne receiver and to compare this with the line loss indicated by the direct detector. If the transmission loss indicated by the heterodyne receiver is greater than the loss measured by a direct detector the difference can be attributed to a decrease of detection efficiency for the heterodyne receiver resulting from distortion of the wavefront of the beam by transmission through the line. Such distortions should not affect the direct detector.

This can be stated differently by noticing that transmission over the long path results in some of the light power being converted into higher order modes by any existing line imperfections. That power converted to very high order modes cannot propagate in the line and is lost. Some of the lower order modes can propagate and will be detected by the direct detectors, whereas the heterodyne detector can respond to only one mode so that any power not in that mode is lost as far as this detector is concerned.

If transmission of the beam through the required external optical components produced wavefront distortion before the beam was launched into the line, the distortions introduced by the line might be masked. Because of this we need to determine the absolute detection efficiency (at least with respect to the beam as it is launched into the line). Detection efficiency is defined as the ratio of actual, or measured, beat-note current to the current calculated for an ideal detector. This efficiency can be measured in several ways. First if the signal power, P_s , the quantum efficiency, η , and the bandwidth B are known, we can calculate the expected carrier-to-noise ratio from the relationship $C/N = \eta P_s / hfB$ where h is Planck's constant and f the light frequency. By comparing the measured C/N to the calculated value we can determine detection efficiency.

The C/N for the heterodyne receiver was measured by means of a detector at the IF amplifier output. With a light signal into the receiver the IF gain was adjusted to produce some value of detected current to serve as a reference. The signal was then removed and the IF gain increased until the detector produced the same amount

of current, this time resulting from the noise. This change of gain was equal to the C/N. Because the equation applied only to shot noise, it was necessary to correct for the presence of thermal noise. With the heterodyne gain obtained for this experiment shot noise and thermal noise powers were comparable.

The second method of determining efficiency was to compare the amplitude of a low-frequency beat note (produced by beating the unmodulated signal with the local oscillator) with the dc output of the diode detector. It can be shown that, ideally, when the oscillator power, P_o , is much greater than the light signal power, P_s , the ratio $I_B/I_{DC} = 2(P_s/P_o)^{1/2}$. I_B is the peak current at the beat note frequency. The ratio of measured I_B to calculated I_B is the detection efficiency.

The beat note method consistently indicated a higher efficiency than the C/N method. Unaccounted-for noise or the cumulative effect of errors in determining the light-signal power, quantum efficiency, and bandwidth, could account for the pessimistic C/N measurements.

We also made some optical tests to determine the heterodyne efficiency. Because heterodyne deficiencies result from lack of phase coherence between the signal and local oscillator beams, it is evident that restricting the area of the superimposed beams to a small region near their centers improves coherence. And if this area is made sufficiently small, perfect coherence must result. Hence, if the combined beams are passed through a small iris before being applied to the detector, a reference value corresponding to 100 per cent efficiency should be obtained. If perfect coherence exists over the total area of the beams, increasing the diameter of the iris will not affect the efficiency and the manner in which the IF output increases with iris area is predictable. If perfect coherence does not exist, increasing the iris diameter will reduce the detection efficiency and the IF output from the detector will not increase as much as expected.

For our experiment we used detected direct current as a criterion of iris area. Direct current from the detector increases linearly with combined local oscillator and signal power. The intermediate frequency power is proportional to the product of the two light powers and should vary as the square of the direct current. Fig. 3 is a plot of the results of varying the iris diameter in such an experiment. The dashed line shows the ideal variation of IF power with detected direct current, that is, the variation which should be expected for perfect phase coherence between the two beams.

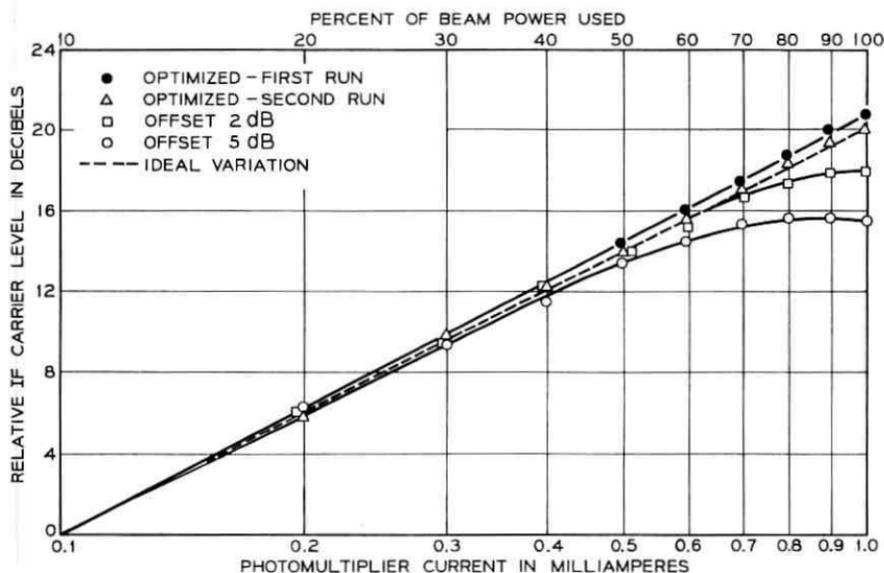


Fig. 3—Phase coherence tests comparing 2 and 5 dB offsets.

The solid curve with dots shows actual performance after the detector had been adjusted to produce maximum IF power. The performance is close to ideal. Next, the received beam was tilted from the local oscillator beam to reduce IF Power by 5 dB with the iris fully open. The curve with circles shows how IF power varied with diaphragm opening. The effect of lack of coherence is quite evident from the saturation of IF carrier as the iris opening was increased. This test was repeated with a 2 dB offset, corresponding to 80 per cent efficiency. The curve with squares indicates the results.

III. EXPERIMENTAL RESULTS

3.1 One Round Trip

To evaluate the effect of components outside the line, we first measured the detection efficiency for only one round trip through the line. For this arrangement the signal beam traversed one KDP modulator, one Glan-Thompson prism, one short focal length lens, two spherical mirrors, two flat mirrors, two quarter-wave plates and two beam splitters. All of these components were stock items with no attempt to obtain exceptional quality. The C/N method indicated a detection efficiency of 81 per cent, the beat note method 98 per cent; the true value is probably somewhat between. Evidently there

was no serious deterioration of performance even though the beam returned from the line was noticeably elliptical in contrast with the circular beam from the beating oscillator.

3.2 One-Mile Line-Losses

By adding 14 more spherical mirrors to the optical path mentioned above the path length was extended to eight round trips for a total distance of one mile. The efficiency, determined from a number of measurements of C/N, averaged 81 per cent; the beat note method averaged 87 per cent. These data indicate that addition of the seven round trips to the optical path did not appreciably reduce the detection efficiency. This is borne out by the results of another series of experiments in which the line losses, as indicated by the heterodyne receiver, were found to be the same as the losses measured with the direct detector. Table I summarizes the results of this experiment.

Optical tests of coherence, as we described, were made by passing the combined signal and local oscillator beams through an adjustable iris. Fig. 4 compares the coherence performance for one and for eight round trips through the line. The results agree with those of the other tests in indicating no measurable loss of coherence resulting from the additional 14 reflections. Within the accuracy of measurement it appears that the external optical components produce more beam distortion than the one-mile line.

3.3 Four-Mile Line Configuration

The round trips through the line were increased by two optical circulators to 32 (63 reflections) for a total length of four miles. Figure 5 shows the experimental setup. Light from the signal laser was first circularly polarized by a quarter-wave plate and then transmitted through a KDP rod where it was modulated at 84 KHz. One component of the modulator output was transmitted through the first analyzer prism, the other absorbed.

TABLE I—One-Mile-Line

Path (round trips)	Indicated Detection Efficiency (%)			Indicated line Loss (dB)	
	C/N	Beat	Variable Iris	Heterodyne	Direct Det.
1	81	98	100	—	—
8	81	87	112	0.9	0.95

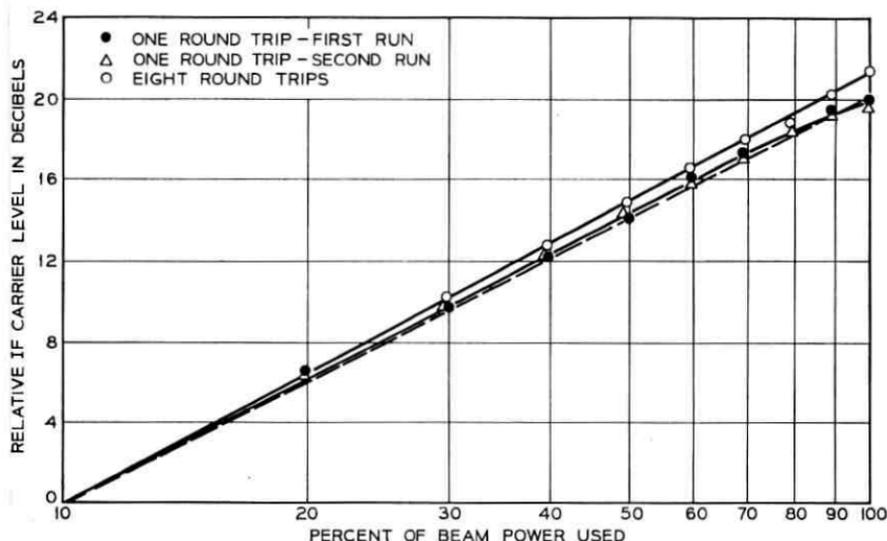


Fig. 4 — Phase coherence test comparing one and eight round trips.

The field of the Faraday-rotation isolator was adjusted to produce a 45° rotation of the plane of polarization of the amplitude-modulated output of the prism. A second analyzer prism, at the output of the isolator, was set to pass this polarization into the line. A quarter-wave plate in the line converted the input to circular polarization. After traversing the line for eight round trips the beam was reflected directly back on itself and returned to the quarter-wave plate. It emerged from the plate linearly polarized in a plane normal to that at which it first entered the line.

Analyzer prism 2 now directed this beam out of one of its side ports where a flat mirror reflected the beam directly back on itself and thereby launched it into the line for a second time. After 16 more round trips through the line the beam returned again to the quarter-wave plate. Upon emerging from the plate this time it had been restored to its original polarization and so was transmitted directly through prism 2 into the isolator. The additional 45° rotation imparted by the isolator caused its plane of polarization to be 90° from that of the original transmission. Upon leaving a side arm of analyzer prism 1, the beam was directed to the heterodyne receiver by a flat mirror and two beam splitters. One beam from the first beam splitter was applied to a photomultiplier direct detector where the 84 KHz modulation was recovered. The other was combined

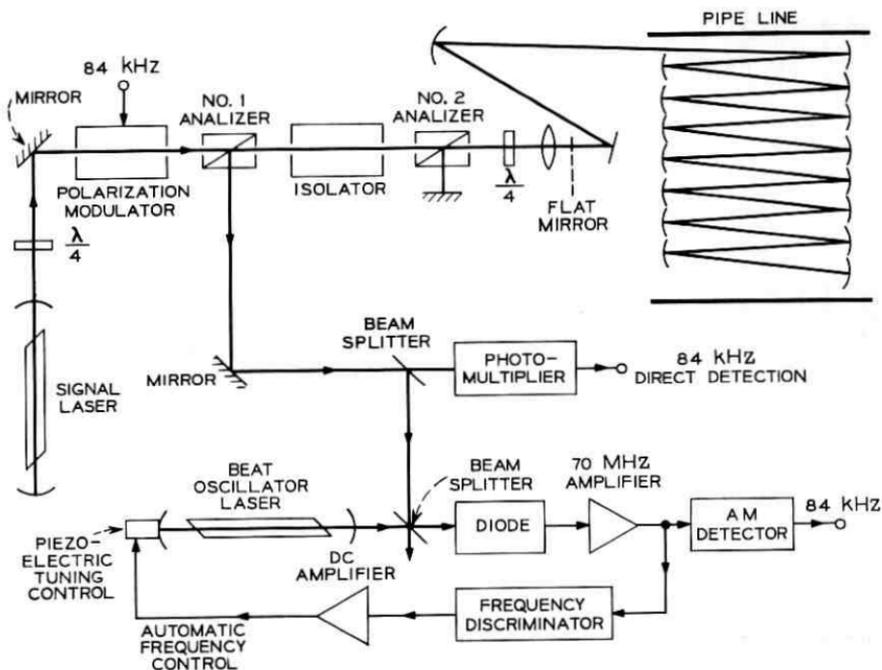


Fig. 5 — Experimental arrangements, four-mile path.

with the output of the local oscillator and applied to a photodiode detector. The 70 MHz output of the photodiode was amplified and detected by an electronic detector to recover the 84 KHz modulation.

The arrangement shown in Fig. 5 made it simple to determine the losses introduced by the four-mile line. By placing a flat mirror, shown dashed, at the focal point of the lens, light was caused to traverse the circulators and other components in the usual manner but transmission through the line was eliminated. Thus the line loss was the difference between the loss when light passed into the line and the loss when the line was blocked by the flat mirror.

To determine the effects of local optical components, the absolute detection efficiency was determined with only these components in the transmission path, that is, with the line blocked by the flat mirror. The average of a number of C/N measurements indicated a detection efficiency of 86 per cent. The beat method indicated 93 per cent.

3.4 Four-Mile-Line Losses

Table II summarizes the results of this experiment. The efficiencies and line losses that the table shows are averages obtained from a

number of measurements. Fig. 6 shows the variable-iris data. Adding the line decreased the detection efficiency from 86 to 78 per cent, as indicated by the change of C/N. This is consistent with the line loss measurements and with the variable-iris check of coherence. The two curves of Fig. 6 show some loss of coherence after transmission over the longer path.

From the results of these experiments one might conclude that the four-mile line losses for a heterodyne receiver were about one decibel greater than for a direct detector. However, our measurements were not accurate enough to make this a reliable figure. It is likely to be somewhat pessimistic. In any case, it appears that heterodyne detection efficiencies near 80 per cent are obtainable even with light signals which have been transmitted through this long and complicated optical system.

We used no really elaborate beam-alignment or beam forming procedures for these experiments. We made the signal beam approximately plane-parallel by means of a telescope. The local oscillator beam remained slightly divergent, just as it emerged from the laser. Beam positions and angles were simply adjusted to produce maximum IF output from the detector. Although most of the optical components were stock items, the mirrors were rated good to 1/50 wavelength by the manufacturers, Laboratory Optical Co.

IV. STABILITY

Once it is determined possible to operate a light heterodyne receiver at the end of a long transmission line efficiently, the next question is whether this efficiency can be maintained for long. Our experiments indicate that, with reasonable care, it can.

There are two types of instability which can cause the system to malfunction: frequency variation and beam wander resulting from

TABLE II — Four-Mile-Line

Path	Indicated Detection efficiency (%)			Indicated line Loss (dB)	
	C/N	Beat	Variable Iris	Heterodyne	Direct Det.
Local Components Only	86	93	100	—	—
Local Components Plus 4-Mile Line	78		87	7.5	6.6

mirror tilt. Rough temperature compensation in the laser cavities was sufficient to keep our transmitting oscillator operating near the center of its doppler line. Automatic frequency control kept the average frequency of the local oscillator differing from that of the receiver by 70 MHz. The control voltage for this AFC came from a frequency discriminator operating at the intermediate frequency. Acoustical vibrations acting on the two lasers produced rapid deviations of the intermediate frequency obtained from the detector. These deviations, with peak values of a few megahertz, were passed without loss by the IF amplifier which had a bandwidth of 20 MHz. We have since practically eliminated these frequency shifts by using a different type mirror mount in the laser cavities.

Because all of our components were mounted on high temperature coefficient materials such as aluminum, brass, and steel there were some deviations in the position of the transmitted and received beams resulting from tilts produced by temperature variations. Our line, which was made up of single mirrors, was very much less stable than a practical line, which would have lenses or pairs of mirrors. Also, our experimental line was probably subjected to greater temperature changes than an underground line would be.⁴ In any case it is evident that some form of beam position servo control will be necessary in a long transmission line.

4.1 One-Mile Line

For this experiment both transmitting and receiving equipment were mounted on the same steel plate at one end of the line. The steel

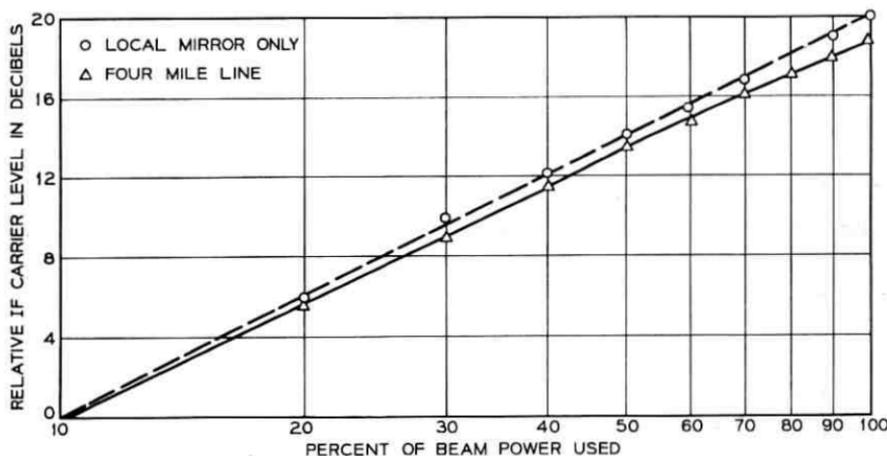


Fig. 6—Phase coherence test comparing local and four-mile paths.

plate was on a wooden table which stood on a concrete floor. The presence of active components, such as the lasers, on this plate resulted in rather large point-to-point differences in temperature. As a result, this plate was the most unstable part of the transmission path and, if uncontrolled, the beam received at the far end of the path wandered about, sometimes going off the receiving mirror.

To stabilize the projected beam, four light sensors were installed behind the first mirror at the far end of the line. These sensors gave error signals which were amplified and applied to piezoelectric cylinders which in turn controlled the tilt of one of the launching mirrors. See Fig. 2. This system improved stability considerably.

A similar servo system was set up to control the position of the received beam to assure proper alignment of signal and local oscillator beams. The sensors were at the exit port of the line and the error voltages controlled the tilt of the last mirror at the far end of the line. With such an arrangement it should be possible to maintain the position of the received beam within a few thousandths of an inch—a fraction of a second of arc in our line. The remaining mirrors in the line had electrical tilt controls but these were all adjusted manually.

We connected a number of monitors to the system to study its stability. The outputs of these monitors were continuously recorded. Some of the quantities recorded were: (i) Vertical and horizontal error voltages from the output beam sensors. These voltages indicated the position of the received beam. (ii) The 84-KHz signal recovered by the heterodyne receiver. This indicated any fading of heterodyne signal which might have resulted from such factors as beam wander. (iii). The 84-KHz signal recovered from the direct detector. By comparing (ii) and (iii) we can measure any excess loss resulting from decrease of detection efficiency in the heterodyne receiver. (iv) Transmitted signal power. (v) Local oscillator power. (vi) Automatic frequency control error voltage.

From the recordings and from direct observation it was evident that the heterodyne receiver was noticeably less stable than the direct detector. However, the fluctuations of output of the coherent detector were usually less than ± 0.5 dB and occurred at low rates which could easily be removed by an AGC circuit. We should emphasize that the heterodyne detector diode was only 0.01 inch in diameter whereas the direct detector cathode was over an inch in diameter and could accommodate more movement of the received beam. Except for the small fluctuations just mentioned, the heterodyne receiver operated as well as the direct detector. The link operated all day, and at times

for several days, without any readjustment and without serious loss of signal. Most times when signal loss did occur it was the result of detuning. A better AFC circuit should take care of this.

The charts in Fig. 7 indicate the stability during a typical six-hour period. The important consideration here is the comparison between the signals recovered by the direct detector and by the heterodyne receiver. Both follow the same general trend and both are completely free of any serious fades.

At point A, near the end of the chart, the servo loop which controlled the position of the final signal beam was opened. As a result, the beam position shifted enough to produce a 1 dB drop in the signal

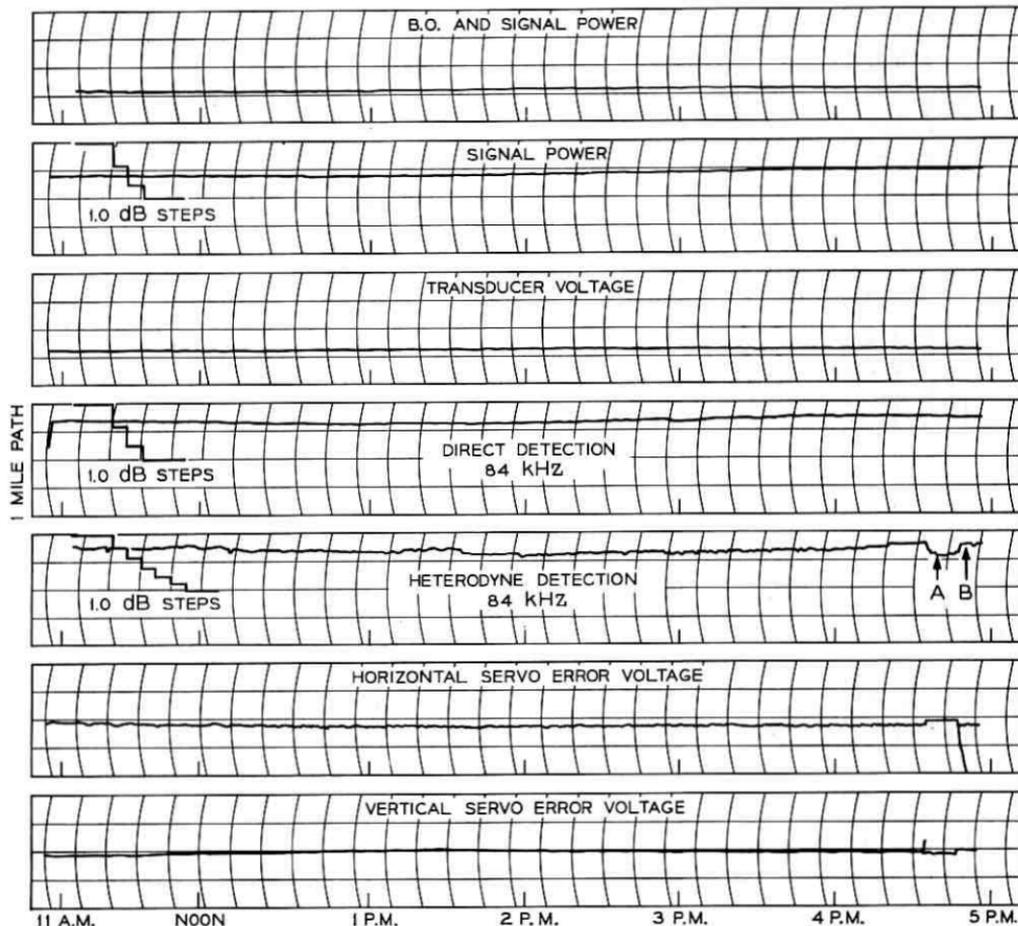


Fig. 7 — Signal received over the one-mile path.

from the heterodyne receiver. The output of the direct detector changed much less. At point B, the servo control had been restored and the signals returned to their former levels.

That the servo control system can be simple is evident from Fig. 8 which shows schematically the control circuit for one coordinate of one mirror. The components consist of two photoconductor cells, two vacuum tubes and two piezoelectric cylinders. The power requirements are 300 volts at a few milliamperes. With the vacuum tube amplifier it would also be necessary to supply heater power. When field-effect transistors capable of handling 300 V become available it should be possible to substitute them for the vacuum tubes.

4.2 Four-Mile Line

Because of the signal reflexing necessary for this experiment it was not possible to control the beam position by servo systems. In spite of this the configuration was quite stable. Fluctuations of signal level from the heterodyne receiver were only slightly greater than for the one-mile transmission but were considerably more rapid. These fluctuations, which had periods of several seconds, appeared to result from reflection of energy from the line back into the laser.

The isolator, supplied by C. B. Rubinstein of Bell Laboratories provided 40 dB of isolation. But because of resonances at very high

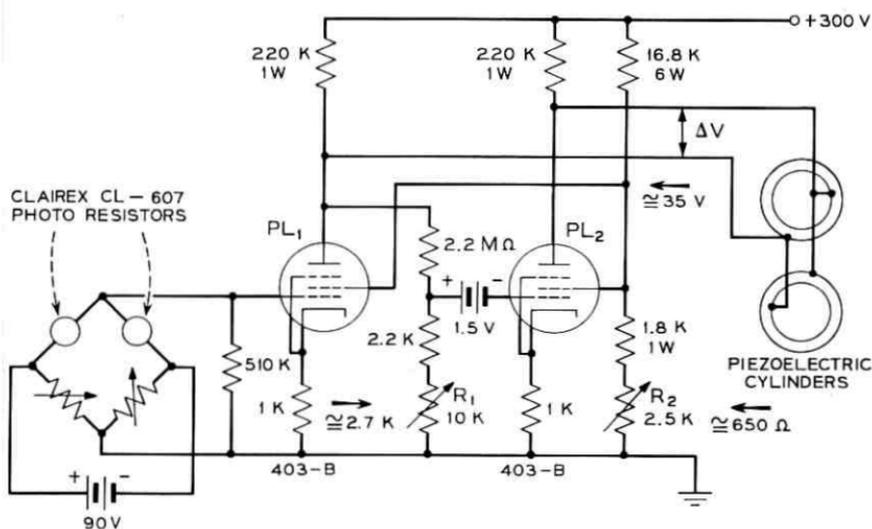


Fig. 8 — Automatic beam positioning control equipment.

Q the light power could build up in the line, and 40 dB apparently was not enough. Tests indicated that these fluctuations resulted from very small changes of length of the transmission path. When part of the path was through open air where the beam could be affected by air currents the fluctuations were considerably greater.

The charts in Fig. 9 show typical performance of the four-mile path for eight hours. The output of the heterodyne receiver decreased about 1 dB during this time. At the same time the output of the direct detector increased almost 1 dB as a result of increased power from the transmitting laser. This would indicate that the heterodyne receiver faded by almost 2 dB during this interval. If servo control of beam position had been possible it probably would have prevented this signal loss.

We do not wish to give the impression that the stability and reliability of our experimental setup did more than approach that required for a practical operating system. However, our observations lead us to believe that the required stability can be attained at reasonable expense by well known methods. To begin with, a practical line would have lenses or pairs of mirrors as beam directors which are much less sensitive to mechanical displacement than the single mirrors of our line. The practical line would also be underground where it would be much less subject to temperature variations.⁴ Having the beam-director mounts made of low-temperature-coefficient material would help a great deal. Temperature control of the director environment could add stability but should not be necessary if beam position is controlled by servos.

Temperature control of the transmitting and local-oscillator lasers, along with the best AFC circuit should solve the frequency stability problem. Laser tube life would need to be extended.

V. CONCLUSIONS

We found that reduction in received signal caused by loss of coherence which resulted from transmission through a complicated system (including as many as 70 reflections) was small, comparable with the measuring errors. The losses in a four-mile transmission path, as measured by a heterodyne receiver, were about 1 dB greater than the losses in the same path measured by a direct detector.

We also found that even without complicated controls the system was sufficiently stable to maintain near-optimum operation for fairly long periods of time. However, our observations indicate that for long-

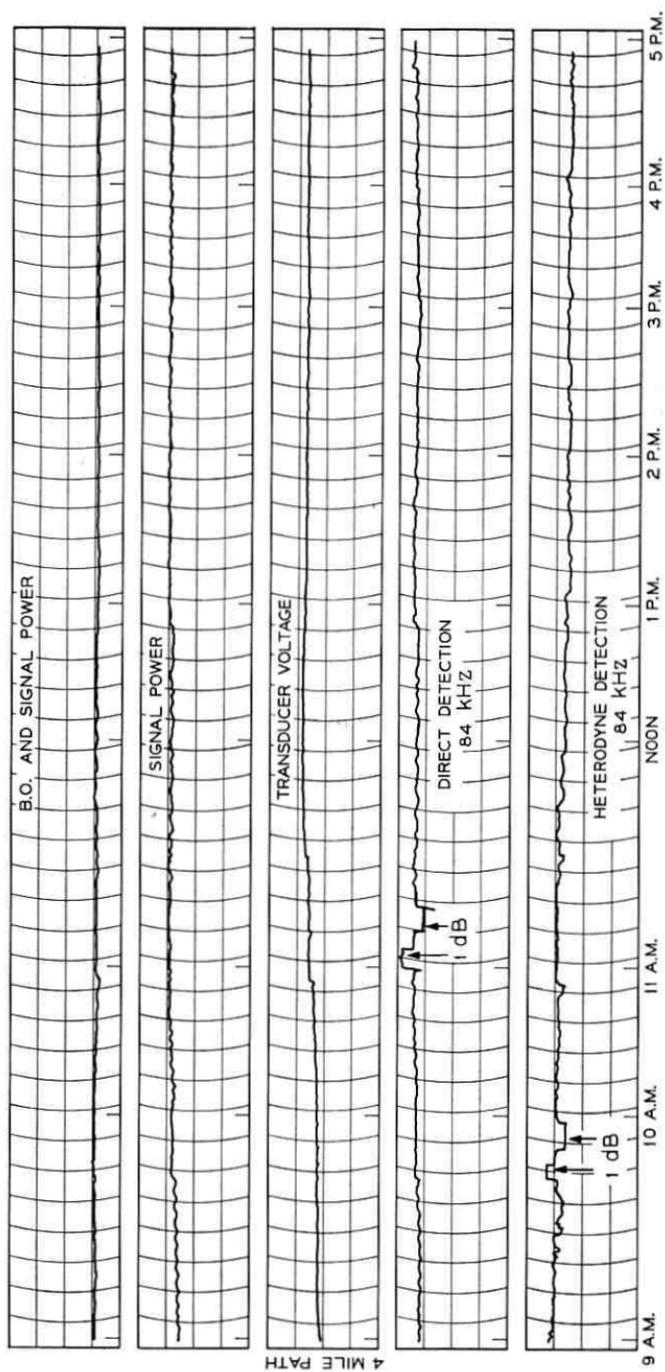


Fig. 9—Signal received over the four-mile path.

distance systems which must operate for extended periods with great reliability it would be necessary to use more sophisticated control circuits. Stability, aside from frequency control, is simply a matter of proper beam alignment. This alignment is necessary to obtain reliable transmission through the line so that the heterodyne receiver requirements are little more severe than the requirements for any other type of receiver.

Although most likely there would be some improvement in efficiency through using better optical components, the stock components that we used gave quite satisfactory performance.

It is interesting that even though it is visually evident that the mirrors we used had deteriorated and accumulated dust over a period of years, the line loss measured now is 0.85 dB per mile in comparison with an average of 0.9 dB per mile measured by a different method when the mirrors were new. That is, any change of line loss is less than the error of measurement.

REFERENCES

1. Siegman, A. E., The Antenna Properties of Optical Heterodyne Receivers, *Proc. IEEE*, *54*, October 1966, pp. 1350-1356.
2. Goldstein, I., Miles, P. A., and Chabot, A., Heterodyne Measurements of Light Propagation through Atmospheric Turbulence, *Proc. IEEE*, *53*, September 1965, pp. 1172-1180.
3. DeLange, O. E., Losses Suffered by Coherent Light Redirected and Refocused Many Times in an Enclosed Medium, *B.S.T.J.*, *44*, No. 2, February 1965, pp. 283-302.
4. Gloge, D., Experiments with an Underground Lens Waveguide, *B.S.T.J.*, *46*, April 1967, pp. 721-735.

Measured Beam Deformations in a Guide Made of Tubular Gas Lenses

By P. KAISER

(Manuscript received September 21, 1967)

We discuss problems connected with the alignment of gas lenses and present experimental results showing the effects of off-axis injection of a light beam into a beam waveguide (offset $a/r \leq 0.315$, r = radius of tubular lens) and of the introduction of various radii of curvatures ($R \geq 500$ m) on the shape and position of the transmitted Gaussian beam. The theoretical prediction of reduced beam distortions for nonconfocal geometry was confirmed experimentally. We demonstrated low losses at small bending radii by using 12 lenses as focusing elements within a laser cavity in the bent state.

I. INTRODUCTION

Tubular gas lenses have possible use as focusing elements in an optical beam waveguide.^{1, 2} Theoretical analysis has shown that thermal gas lenses are not ideal; beam distortion from the cumulative effect of lens aberrations²⁻⁶ and the beam wandering resulting from lens misalignments^{7, 8} limit the number of lenses that can be used in a beam waveguide. Beyond this number, corrective devices such as redirectors, refocusers, and mode filters must be used.^{9, 10} Redirectors sense the position of the beam and introduce deflections which tend to realign the beam with the guide axis. Refocusers sense the increase of beamwidth and tend to focus the beam back to its ideal profile. Mode filters reduce the content of unwanted higher order modes.

Previous experiments indicated difficulties with the accurate alignment of the lenses.¹¹ From this experience we worked out a method of alignment by which we achieved satisfactory accuracy and beam stability. Subsequently we measured the deformation of the beam transmitted off axis and around a bend and compared it with theory. We compared beam distortions for nonconfocal lens spacings with those of confocal geometry.

II. EXPERIMENTAL SETUP

Fig. 1 shows an experimental beam waveguide made of 10 gas lenses. Each lens was suspended within a 12-meter-long steel U-channel, and kept in place and aligned by two sets of screws as Fig. 2 shows. Because the lenses were not rigidly connected among themselves except through the U-channel, each could be aligned without impairing the alignment of previously-mounted lenses. An air-tight interconnection between lenses was accomplished by rubber hoses. The gas lenses used are described in Ref. 11. The actual lens elements (6.34 mm I.D., 15 cm long) were mounted within a 5.08 cm diameter waveguide tubing ($\frac{1}{2}$ meter long), which was connected rigidly to a 5.08 cm diameter, 0.5 m long spacer.

Focal length measurements were performed for a single lens as a function of gas flow rate F and temperature rise ΔT between the input gas and the heated cylindrical wall of the lens. A collimated laser beam was focused through the lens. The beam size as a function of the distance d from the geometrical center of the lens was measured (Fig. 3) and graphically extrapolated back into the lens until it intersected the half-power beamwidth of the collimated beam. The relative position of the intersection was a function of F and ΔT : for low flow rates the intersection occurred before, for higher flow rates beyond the geometrical center. For constant gas flow rate an increase in temperature shifted the intersection toward the beginning of the lens. This behavior had been predicted by theory.^{3, 4} The position of the intersection was obtained with limited accuracy because it was not possible to follow the converging beam too far back into the lens. Hence, the focal length was determined from the distance d between the waist and the geometrical center of the lens. The average width

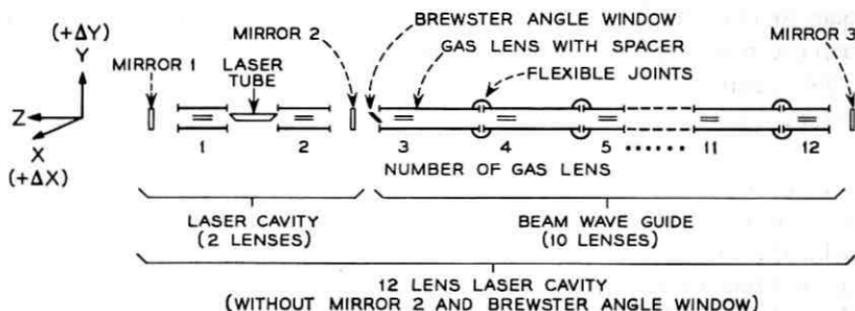


Fig. 1 — Experimental arrangement of beam waveguide.

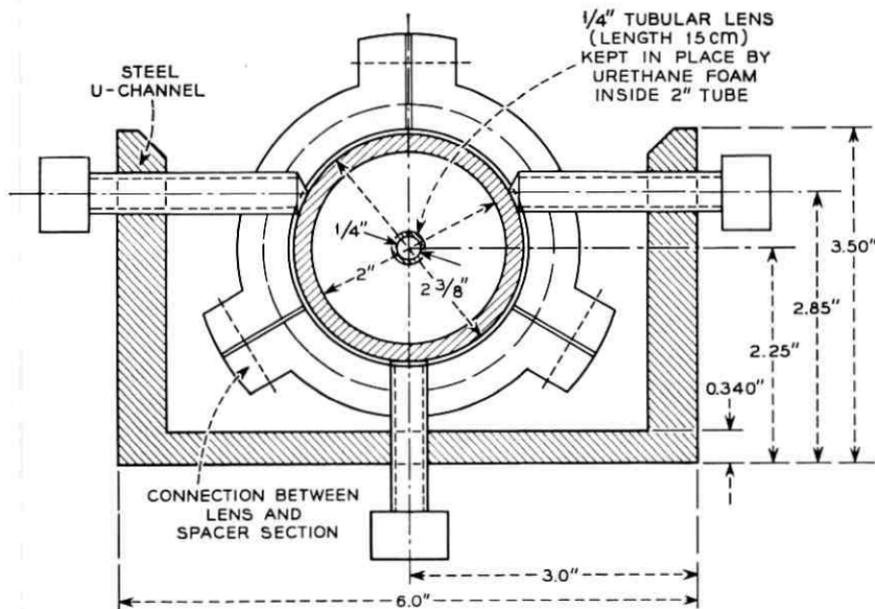


Fig. 2— Gas lens mounted inside U channel.

of the collimated beam was w and the width of the waist w_0 . From this the focal length was computed according to

$$f = \frac{d}{1 - (w_0/w)^2}$$

Fig. 4 shows the focal length as a function of temperature rise for different gas flow rates. For $F = 1.5$ L/min and $\Delta T = 103^\circ\text{C}$, the principal plane coincided with the geometrical center of the lens. We used the resulting focal length of 50 cm for the confocal operation of the beam waveguide. There was no noticeable difference between the focal lengths of the vertical and horizontal plane.

To guarantee the proper matching, we used two gas lenses as the focusing elements in the confocal laser cavity. At first the laser cavity was mounted on an optical bench, separate from the transmission line of 10 lenses. However, the change of the target spot position in time, probably caused by motion of the respective supports, was not tolerable. We achieved much better stability by mounting the launching system, together with the transmission line, into the U-channel itself.

In order to guarantee optimum alignment of the laser cavity, we

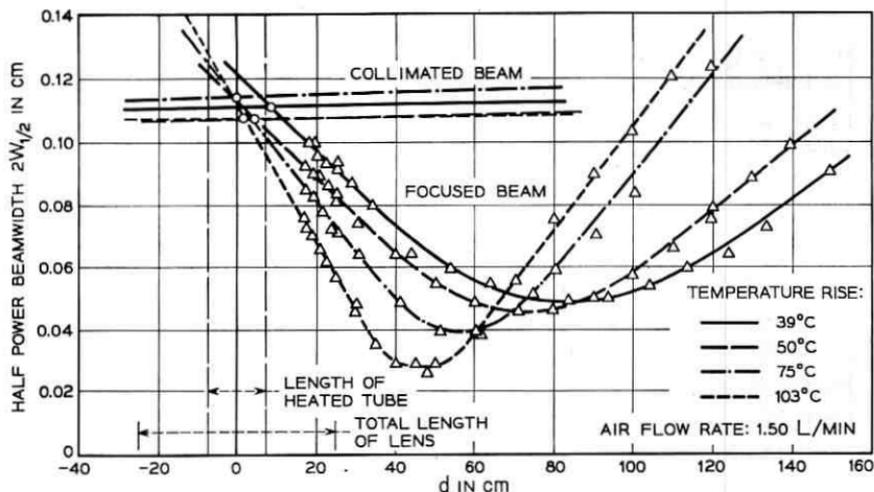


Fig. 3—Beamwidth of collimated and focused light beam as a function of distance from a tubular gas lens.

sent a reference beam through the channel on a predetermined path and with its help the optical axes of the two gas lenses of the laser were aligned. We placed the glass tube of the laser between the lenses and adjusted for minimum scattering of the passing reference beam. We placed a plane mirror (mirror 2 in Fig. 1) at the far end from the reference laser, so that the energy was coupled back into this laser and resonance could be observed by an increase in power output. The transmittance t of this mirror was 0.95 percent. After mounting a plane parallel mirror at the other end of the laser (mirror 1 with $t = 0.20$ percent), oscillations could be obtained easily and the reference beam was shut off. As intended, the aim of the assembled laser was close to the predetermined path, and undulations of the beam within the cavity were kept to a minimum. The first gas lens of the transmission line was closed on one side with a thin glass plate mounted at the Brewster angle to block the gas flow. It did not displace the beam noticeably, a fact which was important for the eventual removal of the plate in order to form a single laser cavity out of a larger number of gas lenses.

The accurate alignment of the lens required the knowledge of the beam position inside the lens. For a null indicator we used a probe consisting of a differential detecting device with 4 photocells arranged as quarter segments of a circle as Ref. 11 describes. In this way, we

avoided several difficulties such as calibration inaccuracies, dependence upon absolute intensity, and relative low sensitivity for small displacements of the beam from the probe's center. We performed the alignment as follows. We oriented the cold lens in such a way that the probe indicated null readings at each end of the tubular lens. Four external indicators marked the horizontal and vertical position of the lens within ± 1 mil. With the probe removed, we determined the position of the beam on a target 6 to 12 meters from the lens. The target was a micropositioner containing a photodiode and a metal plate with a 3-mil hole in front of it. The position of the beam maximum intensity could be defined to approximately ± 8 mils. Then we heated the lens to normal temperature. This deflected the beam downward, since the optical axis is below the geometrical axis of the tube because of the gravity effect. By means of the vertical indicators, we introduced a parallel displacement of the lens until the target spot coincided with the position before heating. By doing this, we made sure that the heating process of the lens did not deflect the beam. Henceforth, its trajectory coincided with the optical axis.

Because of symmetry, no horizontal adjustment of the lens should

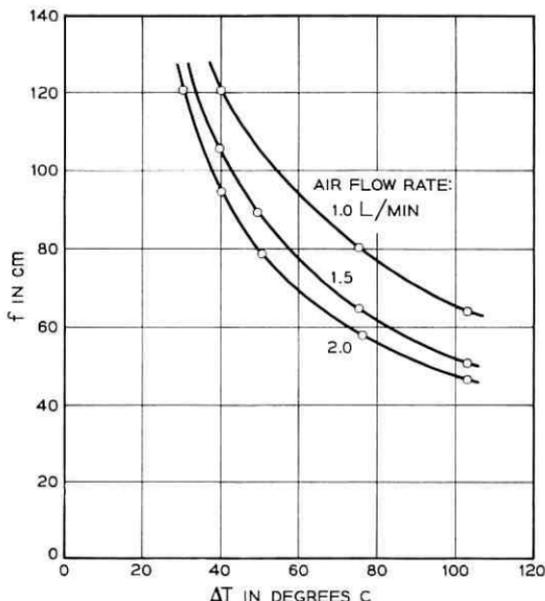


Fig. 4—Focal length of the tubular gas lens as function of temperature rise between wall and inlet air.

have been necessary. Even though this was true for most of the lenses, a substantial horizontal displacement had to be introduced in some of them, apparently to compensate for lens inaccuracies. The vertical displacement required for the compensation of the gravity effect varied between 13 and 23 mils, with values around 17 occurring most. The horizontal displacement necessary, if all at, amounted to a few mils and only in one case reached 12 mils.

One particular difficulty arose in carrying out this procedure. By inserting the probe, a back pressure was established and the reduced air flow resulted in higher temperatures in the preceding lenses. As a consequence, the ensuing variation of the optical axes of these lenses deflected the beam from its original target spot. We could make temperature variations negligible by fast probe insertion. We neglected the beam variation caused by reduced air flow and its influence on the focal length. Air conditioning machinery vibrated the building and caused random displacements and amplitude fluctuations in the laser beam. This was particularly detrimental for the centering procedure and prevented our using a more sensitive range in the probe's differential amplifier. The fluctuations disappeared after the motors were stopped for a short period. Taking the various factors into account, we estimated the alignment accuracy of the lenses' optical center to be within ± 3 mils, which is $\frac{1}{6}$ of the theoretical $1/e$ beam halfwidth $w (= 17.75$ mils) at each lens.

III. OFF-AXIS INJECTION

We injected the light beam off axis by passing it through a tilted plane parallel glass plate 83.5 mils thick. Figs. 5 through 8 show the power profiles of the beam transmitted through the guide on axis and with different horizontal and vertical offsets. We took about 25 points for each power profile and drew a smooth curve through these points as Fig. 5 shows. The target was 6.25 meters from the center of the last lens. For a beam injected on axis, the half-power beamwidth changed from day to day. It varied between 0.13 and 0.16 inch vertically, and between 0.15 and 0.17 inch horizontally. The center of the beam on the target also moved slowly a few tens of mils as a function of time. We found no satisfactory explanation for this behavior.

We normalized both the vertical and horizontal beamwidths for off-axis injection with the horizontal beamwidth of the on-axis beam. We obtained this reference width at the beginning of each series of measurements. We normalized the distance between the maximum

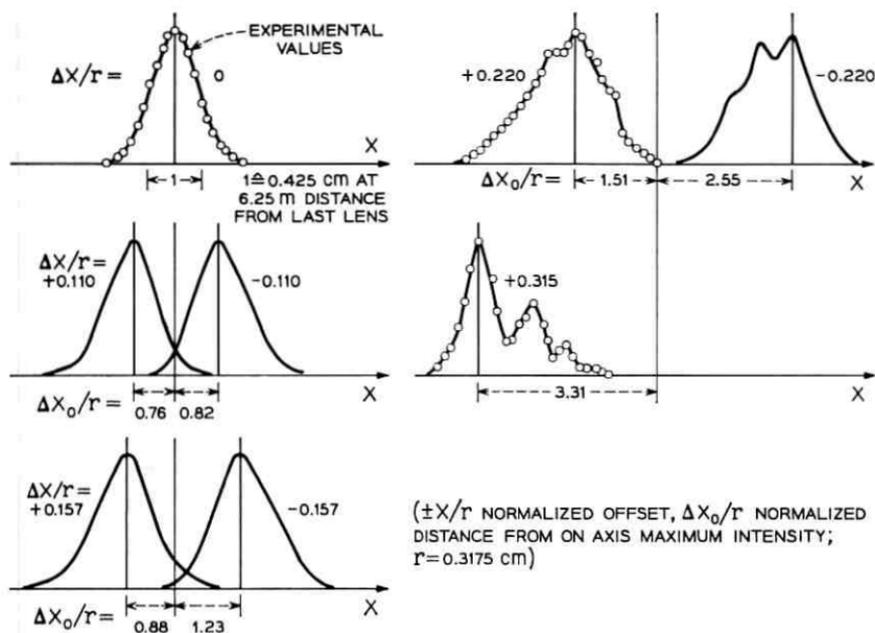


Fig. 5 — Horizontal power profile for horizontally-displaced beam.

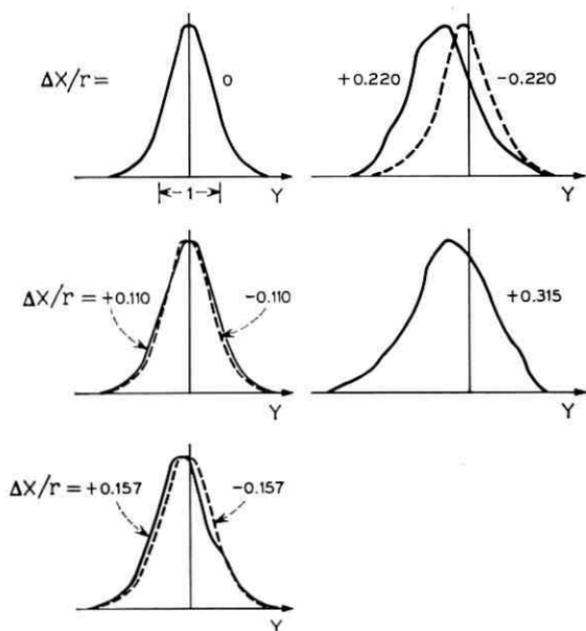


Fig. 6 — Vertical power profile for horizontally-displaced beam.

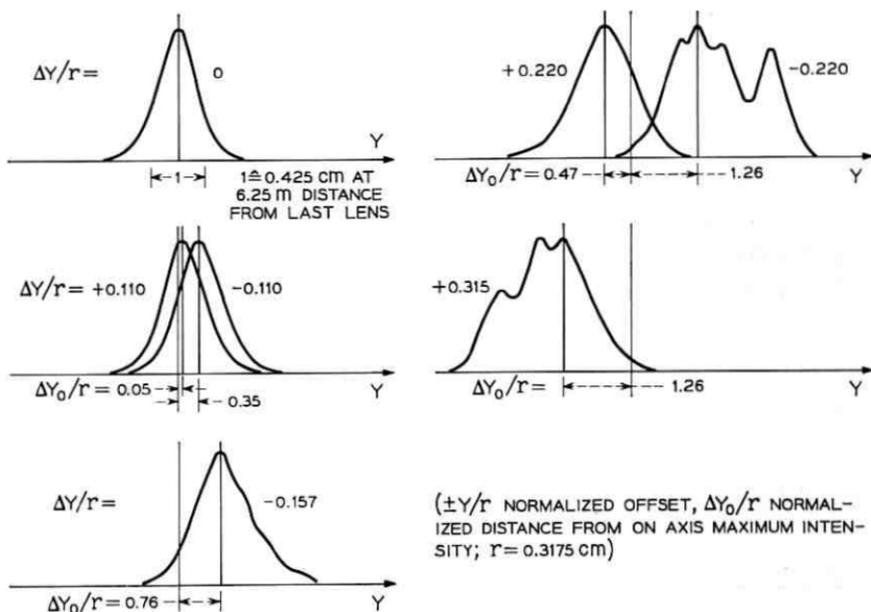


Fig. 7 — Vertical power profile for vertically-displaced beam.

intensities for off-axis and on-axis beams with the radius of the tubular lens and set the maximum intensities to unity. As could be expected from theoretical considerations, the deformations in the vertical and horizontal plane were independent of each other to the first order. However, we noticed a stronger deformation of the vertical beam shape in case of larger horizontal displacements, as compared with the horizontal beam shape for large vertical displacements (see Figs. 6 and 8). The most important result shown in Figs. 5 and 7 is the increase in beamwidth for growing offsets and eventual beam deformations for large offsets. Fig. 9 shows the increase in horizontal and vertical beamwidths relative to their respective values at zero offset. Points obtained for different experimental runs and, in general, resulting in different beamwidths, fitted well into the graph, once they were related to their respective horizontal zero offset widths. The uncertainty of the result owing to limited accuracy of the measurements and ambiguity of interpolation becomes larger for increasing beamwidths and is indicated in the graphs. The growth appears to be quadratic to the first order. Whereas this dependence is approximately equal for left, right, and high injected beams, the growth is larger for

the beam injected low. We found no satisfactory reason for this discrepancy.

The good symmetry in the position and the power profile of horizontally deflected beams appear to be the result of good alignment and horizontal symmetry of the lenses. The asymmetric deformations of the lenses owing to gravity apparently result in asymmetry of the position and profile for vertically-displaced beams.

For confocal geometry the normalized width w_n of a light beam which undulated with amplitude a about the axis of a sequence of n lenses is given by¹⁰

$$w_n = 1 + \frac{3}{2} \delta n \left(\frac{a}{w} \right)^2$$

where δ is a measure of the focal length distortion. At distance w from the center of the lens the relative increase of the focal length is δ . D. Marcuse's analysis of the tubular gas lens⁴ yields a value δ of approximately 0.012 for the geometry, gas flow rate, and temperature rise being considered. The theoretical prediction for the beam size increase

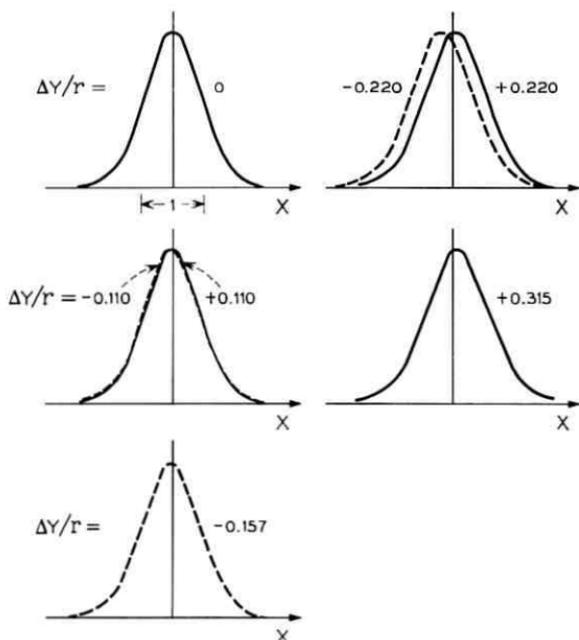


Fig. 8—Horizontal power profile for vertically-displaced beam.

according to above equation corresponds closely to the experimental results. (See Fig. 9.) Ten per cent beam size increase occurs after 10 lenses with a beam displacement a/w of only 0.7.

IV. BENT BEAM WAVEGUIDE

Because the lenses were mounted within a U channel it was easy to put bends of varying radii into the vertical plane. We lifted the channel to calculated heights for this purpose. The curvature started at the first lens of the optical guide (third lens in Fig. 1). The laser cavity with its two gas lenses remained straight. The U-channel was sufficiently elastic to permit following the desired curvature. The light beam entered the bend on the optical axis of the first guide lens. Figures 10 and 11 show the beam profile transmitted through the line for five different radii between 2500 and 500 meters. The half-power widths are normalized with 0.425 cm. No significant change of the Gaussian beam shape occurred down to a radius of 834 meters. A small hump appeared on one side of the beam at the 625 meter radius, but we measured no noticeable power loss. At 500 meters, however, the beam showed characteristic diffraction rings and a power loss of approximately 1 dB.

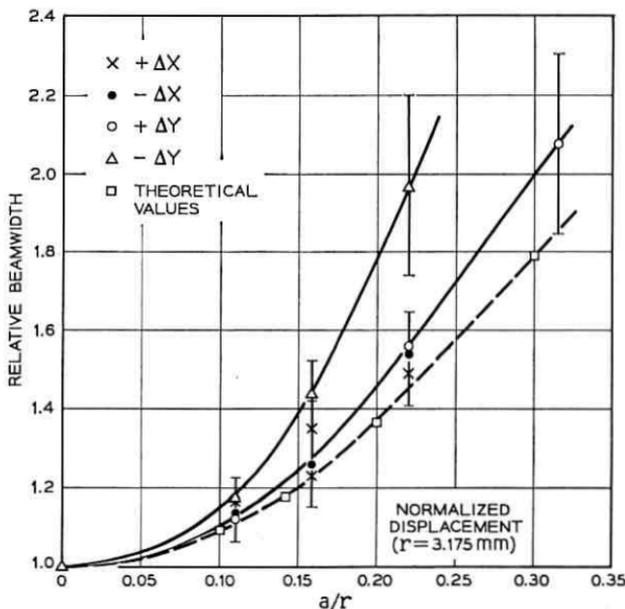


Fig. 9—Relative increase of beam width as a function of horizontal and vertical offset.

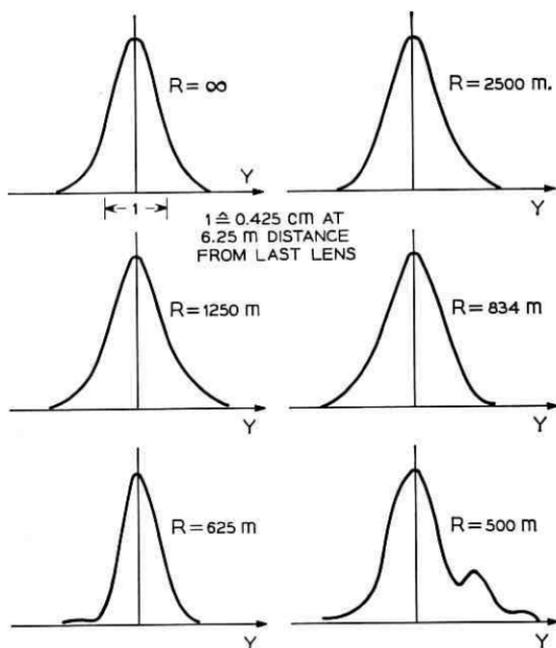


Fig. 10 — Vertical power profile as a function of radius of curvature.

According to the theory of a circular bend,¹² a beam injected off axis and with a particular tilt does not undulate, whereas on-axis injection causes undulations. We found that a small offset (without tilt) of approximately 0.35 mm for a 500-meter radius resulted in optimum transmission (power loss approximately 0.2 dB). For a parallel displacement without tilt a theoretical offset of 0.5 mm results in undulations around the ideal trajectory with minimum amplitude. As a consequence of this, the distortions are also minimized. Just as for vertical displacements, the horizontal shape did not change significantly for vertical bends.

In order to confirm the low-loss guidance properties of the 10-lens bent beam waveguide, we transformed the 2-lens laser cavity and the 10 guide lenses into a 12-lens laser, whereby the discharge tube of the 2-lens laser remained the only gain medium (see Fig. 1). Lasing was obtained at $R = 834$ meters with power output equal to that of the laser with only 2 lenses. The alignment procedure was as follows. With the 2-lens laser oscillating, we mounted a plane parallel mirror behind the 12th lens (mirror 3 in Fig. 1) and adjusted this mirror for resonance. Then we removed mirror 2 as well as the Brewster angle

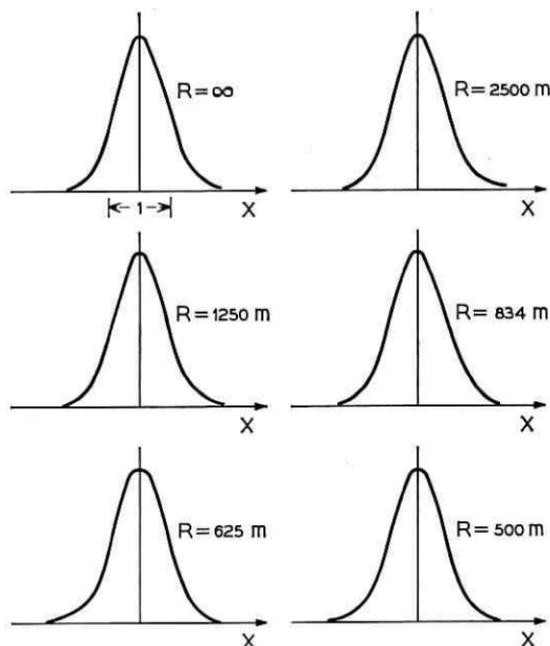


Fig. 11 — Horizontal power profile as a function of radius of curvature.

window between the laser and the guide and connected the two adjacent lenses by a hose. The 12-lens laser started lasing immediately. We replaced mirror 3 by mirror 2 in order to obtain identical operating conditions. Subsequently, the 12-lens laser was deformed into a circle with a radius of 625 meters (the first two lenses remained straight). Even though we did not obtain stable lasing, the proximity to oscillations existed; since small changes of air flow were enough to produce light flashes.

By changing the alignment of the mirrors, we created various paths through the 12 lenses of the laser cavity. This affected the field configuration of the fundamental mode. We observed strongly deformed beam shapes whose single mode configuration could be confirmed by placing a circular aperture into the path of the beam. We recorded no change of these shapes when the aperture was closed until laser action vanished.

V. BEAM WAVEGUIDE WITH NONCONFOCAL SPACING

Theoretical work by E. A. J. Marcatili¹⁰ indicated that distortions of the Gaussian beam passing through a sequence of lenses are reduced

if a nonconfocal geometry is chosen. To check this prediction experimentally, we aligned a beam waveguide consisting of 10 gas lenses with nonconfocal spacing (a distance-to-focal-length ratio (d/f) of 1.82). At the same time, we could make several improvements in the alignment procedure as a consequence of the experience gathered with the beam guide aligned before. We strongly reduced the influence of building vibrations by working farther away from the maintenance machinery at the far end of the waveguide corridor. The ensuing reduced fluctuations of the light beam allowed a more accurate coaxial alignment of the tubular lens with the light beam. Also, we eliminated the negative influence of the back pressure due to probe insertion on the alignment by leaving a small air gap between the lens (which was to be centered around the beam) and the preceding lens. This permitted the gas to escape. After the centering procedure we connected the two lenses air tight and then compensated for the gravity effect. Because it was a different set of gas lenses than in the experiment before, the vertical displacement to compensate for the gravity effect varied between 17.5 and 20.5 mils, but reached extremes of 15 and 26 mils.

We left the laser cavity untouched, and adjusted the temperatures of the guide lenses to yield focal lengths of 60 and 50 cm, corresponding of $d/f = 1.67$ and 2.0 , respectively (Figs. 12 and 13). This did not deflect the on-axis beam in the horizontal direction. However, small deviations in the vertical position are apparently the result of the temperature dependence of the gravity effect. The horizontal and vertical beam widths of the on-axis beam for confocal geometry were virtually identical with those we measured previously. For nonconfocal operation, the horizontal beamwidth of the on-axis beam remained approximately constant, whereas the vertical width became larger for nonconfocal geometry. We recorded good symmetry of the target beam position with respect to symmetric off-axis injection of the beam in all three cases. The normalized offsets are the same as in Figs. 5 through 8. Again we measured the power profile 6.25 meters behind the last lens of the guide. However, we chose arbitrary units at this time whereby the power into the guide was held approximately constant. We could not confirm the asymmetry observed in Fig. 7 for small vertical offsets.

If you compare the power profiles of light beams passing through a sequence of lenses with different distance-to-focal-length ratios, you notice a marked increase in beam distortions as you approach confocal geometry. This is true for beams undulating horizontally as well as vertically. This behavior was predicted by theory¹⁰ and confirms that

in beam waveguides with nonconfocal spacing there is less mode conversion than in those with confocal spacing.

VI. CONCLUSION

We achieved accurate alignment of a sequence of gas lenses. We found that a beam undulating in an optical waveguide increases its beamwidth as a function of the initial offset, but maintains its Gaussian profile up to a certain maximum displacement. With 10 lenses we recorded maximum offsets of approximately 1.0 to 1.5 (normalized with the theoretical $1/e$ beamwidth $w = 0.45$ mm at the positions of the lenses). Larger offsets resulted in a change of the Gaussian beam shape and the appearance of diffraction rings. Observed symmetry in the position of the maximum intensity and the power profile for left and right horizontal displacements indicate symmetry of the lenses with respect to the vertical plane. Unsymmetrical vertical deformations are apparently the consequence of gravity aberrations.

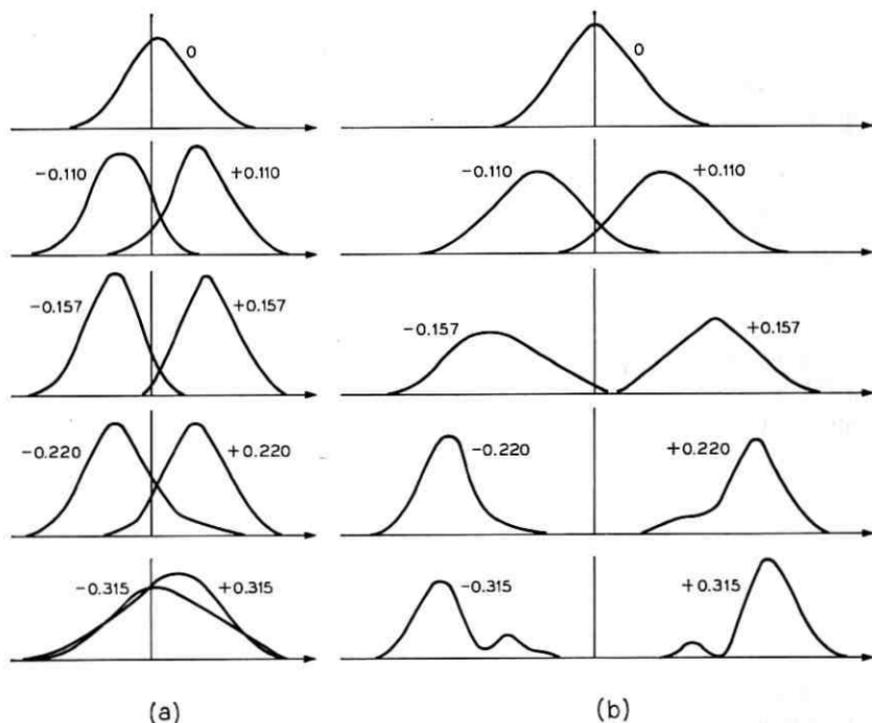


Fig. 12—Horizontal power profile for horizontally-displaced beam (arbitrary units, offset normalized as in Fig. 5). (a) focal length: 60 cm ($d/f = 1.67$). (b) focal length: 50 cm (confocal).

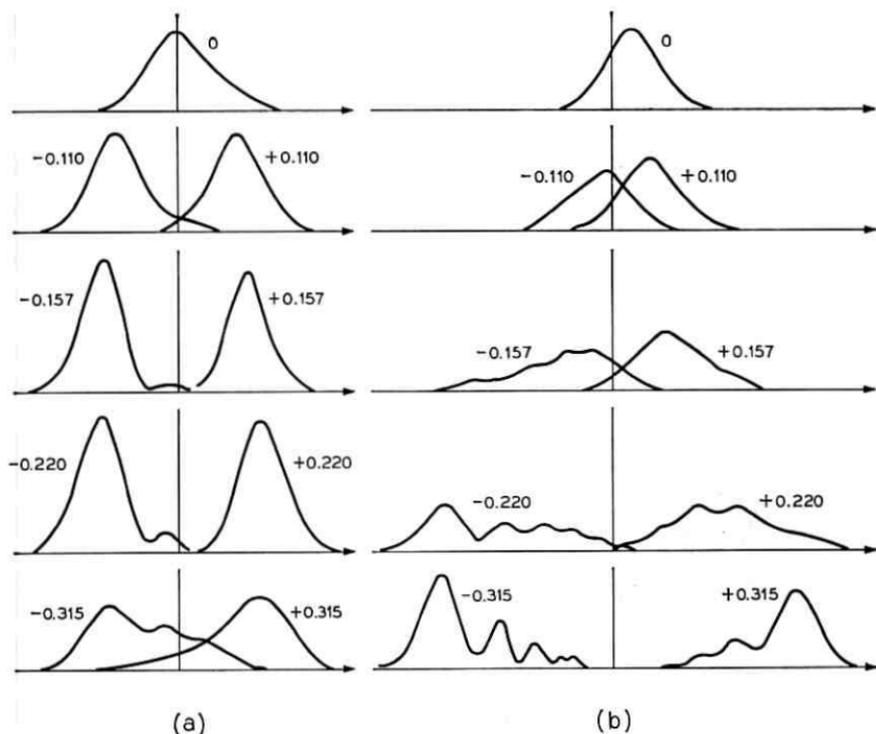


Fig. 13— Vertical power profile for vertically-displaced beam (arbitrary units, offset normalized as in Fig. 7) (a) focal length: 60 cm ($d/f = 1.67$). (b) focal length: 50 cm (confocal).

Thus, averaged aberration properties of gas lenses can be ascertained from offset injection of a light beam into a small sequence of lenses.

We demonstrated effective guidance of a light beam around a circular bend. The minimum radius below which changes of the beam shape occurred was approximately 625 meters for 10 lenses. We confirmed the low-loss guidance properties by using 12 lenses as focusing elements within a laser cavity in the bent state.

We experimentally confirmed the theoretical prediction of reduced beam distortions for nonconfocal geometry. Thus, nonconfocal spacing of the lenses is more advantageous in optical transmission line design than confocal spacing.

VII. ACKNOWLEDGMENT

I gratefully acknowledge the suggestions of E. A. J. Marcatili, as well as the discussions with him. I appreciated the help of H. W. Astle in

carrying out part of the measurements and his assistance in evaluating the data.

REFERENCES

1. Marcuse, D., Comparison Between a Gas Lens and Its Equivalent Thin Lens, *B.S.T.J.*, *45*, October 1966, pp. 1339-1344.
2. Marcuse, D., Deformation of Fields Propagating Through Gas Lenses, *B.S.T.J.*, *45*, October 1966, pp. 1345-1368.
3. Marcuse, D. and Miller, S. E., Analysis of a Tubular Gas Lens, *B.S.T.J.*, *43*, July 1964, pp. 1759-1782.
4. Marcuse, D., Theory of a Thermal Gradient Gas Lens, *IEEE Trans. on Microwave Theory and Techniques*, *MTT-13*, No. 6, 1965, p. 734.
5. Gloge, D., Deformation of Gas Lenses by Gravity, *B.S.T.J.*, *46*, February 1967, pp. 357-365.
6. Marcatili, E. A. J., Off-Axis Wave-Optics Transmission in a Lens-like Medium with Aberration, *B.S.T.J.*, *46*, January 1967, pp. 149-166.
7. Hirano, J. and Fukatsu, Y., Stability of a Light Beam in a Beam Waveguide, *Proc. IEEE* *52*, 1964, p. 1284.
8. Steier, W. H., The Statistical Effects of the Random Variations in the Components of a Beam Waveguide, *B.S.T.J.*, *45*, March 1966, pp. 451-471.
9. Marcatili, E. A. J., Ray Propagation in Beam-Waveguides with Redirectors, *B.S.T.J.*, *45*, January 1966, pp. 105-115.
10. Marcatili, E. A. J., Effect of Redirectors, Refocusers, and Mode Filters on Light Transmission Through Aberrated and Misaligned Lenses, *B.S.T.J.*, *46*, October 1967, pp. 1733-1752.
11. Beck, A. C., an Experimental Gas Lens Optical Transmission Line, *IEEE Trans. Microwave Theory & Techniques*, *MTT-15*, No. 7, July 1967, pp. 433-434; also unpublished work.
12. Marcuse, D., Propagation of Light Rays Through a Lens-Waveguide with Curved Axis, *B.S.T.J.*, *43*, March 1964, pp. 741-753.

Silicon Schottky Barrier Diode with Near-Ideal I-V Characteristics

By M. P. LEPSALTER and S. M. SZE

(Manuscript received October 19, 1967)

Metal-semiconductor diodes with near-ideal forward and reverse I-V characteristics have been fabricated using PtSi contacts and diffused guard rings. Typically, for a device with an area of $2.5 \times 10^{-6} \text{ cm}^2$ made on an n-type (111) oriented, 0.35 ohm-cm silicon epitaxial substrate, the forward current follows the expression $I_f = I_s \exp(qV/nkT)$ over eight orders of magnitude in current with $I_s = 10^{-12} \text{ A}$ and $n = 1.02$. The reverse breakdown is sharp and occurs at the theoretical breakdown voltage of p⁺n silicon junctions of the same n-type doping. The premature breakdown observed in nearly all previous Schottky barrier diodes has been shown to be caused by electrode sharp-edge effects. Besides giving sharp breakdown voltage, the guard ring also eliminates anomalously high leakage currents, yet still retains the fast recovery time characteristic common to other Schottky barriers. Typically, the recovery time measured at 10 ma is less than 0.1 ns, the resolution of the measurement.

I. INTRODUCTION

Metal-semiconductor (or Schottky) barriers have been studied extensively in the past few years. The main concerns are the barrier height and the current transport in the metal-semiconductor system; the former has been reviewed recently by Mead,¹ and the latter discussed by Crowell and Sze.² Most emphasis has been placed on the forward current-voltage characteristic, which relates intimately to the electronic applications of Schottky diodes such as varistors for logic gates and microwave downconverters. The reverse current-voltage characteristic of a conventional planar Schottky diode,³ usually has had greater leakage current and lower breakdown voltage than a diffused p-n junction. Because of this "soft" reverse characteristic, Schottky diodes have not been considered for power application, as well as IMPATT oscillators.

It should be pointed out, however, that if the reverse characteristic can be improved substantially, a Schottky diode with its inherent majority transport property, can be used as a high-speed switch in which there is virtually no minority carrier storage, and as a high-power rectifier or high-frequency oscillator where the region of greatest heat dissipation is located right at the metal-semiconductor interface and therefore the heat can be more readily conducted away.

The soft breakdown of conventional Schottky diode is not caused by the electron tunneling effect but results mainly because of the "edge effect" shown in Fig. 1(a), where high-field concentration gives rise to excess leakage current and low breakdown voltage. The tunneling effect is ruled out because of the fact that even at very large electric fields ($\sim 5 \times 10^5$ V/cm), there is negligible contribution of the tunneling current component to the total conduction current,² owing to the relatively large electron effective mass in silicon such that the tunneling probability of electrons from the metal to the semiconductor conduction band is very small.

This paper presents one method to eliminate this edge effect: the diffused guard-ring method. This method has been used on planar p-n junction devices⁴ to eliminate the junction curvature effect. It is shown in this paper that the guard ring improves both the forward and reverse characteristics of the Schottky diode. It is also demonstrated for the first time that the measured breakdown voltage of a Schottky junction is equal to the theoretical value of a one-sided abrupt p-n junction with the same background doping concentration. In addition, the surface field effect associated with the metal-insulator-semiconductor structure on the junction breakdown is studied.

II. EXPERIMENTAL PROCEDURE AND RESULTS

2.1 *Device Fabrication*

In order to study the effect of a junction guard ring on the I - V characteristics of a metal-semiconductor barrier, three kinds of structures were fabricated on a single wafer as shown in Fig. 1. Fig. 1(a) is a planar PtSi-Si Schottky diode alone; Fig. 1(b) is the diffused p-n guard ring alone; and Fig. 1(c) contains both a PtSi-Si Schottky barrier and a p-n junction which serves as its guard ring.

Most of the wafers used in the study were n -type silicon materials, with 2 to 8 μm thick epitaxial layers, $\langle 111 \rangle$ oriented, 0.3 to 1.2 ohm-cm. The basic steps of fabricating the Schottky diode with a guard ring were: the wafers were first cleaned and degreased; silicon dioxide

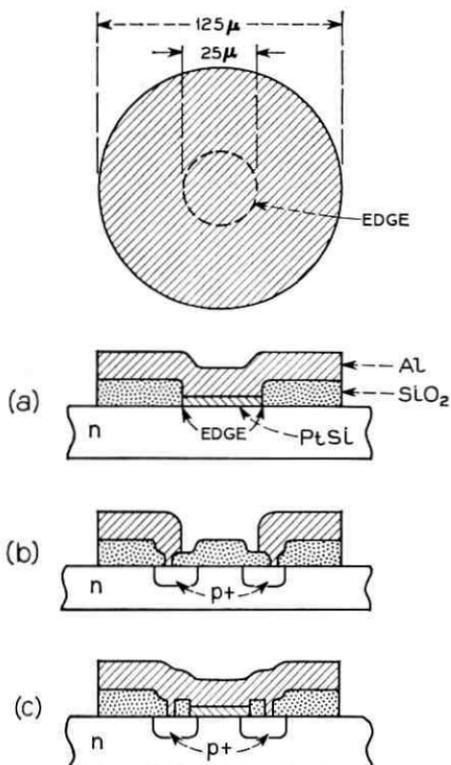


Fig. 1—(a) Planar PtSi Schottky diode. (b) Diffused p+n junction ring. (c) PtSi Schottky barrier with a p+n guard ring as in b.

layers of about 5000 Å were then grown thermally. Diffusion rings were then cut in the oxide by the standard photoresist technique. After the junctions were boron-diffused, another SiO₂ layer was grown on the surface, and photoresist was used to define the center hole for platinum deposition.

In order to produce a metal-semiconductor interface comparable in uniformity to a diffused p-n junction it is required to start with a "clean" silicon surface. Even the most efficient chemical methods leave several atomic layers of inorganic films. In this study, the silicon surfaces were back-sputtered with argon ions (at a pressure of 20 microns) using the setup shown in Fig. 2. Cathode 1 was excited with an RF oscillator, producing a plasma over the sample, and removing any films previously formed. Then a DC voltage was applied to cathode 2, and a 500 Å platinum film deposited. Heating the sample in the

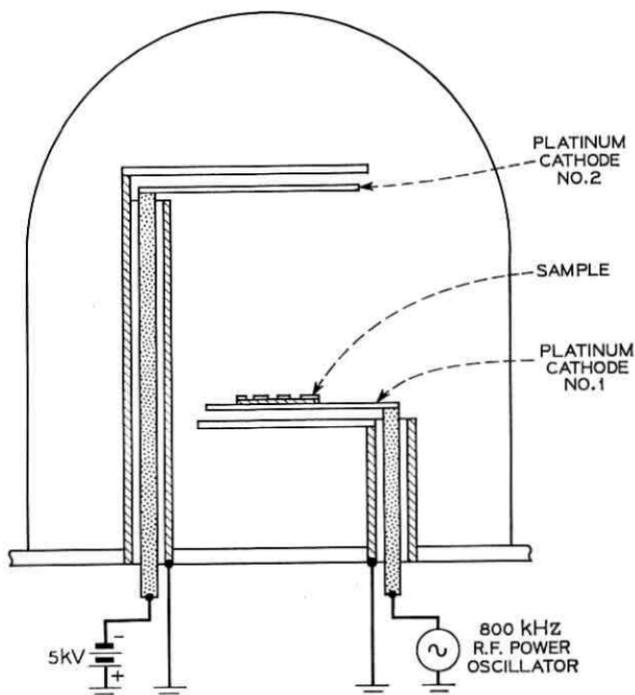


Fig. 2 — Platinum sputtering system.

vacuum chamber to 600°C caused the platinum film to react with the exposed silicon surface. For a clean surface, the reaction proceeded to the most stable phase of the system Pt-Si, forming the compound PtSi. This reaction, verified by X-ray analysis, is strongly influenced by the actual cleanliness of the interface.

The platinum layer deposited on the SiO_2 , which did not react with the exposed silicon, was then removed by aqua regia. The final overlay metal contact was made by evaporation of aluminum and another photoresist to isolate each device. The back contact was formed by evaporation of gold and alloyed at 400°C. The three different structures shown in Fig. 1 were made at the same time on the same wafer by using multiple patterns on the photoresist masks.

2.2 Forward I - V Characteristics

Figure 3 shows the measured forward I - V characteristics for the three structures as shown in Fig. 1. The epitaxial thickness is 2 μm , and the background doping is $2.2 \times 10^{16} \text{ cm}^{-3}$. The diffused junction

depth is about $1 \mu\text{m}$. The guard ring has an area of $2.9 \times 10^{-5} \text{ cm}^2$ and the Schottky junction of $2.5 \times 10^{-6} \text{ cm}^2$. One notes that the current of the Schottky diode with a guard ring (from now on, we call it diode *c* referring to Fig. 1c) follows the expression $I_f = I_s \exp(gV/nkT)$ over eight orders of magnitude with $I_s \cong 10^{-12}$ amp and $n = 1.02$. From the saturation current and the area of the diode ($2.5 \times 10^{-6} \text{ cm}^2$) a barrier height of 0.85 eV for the PtSi-Si barrier is obtained, in agreement with the previously reported results.³

For the planar Schottky diode without a guard ring (diode *a*) it is apparent that the I - V characteristic is much inferior and gives about two to four orders of magnitude excess current at lower bias. The forward I - V characteristic of the p-n junction guard ring is shown in Fig. 1(c). It is clear that the guard ring helps to eliminate the edge leakage current which has normally existed in a planar Schottky diode. By the use of lower resistivity and a thinner epitaxial layer, it is ex-

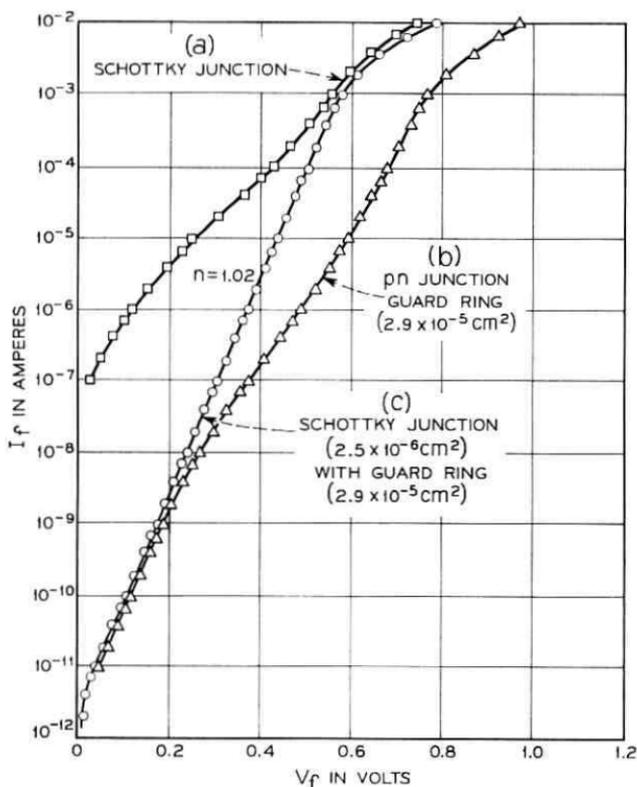


Fig. 3—Forward current-voltage characteristics of devices shown in Fig. 1.

pected that one can obtain an exponential current range (with constant n -value) even larger than the present eight orders of magnitude.

2.3 Reverse I - V Characteristics

The Schottky diode with a guard ring also has a superior reverse I - V characteristics. For diode a the breakdown voltage is only 5 V (at 1 ma). The breakdown voltage for diode b is 27 V. About the same breakdown voltage is obtained for diode c which is shown in Fig. 4 for two different areas of the Schottky diodes. For small area, the space-charge generation and recombination current of the p-n junction guard ring dominates the reverse leakage current. For larger area, however, the reverse current approaches the ideal Schottky

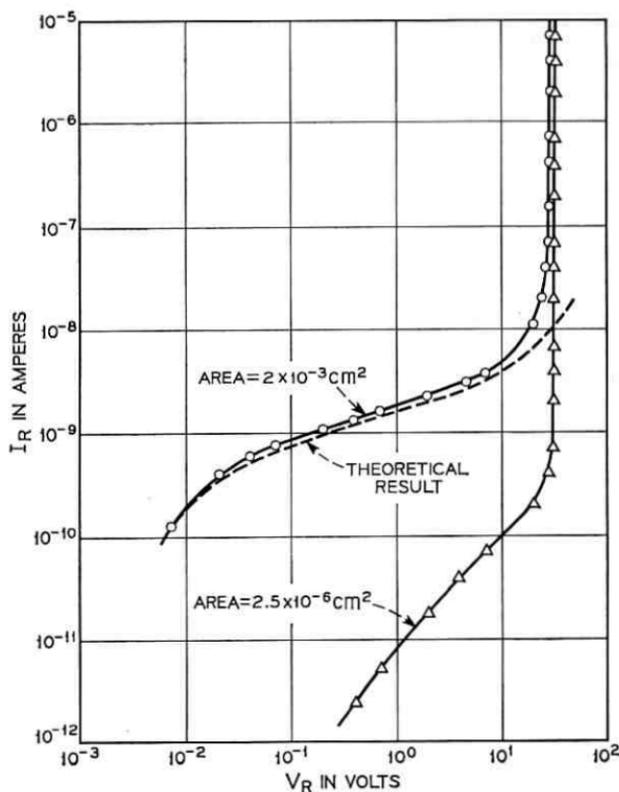


Fig. 4—Reverse current-voltage characteristics of Schottky diodes (type c in Fig. 1). For the small diode, the area of the Schottky diode is $2.5 \times 10^{-6} \text{ cm}^2$ and that of the guard ring is $2.9 \times 10^{-5} \text{ cm}^2$. For the larger one, the areas are $2 \times 10^{-3} \text{ cm}^2$ and $1.5 \times 10^{-4} \text{ cm}^2$, respectively.

barrier current due to image-force lowering (the dotted line is calculated using an image-force dielectric constant of 12).⁵ The breakdown voltages in diodes *b* and *c* are in good agreement with the theoretical calculation shown in Fig. 5. This figure shows the breakdown voltages for Si p^+n junctions as a function of the background doping with the junction depth as a parameter.⁶

For planar diffusion through a photoresist mask, the impurities will diffuse both downward into the bulk semiconductor and sideways under the oxide layer (see insert of Fig. 5). Thus a junction curvature is formed near the corner of the diffusion mask. For a one-sided abrupt junction, this curvature causes a reduction of the breakdown voltage. In the present case with $N_B = 2.2 \times 10^{16} \text{ cm}^{-3}$, the breakdown voltage would be 36 V for a plane junction, and about 25 V for a junction with $1 \mu\text{m}$ curvature,⁶ as shown by the circles in Fig. 5. Therefore the breakdown voltages for diodes *b* and *c* agree with theoretical expectation for a guard ring with $1 \mu\text{m}$ junction depth.

The above result confirms that a Schottky diode with a junction guard ring can have a breakdown voltage many times larger than one without a guard ring. In this example, breakdown occurred at the guard ring; but this result does not imply that the breakdown voltage of a Schottky diode is always limited by the curvature effect of the junction guard ring.

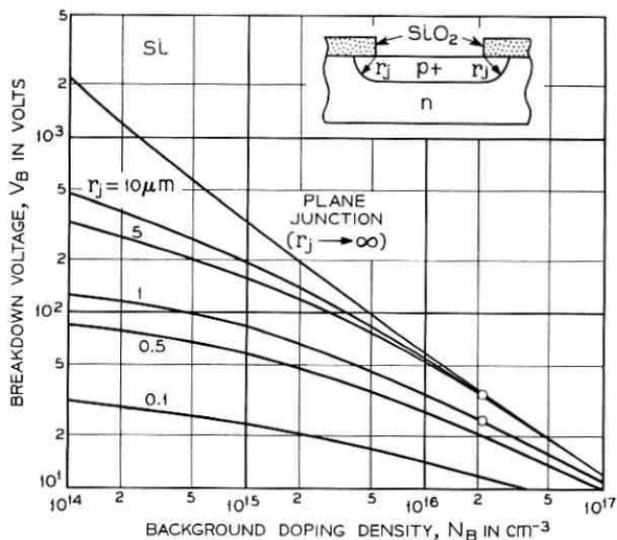


Fig. 5—Theoretical results for silicon p^+n junction with junction depth r_j (insert) as the parameter. (See Ref. 6.)

By proper control of the diffusion process, one can form a linearly-graded junction with V_B which is virtually independent of the junction curvature,⁹ or one can form a composite junction (see insert of Fig. 6); that is, the space charge terminates in a graded region on one side of the junction and in a uniformly doped region on the other. The composite junction has a breakdown voltage always larger than an abrupt junction with the same junction curvature (r_j) and the same background doping (N_B). Some theoretical results of the composite junction are shown in Fig. 6.

In order to make a guard ring which has higher breakdown voltage than that of the Schottky diode, the following experiment was performed. An epitaxial silicon wafer with doping about $2 \times 10^{16} \text{ cm}^{-3}$ and a thickness of $8 \mu\text{m}$ was used to form the three structures as shown in Fig. 1. The diffused guard ring has a doping profile which can be approximated by a composite junction with a junction depth of $3.5 \mu\text{m}$ and an impurity gradient of about $2.5 \times 10^{21} \text{ cm}^{-4}$. The theoretical breakdown voltage for the guard ring is about 50 V and the breakdown voltage for the Schottky barrier is about $42 \pm 2 \text{ V}$ (shown in Fig. 6). Thus the Schottky barrier diode is expected to break down first. Figure 7 shows the I-V characteristics of the Schottky diode with guard ring (diode c) which has a forward characteristic similar to that shown

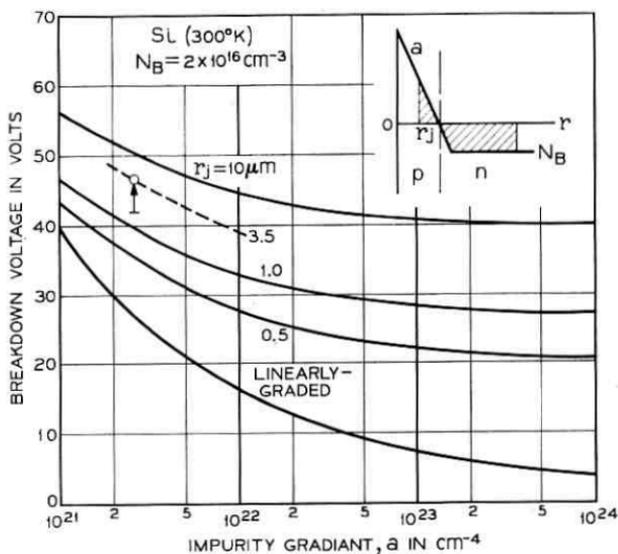


Fig. 6 — Theoretical results for silicon linearly-graded composite junction (insert) with junction depth as the parameter.

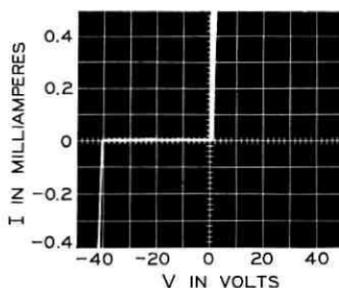


Fig. 7—Oscilloscope current-voltage display of the characteristics of a Schottky diode with the ideal breakdown voltage ($N_D = 2.0 \times 10^{19} \text{ cm}^{-3}$).

in Fig. 3, and a reverse breakdown at about 41 V in good agreement with the expected result.

Figure 8 shows the forward and reverse I - V curves for the three structures. For diode *a* the forward and reverse characteristics are poor. For the p-n junction (guard ring alone, see Fig. 8b) the forward knee occurred at about 0.5 V and the reverse breakdown at 49 V. For diode *c*, the Schottky diode breaks down first at about 43 V. Because of the small junction area there is a large space-charge resistance, R_{sc} , under avalanche conditions. The value of R_{sc} is given by⁷ $R_{sc} \cong [W^2/A(2\epsilon V_d)]$ where W is the width of depletion layer at breakdown, A the junction area, ϵ the permittivity, and V_d the limiting drift velocity. In the present case with $W \cong 3 \mu\text{m}$, $A = 2.5 \times 10^{-8} \text{ cm}^2$, and $V_d \cong 10^7 \text{ cm}$ per second, the value of R_{sc} is 1.5 K Ω , which is in good agreement with the slope shown in Fig. 7(c). (The junction temperature effect and the series resistance effect are calculated to be small compared with the above space-charge effect.) This further confirms that we have uniform breakdown of the Schottky diode. When the voltage reaches about 50 V, the p-n junction guard ring begins to break down similar to case *b* (Fig. 8b). The space-charge resistance in this region is considerably lower because of the larger area of the guard ring.

2.4 Surface Field Effect Study

Notice in Fig. 1(c) that in addition to diffused p^+ guard ring, there is a metal-oxide-semiconductor structure surrounding the device. It is known that when a reverse bias is applied to the metal electrode, the surface space-charge region under the oxide is in nonequilibrium condition.⁸

It has been shown that a surface field has a profound effect on the breakdown of planar p-n junctions.⁹ In order to study this effect on

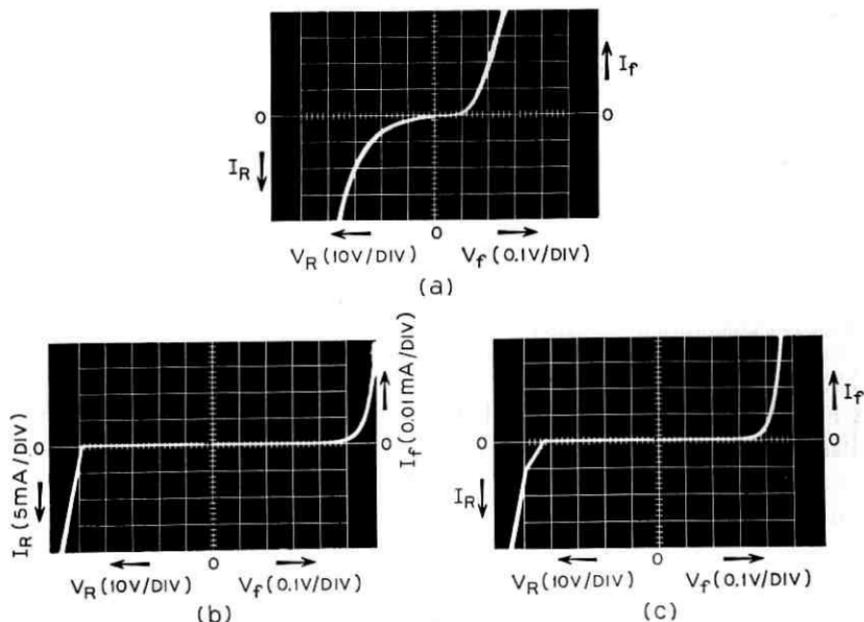


Fig. 8—Oscilloscope display of the current-voltage characteristics for the structures shown in Fig. 1.

the breakdown voltage of Schottky barrier diodes, a separate metal electrode on the oxide (that is, the gate electrode) is fabricated as shown in Fig. 9. The fabrication procedures were the same as described previously. The samples used were (111) oriented, 1.0 ohm-cm, *n*-type silicon wafers. When a negative gate bias is applied, the surface field tends to smooth out the field concentration near the junction edge (see Fig. 10). Thus the radius of curvature, r_j , is effectively in-

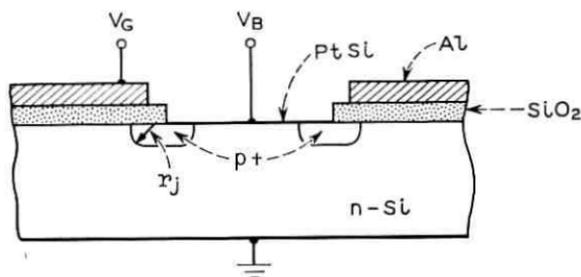


Fig. 9—Schottky barrier diode with separate electrodes on PtSi and on the oxide.

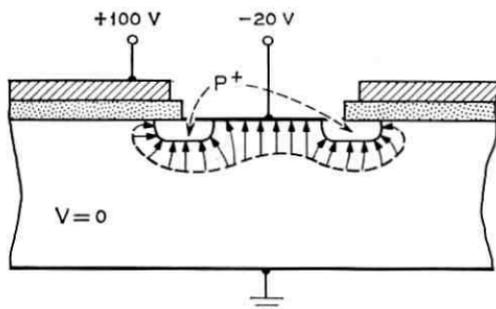


Fig. 10 — Electric field distribution for negative gate voltage.

creased. This in turn increases the breakdown voltage (the relationship between V_B and r_j has been shown in Fig. 5).

Figure 11 shows the field profile near the junction as a positive gate bias is applied, where the radius of curvature, r_j , is effectively reduced resulting in a decreased breakdown voltage. Figure 12 shows the measured reverse $I-V$ characteristics as a function of the gate voltage for a Schottky diode with a guard ring of $0.4 \mu\text{m}$ junctions depth. As expected, the gate voltage does have a profound effect on the junction breakdown voltage. Figure 13 shows the measured breakdown voltage versus gate voltage for four different junction depths. Notice that the breakdown voltages all approach the theoretical value ($\approx 100 \text{ V}$) as $-V_G$ increases. Also at zero bias, the breakdown voltage decreases as the junction depth r_j decreases.

Similar effects are observed on a Schottky diode that has no p-n junction guard ring but has a second metal oxide semiconductor overlay as shown on the Fig. 14 insert, where ZrO_2 is formed near the periphery of the metal, and separate voltages are applied to the

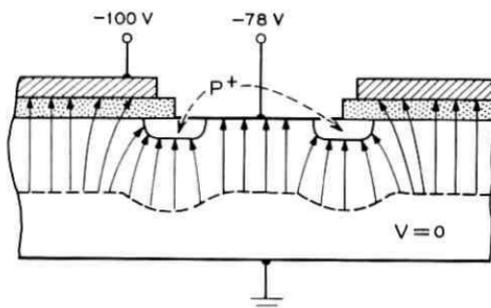


Fig. 11 — Electric field distribution for positive gate voltage.

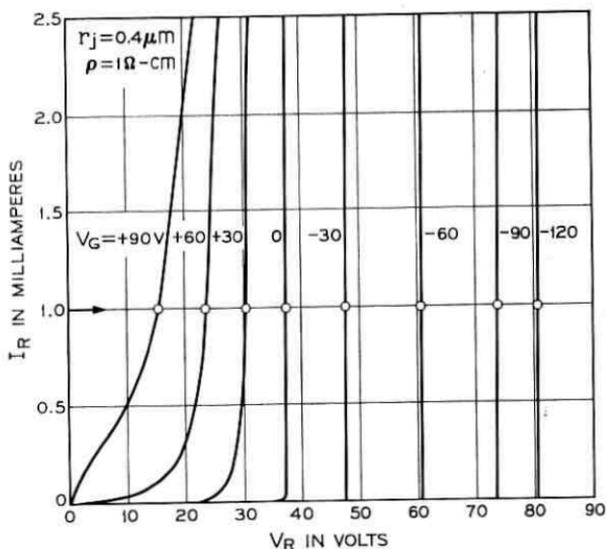


Fig. 12—Reverse current-voltage characteristics as a function of the gate voltage on a Schottky diode with a p-n junction guard ring.

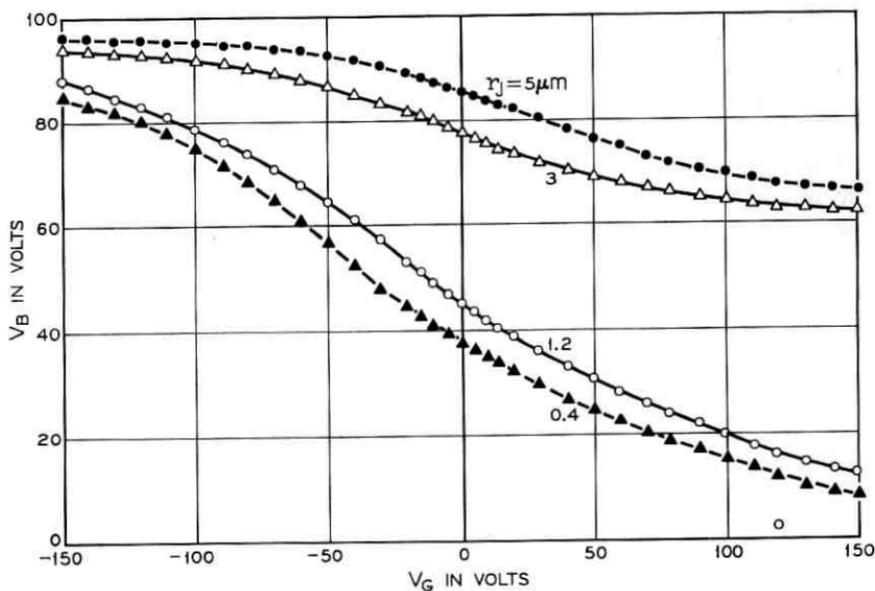


Fig. 13—Measured breakdown voltage (at 1 ma) versus gate voltage for Schottky diode with p-n junction guard ring. The junction depths are 0.4, 1.2, 3, and 5 μm .

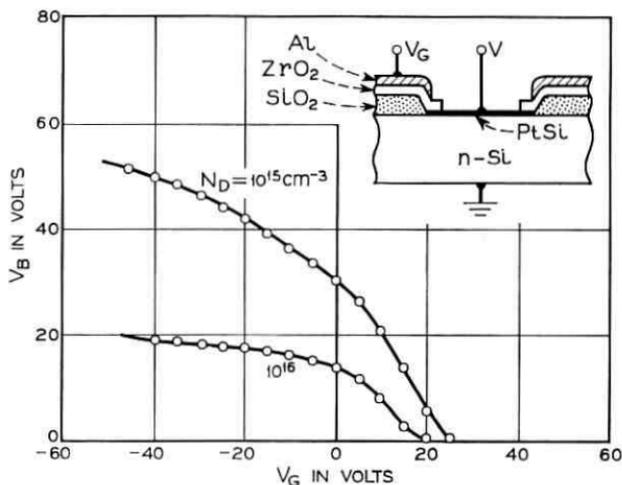


Fig. 14 — Measured breakdown voltage versus gate voltage for Schottky diode with a metal oxide semiconductor overlay near the periphery (insert).

diode and the metal oxide semiconductor guard ring electrodes. As the gate voltage increases in the negative direction, the edge field is gradually reduced, resulting in an increased breakdown voltage (here defined as voltage drawing 1 ma current). When V_G increases in the positive direction, however, the enhanced edge field causes drastic lowering of the junction breakdown voltage as Fig. 14 shows.

III. SUMMARY

It has been demonstrated that the breakdown voltage of a Schottky barrier diode can be made to approach the theoretical value of one-sided abrupt p-n junction with the same background doping. The much lower breakdown voltages usually obtained result from "edge effect" which can be modified or eliminated by proper use of either a diffused guard ring or a surrounding electrode. It has also been shown that the surface field can have profound influence on the guard ring breakdown characteristics. Since any planar diffused guard ring has a finite junction curvature (r_j), to eliminate the junction curvature effect one has to control the diffusion depth and diffusion profile such that the breakdown voltage of the guard ring is larger than that of the Schottky junction.

IV. ACKNOWLEDGMENT

The authors wish to thank D. Kahng for his stimulating discussions. They also wish to thank R. W. MacDonald for his assistance

in processing the samples, Mrs. M. H. Read for her X-ray analysis of the PtSi films, and A. Goetzberger and R. M. Ryder for helpful suggestions.

REFERENCES

1. Mead, C. A., Metal-Semiconductor Surface Barriers, *Solid State Electron*, *9*, 1966, pp. 1023-1034.
2. Crowell, C. R. and Sze, S. M., Current Transport in Metal-Semiconductor Barriers, *Solid State Electron*, *9*, 1966, pp. 1035-1048.
3. Kahng, D. and Lepselter, M. P., Planar Epitaxial Silicon Schottky Barrier Diodes, *B.S.T.J.*, *44*, 1965, pp. 1525-1528.
4. Goetzberger, A., MacDonald, B., Haitz, R. H., and Scarlett, R. H., Avalanche Effects in Silicon pn Junctions II. Structurally Perfect Junctions, *J. Appl. Phys.*, *34*, 1963, pp. 1591-1600.
5. Sze, S. M., Crowell, C. R., and Kahng, D., Photoelectric Determination of the Image Force Dielectric Constant for Hot Electrons in Schottky Barriers, *J. Appl. Phys.*, *35*, 1964, pp. 2534-2536.
6. Sze, S. M. and Gibbons, G., Effect of Junction Curvature on Breakdown Voltage in Semiconductors, *Solid State Electron*, *9*, 1966, pp. 831-845.
7. Sze, S. M. and Shockley, W., Unit-Cube Expression for Space-Charge Resistance, *B.S.T.J.*, *46*, 1967, pp. 837-842.
8. Grove, A. S. and Fitzgerald, D. J., Surface Effects on pn Junctions: Characteristics of Surface Space-Charge Region Under Non-Equilibrium Conditions, *Solid State Electron*, *9*, 1966, pp. 783-806.
9. Grove, A. S., Leistiko, O., and Hooper, W. W., Effects of Surface Fields on the Breakdown Voltages of Planar Silicon pn Junctions, *IEEE Trans. Electron Devices*, *ED-13*, 1967, pp. 157-162.

Design of a Simulator for Investigating Organic Synchronization Systems

By R. H. BOSWORTH, F. W. KAMMERER, D. E. ROWLINSON
and J. V. SCATTAGLIA

(Manuscript received October 24, 1967)

This paper describes circuits and devices designed for a real time simulator that was used to study a network synchronization scheme. The simulator represents a system of four geographically-separated stations with compound interconnections. The system clock rate is a function of the states of each local clock and the interconnecting paths. Component parts for the simulation system were built to specifications exceeding requirements of an actual system so that the operating states could be precisely determined. The simulator performed reliably during the series of experiments which provided the data described in a companion paper.¹

I. INTRODUCTION

The problem of maintaining synchronism in a pulse code modulation network has been of interest for some time.

One technique for synchronizing the geographically separated clocks of a network is known as organic synchronization. In this scheme, the frequency of each clock is controlled to maintain a phase lock between its signal and the average phase of signals arriving from other stations.

The next article in this issue describes a design of such controls and also describes simulations that demonstrate the practicality of the technique.¹ This paper describes the analog simulator used in that work.

The organic synchronizing method compensates the slow phase drift caused by discrepancies in the clock frequencies and the phase changes brought about by changes of transmission delays. The fast phase jitter introduced by changes in the PCM word patterns is not the concern of this study.² Therefore, in the simulator, it is convenient to use the cycles of a constant amplitude sinusoid to represent the

PCM frames. The sinusoids are generated by independent oscillators, one situated at each station, and the transmission paths between them are simulated by delay lines.

The particular network simulated is shown in Fig. 1. It comprises four stations interconnected with 12 links. Each link has a large fixed delay representing a transmission path typical of transcontinental distances. A small variable delay is added for simulation of slow delay changes.

II. A STATION DIAGRAM

The control functions for a single station are illustrated in Fig. 2. This station is designated station i . The other stations are labeled j , k , and l . In Fig. 2, only the link from station j is shown in detail, but similar equipment is provided in all links arriving at each of the four stations. The function of the system blocks will be described while tracing a signal through the network.

The sinusoidal output of oscillator j comes to station i through a cascade of fixed delay elements selected to represent a particular path length. The last delay element in the path is servodriven and is continuously variable under control of an external signal to simulate transmission delay variation.

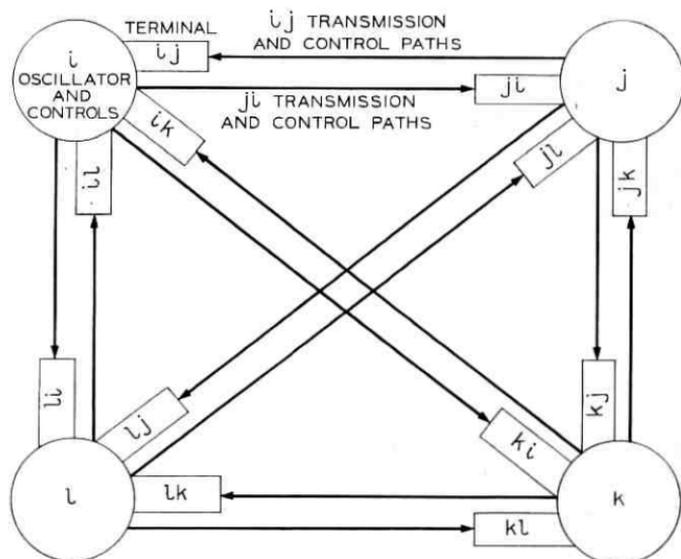


Fig. 1—A network for simulating organic synchronization of four geographically-separated stations.

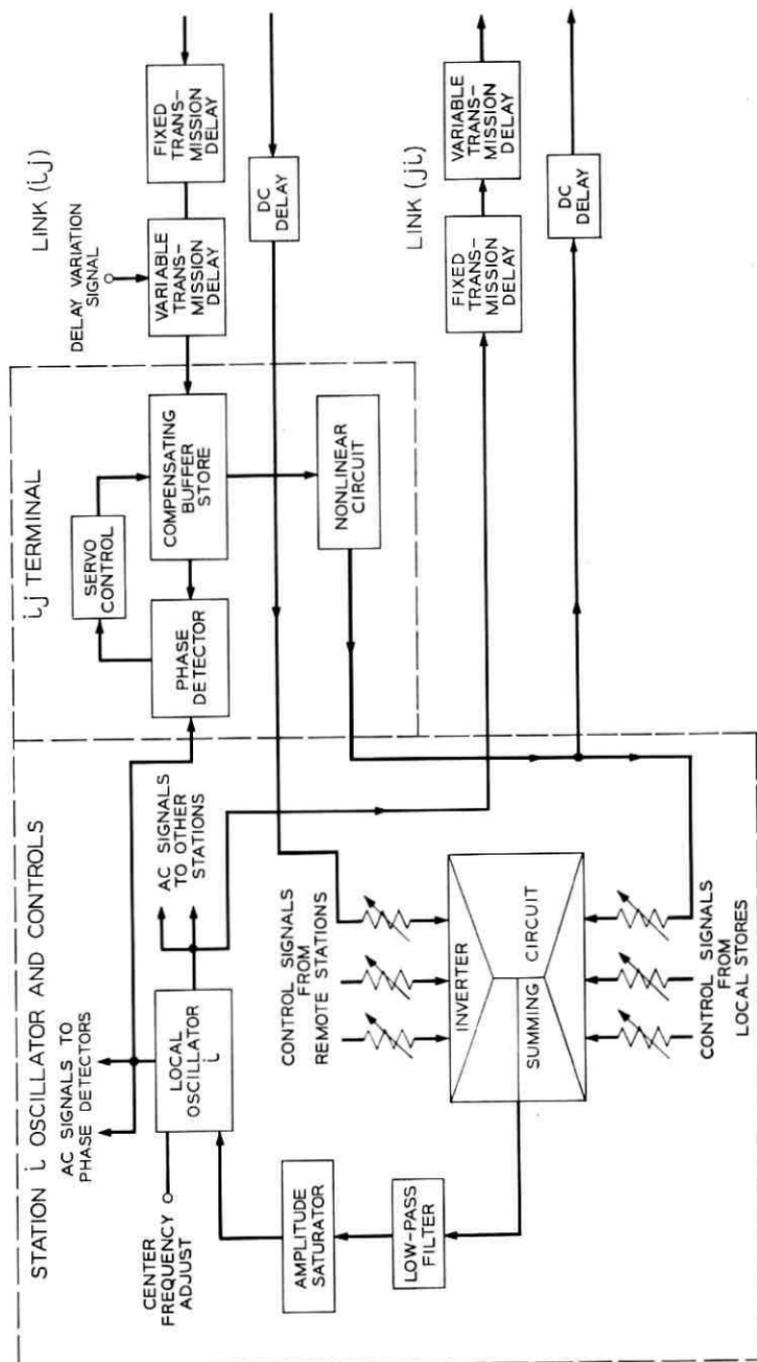


Fig. 2 — Control functions for one of four stations. The illustration shows station i and the simulated transmission links connecting with station j .

At station i the signal is fed to a second continuously variable delay, the compensating buffer store, that is servodriven to maintain the required phase lock. The drive for this delay is derived from the principal phase difference between the output of the delay itself and the local clock. Sufficient gain is provided to keep the phase tracking error less than 3 degrees.

After synchronization of the phases is acquired, change in the delay position is a continuous measure of phase change between the signal arriving from j and the i oscillator. A voltage proportional to this delay position emanates from the compensating buffer store (see box representing this store in Fig. 2). This voltage, designated x_{ij} , is transformed in a series of signal processing operations and used to control the i and j oscillator frequencies in such a way as to reduce the long term frequency differences and minimize the deviation of the compensating delays from their center positions.

The control signal, x_{ij} , may first be modified in a nonlinear circuit that emphasizes the large amplitudes in order to enhance the control influence of delays approaching overflow. Then it is fed separately to two summing circuits, one located at station i and the other at station j . These circuits generate weighted sums of all control voltages derived from links arriving at a station and subtract from the total a weighted sum of the control voltages from links departing the station. The controls returned from remote ends of the departing links are delayed in the simulator, as they would necessarily be in a real system.

The output of the summing amplifier is filtered and applied to a saturating circuit that limits the maximum amplitude excursions. It is then used to adjust the oscillator frequency proportionately.

III. DESIGN SPECIFICATION

In designing simulators, it is unnecessary to copy specifications of realistic systems. Indeed, it is often advisable to use scaling factors that make the simulator convenient to operate. However, for this work, realistic specifications for a possible continental network are followed in order to gain practical experience in designing and operating a network, and to discover any factors of practical importance that may have been overlooked in theoretical studies.

An important component of a PCM center is the local clock which determines the timing of all switching actions at the center. In the simulator, no digital processing is performed, therefore, the clock is simply a 1 megahertz sinusoidal oscillator, one cycle of which represents a frame of information in the digital system. A frequency con-

trol range of one part in 10^6 would be acceptable in practice. A suitable crystal tuned clock would typically have a long term stability of a little less than one part in 10^8 . This sets a limit of about 1 per cent on the realizable control accuracy. To maintain this accuracy, noise and zero drift introduced into the low-pass control signals must be less than 1 per cent of full amplitude.

Transmission delays lie in the range 5 to 8 microseconds per mile, depending on the transmission medium. Thus, delays in a continental system may typically lie in the range 0 to 24 milliseconds. In typical conditions, such delays would vary by amounts up to 10 microseconds, largely because of temperature change. The synchronizing system compensating for this variation would have storage capacity in excess of the expected variation in order to have some control margin.

Components of the simulator have properties similar to those discussed above, but the range of parameter variation and disturbances have been increased in order to provide adequate margin for tests under extreme conditions. For measuring convenience, all sources of noise and drift in the system have been made much less than 1 per cent of signal, so that representative disturbances can be introduced in a controlled fashion.

The servodrive for the compensating delay is designed so that its filtering action on the frequency control signal can be neglected. For this approximation to hold, the low-frequency linear response of the servo has been made equivalent to a delay of less than 2 milliseconds. Hysteresis and other tracking errors are less than 1 per cent of a cycle.

IV. THE OSCILLATORS

It is necessary to have very stable oscillators in order to keep spurious phase errors small. The 1 megahertz crystal oscillators purchased and modified for this application deviate, on average, less than one cycle of phase in ten minutes. This is the approximate duration of an experiment.

Each oscillator is individually mounted in a temperature-stabilized enclosure, and is electrically screened from the other units.

The oscillator frequencies are controlled linearly over a range ± 1 cycle per second (that is, one part in 10^6), by changing the voltage on a varactor diode that is loosely coupled to the crystal. The center frequency and voltage sensitivity are both adjusted by precise potentiometers connected as shown in Figure 3. Figure 4 gives an oscillator control characteristic.

When the simulator is in use, the zero settings of the oscillators are

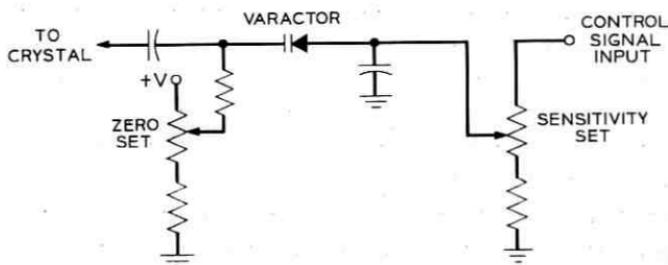


Fig. 3—Circuit for controlling the 1 MHz station oscillator. The oscillator is controlled linearly through a range of ± 1 hertz.

checked before each experiment and their sensitivities are measured about once per week. Realignment, which seldom is necessary, is accomplished by observing Lissajou ellipses which portray the oscillator phases with respect to one another and to an atomic clock.

Zero settings are adjusted to give stationary ellipses when the control voltage is zero. Next, the sensitivities are equalized by adjusting the control voltage to give stationary patterns when all the oscillators are modulated simultaneously with a low-frequency sine wave. In this measurement one oscillator is used as a reference. Its absolute sensitivity is set using calibrated dc inputs and timing beats with respect to the atomic clock.

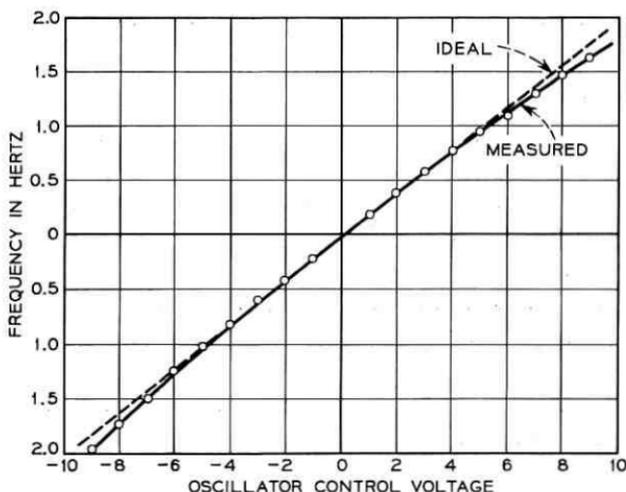


Fig. 4—Normalized frequency response to a control signal for a typical station oscillator.

V. SIMULATION OF FIXED TRANSMISSION DELAY

Fixed transmission path delays are simulated with commercially available magnetostrictive lines of 2, 6, and 8 milliseconds. Each link uses three of these in tandem. Thus, values of delay from 0 to 24 milliseconds are available in increments of 2 milliseconds. Each delay line is accompanied by an amplifier that makes up the amplitude loss in the magnetostrictive line and its transducers. The net loss is about $(40 + 3\tau)$ dB where τ is the nominal delay in milliseconds. A representative delay unit is shown in Fig. 5.

The gain of each of these amplifiers is tailored to match the associated delay loss. A patching coaxial cable is used to couple the amplifier output to the next delay unit. A filter circuit is incorporated that reduces the effect of using cables of different lengths.

To simulate the fixed delay in the return dc control signal paths, lumped constant LC delays are used. Values of 2, 6, and 8 milliseconds are provided by π -sections selected in sets to equal the corresponding ac delay.

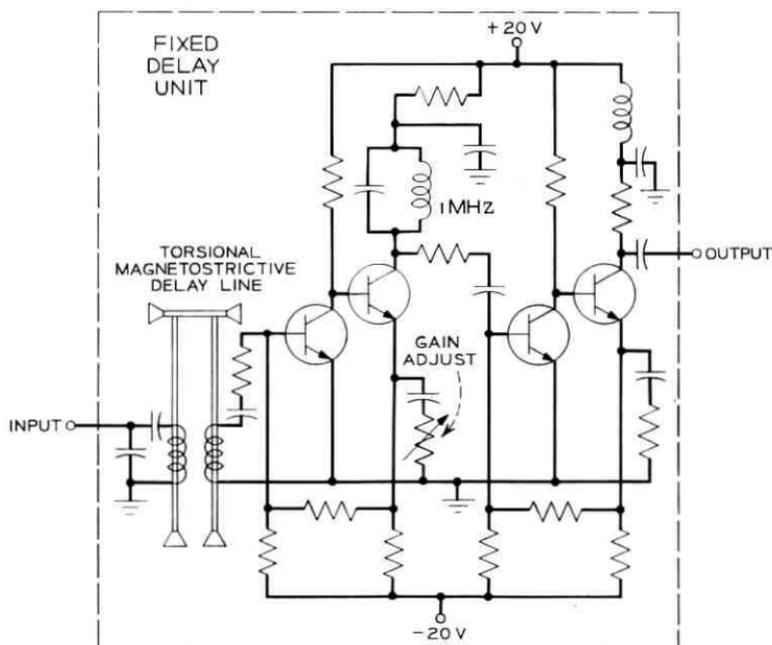


Fig. 5—Transmission delays are simulated with fixed magnetostrictive delay lines. Associated amplifier compensates for insertion loss.

VI. THE CONTINUOUSLY VARIABLE DELAYS

Continuously variable magnetostrictive delay lines are used to simulate the compensating buffer stores that would be used in a real PCM network. Similar delay lines are also used to introduce variation in the length of the simulated transmission links.

Fig. 6 is a simplified diagram of the compensating buffer store. Each sonic delay wire is mounted horizontally and is paralleled by two resistive potentiometers. Potentiometer A is used to monitor the state of the delay. Potentiometer B is used to generate control signals. Terminal C is the output for the delayed ac signals. Correspondingly-labeled terminals are joined by a flexible cable. A sliding carriage holds the potentiometer wipers, a magnetostrictive receiving transducer, and signal preamplifier. It is driven along a steel rail by a 400-cps motor through a gear train, windlass, and steel cable.

The preamplifier, together with a main amplifier, compensates for approximately 66 dB of loss in the transducers. The loss in the sonic wire is insignificant; therefore, the output is virtually independent of delay position. The variable line has a larger transducer loss than a fixed line because of the inefficiency of the mobile receiver transducer.

Torsional mode lines, rather than conventional longitudinal lines were selected for their greater delay per unit length, that is, approximately 3.3 microsecond per centimeter. The maximum delay is a little more than 200 microseconds. Because of the need to resolve small fractions of a cycle (that is, 1 per cent), the electromechanical drive requires great precision. This is further complicated by the need, already mentioned, for a short time constant in its linear response to slow disturbances. Design details of the servomechanism for the compensating delay shown in Fig. 6 are given in the Appendix.

The "variable transmission delay" simulator has less stringent requirements. It is controlled by a conventional positioning servomechanism shown in Fig. 7. The motor drives these delays so that the potentiometer voltage at B approximately equals the applied voltage V_1 . Six volts are needed to fully deflect the lines and the tracking error corresponds with an input of less than 50 mV.

Fig. 8 shows a completed unit which is representative of the 24 that are used for both transmission delay variations and for compensating buffer storage.

VII. THE NONLINEAR FUNCTION

Nonlinear shaping of the frequency control signal is conveniently accomplished by loading the potentiometer output at B with a

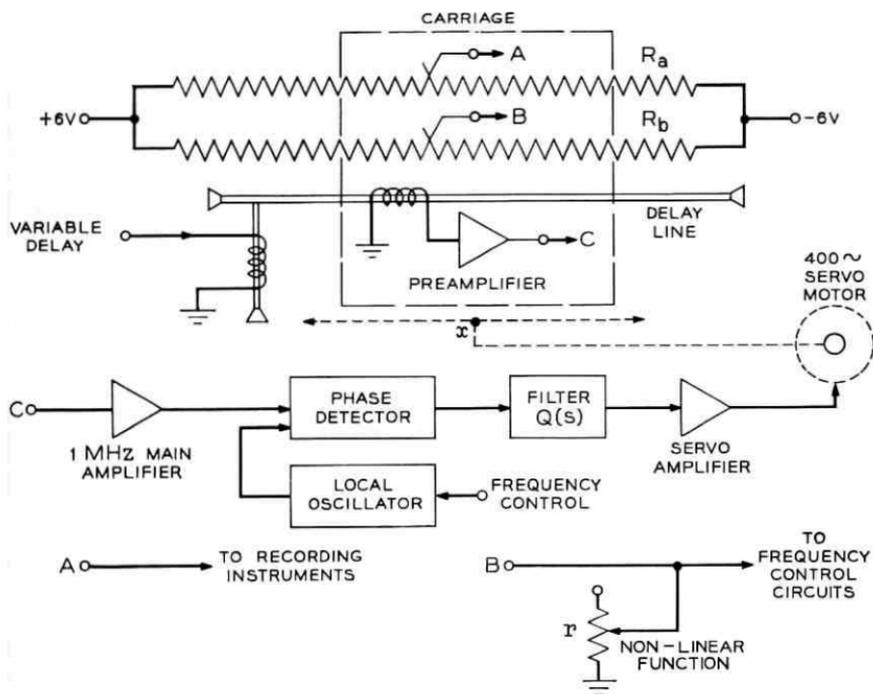


Fig. 6—The compensating buffer store is a servo-controlled, continuously variable magnetostrictive delay line. The maximum delay is slightly more than 200 microseconds.

variable resistor, as shown in Fig. 6. The control voltage is then given by

$$\delta(x) = x \frac{(\mu - 1)}{(\mu - x^2)} V$$

where

$$\mu = \frac{4r}{R_b} + 1,$$

x is the fractional deviation of the delay from center, $+V$ and $-V$ are the voltages applied to the potentiometer, R_b is the potentiometer resistance (5K), r is the load resistor which is variable from 220 Ω to 10K.

Use of this potentiometer loading introduces contact noise into the control, but this has not been a serious problem during the 6 months the system has been in use.

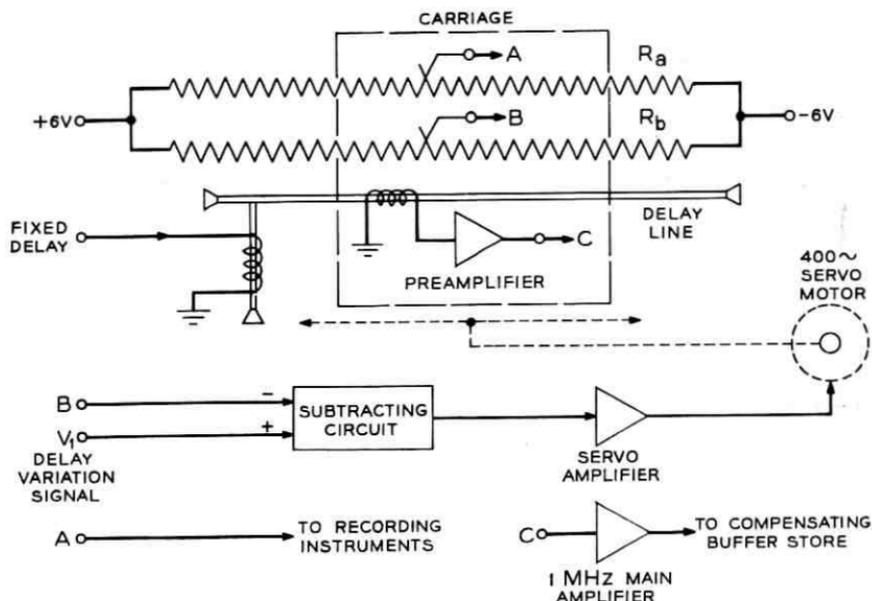


Fig. 7—Variable transmission delay simulator. The variable delay line assembly is identical to the type used in the buffer store but differs in the control circuits.

VIII. THE PHASE DETECTORS

The phase detector circuit is intended to measure the small phase discrepancy between two supposedly locked sinusoids. For this purpose, a linear response is needed only over a small part of a cycle. However, it is important that the phase indication be independent of input amplitude. Otherwise, spurious amplitude changes, such as might be caused by delay line imperfections, would perturb the control and could possibly cause instability.

Fig. 9 shows the phase detector schematically. The two input sig-

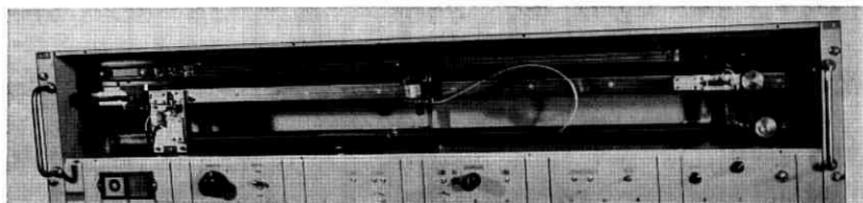


Fig. 8—One of 24 variable delay units. Twelve are used as buffer stores, the remainder serve as transmission delay simulators.

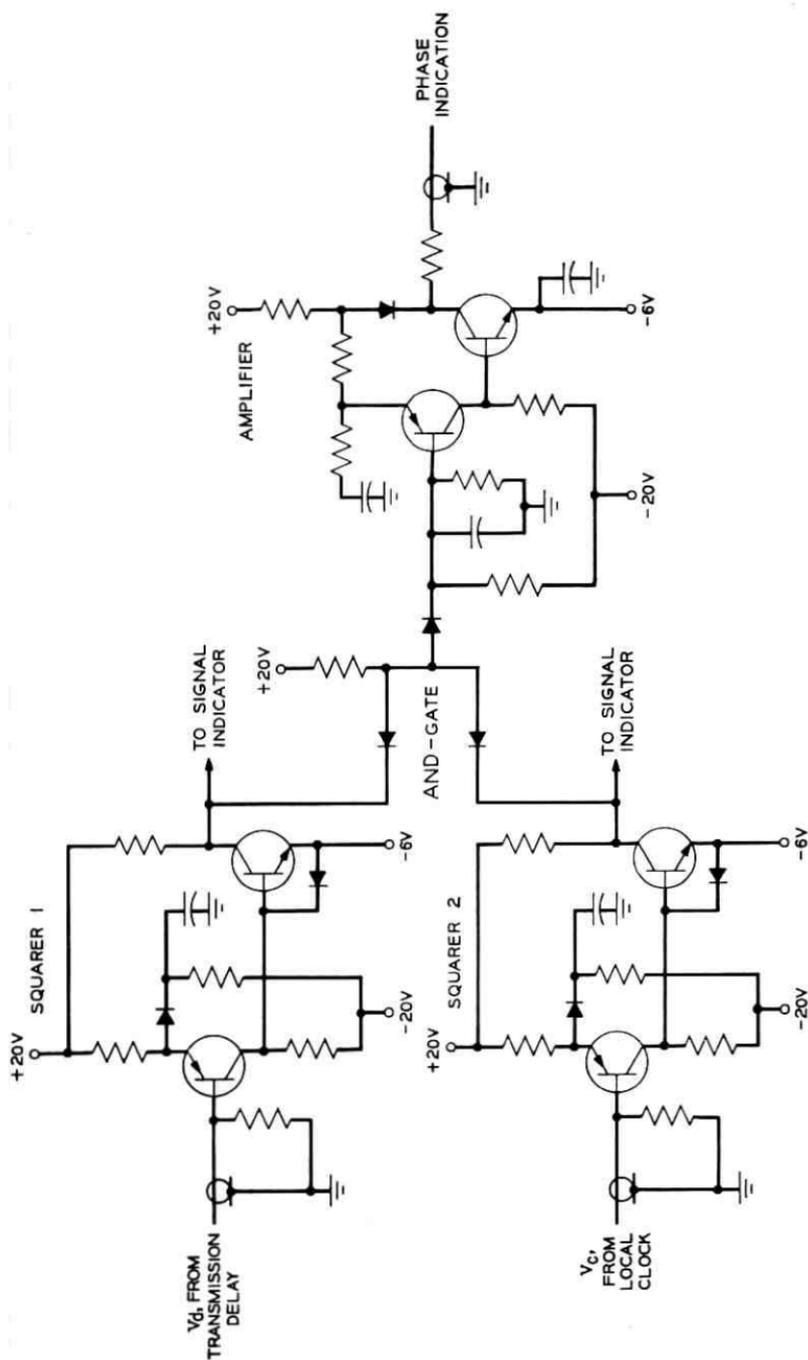


Fig. 9—Phase detector for measuring small phase discrepancies between two sinusoidal signals.

nals are applied directly to individual waveform squaring circuits.³ The phase difference between the resulting squarewaves is measured by determining the duration of their time overlap in an *and* gate. This gate is biased to give zero mean output when the two inputs are in quadrature, that is, have 50 per cent overlap, as seen in Fig. 10(a). This figure shows the net output which has been smoothed and fed through the dc amplifier. The amplifier incorporates the filter needed to stabilize the servo, specifications of which are derived in the Appendix.

The fact that the system synchronizes with the signals in quadrature at each phase detector is inconsequential to the simulator experiments. A linear phase response is available for almost ± 90 degrees deviation from the quiescent phase. This is more than ample because 3 degrees phase displacement drives the motor to full speed.

Fig. 10(b) shows that the phase indication is sensibly independent of input amplitudes within a large range.

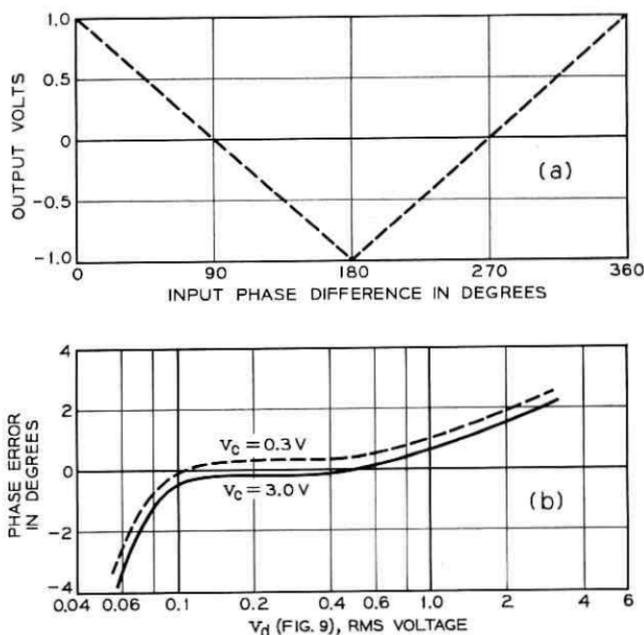


Fig. 10—(a) Output of the phase detector as a function of phase difference between two input signals. (b) Phase errors resulting from amplitude variations in the local clock, v_c , and delayed signals, v_d , are negligible over the required operating range.

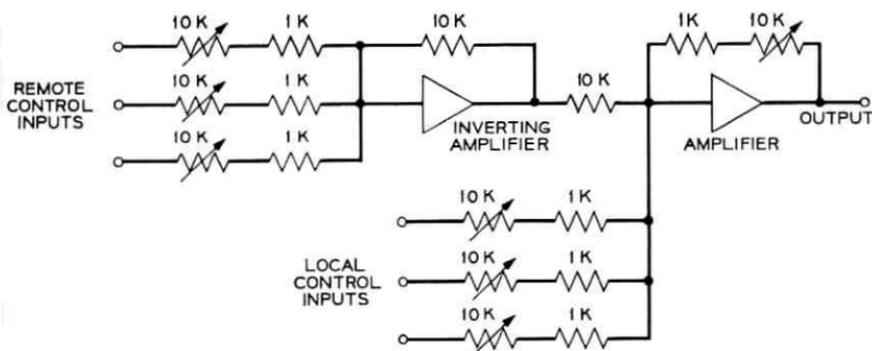


Fig. 11 — Summing of weighted amplitudes.

IX. THE OSCILLATOR CONTROL PATH

The signals used to control the oscillator frequency are combined and shaped in a series of operations depicted here. The summing of weighted amplitudes is performed by conventional operational amplifier circuits, Fig. 11. Similar amplifiers are also used in the filter circuit, Fig. 12, and the amplitude saturator, Fig. 13.

The properties of these circuits can be summarized as follows. The electrical gain in each limb of the summing amplifier is continuously variable over the range $1/11$ to 11 . The filter has a simple low-pass characteristic, $1/(1+pT)$, for signals below 100 cycles per second. The time constant T is continuously variable from 0.2 to 200 seconds. The saturator clipping levels are continuously variable from 0.1 to 10 volts and they automatically remain symmetrical about zero.

The zero drift and low-frequency noise in these circuits is less than

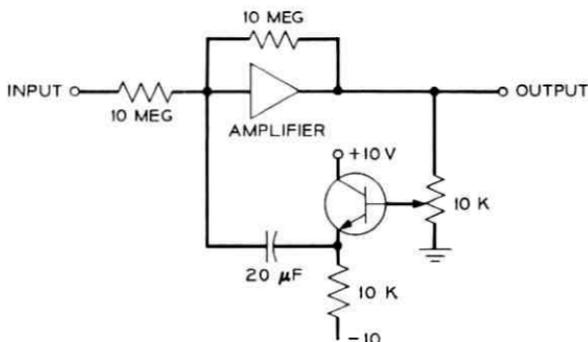


Fig. 12 — Low pass filters.

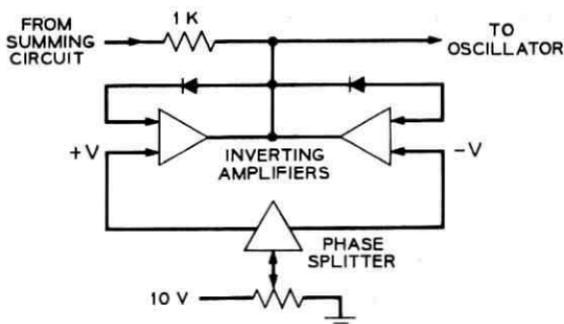


Fig. 13 — Amplitude saturator.

50 mV at the oscillator control inputs and they are designed to accommodate 10 volt peak amplitudes.

X. OPERATION OF THE SIMULATOR

The simulator is arranged to have two active states: standby and run. In the standby condition, the external disturbances are disconnected from all the variable delay simulators and from the oscillators. The compensating delay control servos are disconnected from the phase detector outputs and are connected instead as simple position servos. In this quiescent state the positions of all delays and the oscillator frequencies are determined by preset controls.

To operate the system in the test mode two switches are provided. One connects the externally-provided disturbance signals to the variable delays (Fig. 7) and to the oscillators. The other switch connects the compensating buffer stores (Fig. 6) in the normal run position.

XI. REACTION TO FAILURES

In a real PCM network it is imperative that a failure in any one section of the system not disable the entire network. One assurance against such a catastrophe is to provide an interlock that requires an output from both square-wave generators in the phase detector before the associated servodrive can take the delay from its quiescent setting. Then, if a transmission link fails or an oscillator signal is lost, all the associated compensating delays immediately go to their quiescent states and have no further influence on the frequency control of the remaining operating system.

This facility is part of the simulator. It is used to simulate response in various fault conditions.

XII. SUMMATION

The simulator has been used to investigate a large number of networks and control parameters. It has performed reliably and functions close to theoretical prediction. Some specific results that have been obtained are reported in the article which follows this.¹

Figs. 14 and 15 give a general view of the complete simulator. Fig. 14 shows the array of 24 servodriven transports which serve as variable delay units and Fig. 15 shows the master control and recording equipment.

ACKNOWLEDGMENTS

This work was done under the supervision of M. Karnaugh. We are grateful to J. C. Candy for his counsel with much of the design.

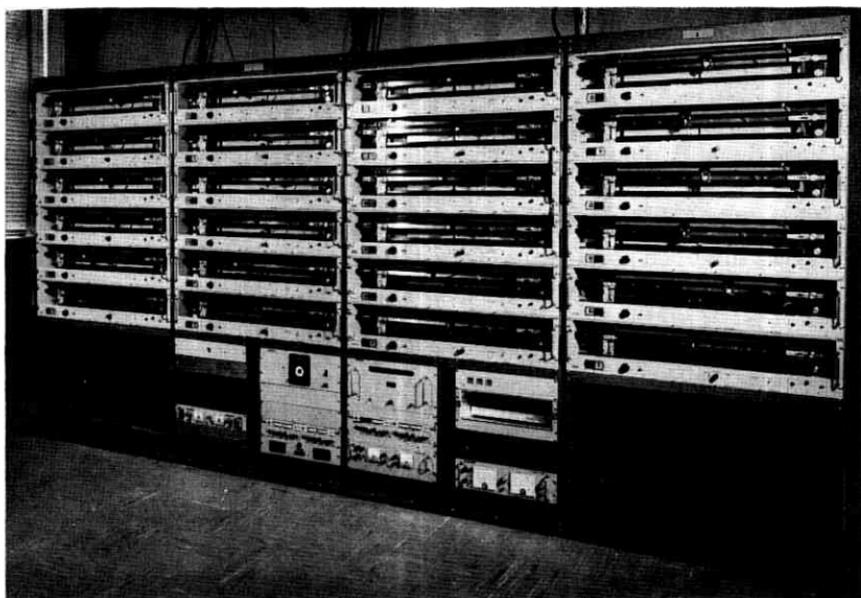


Fig. 14 — Bays containing the 24 servo-controlled variable delay lines.

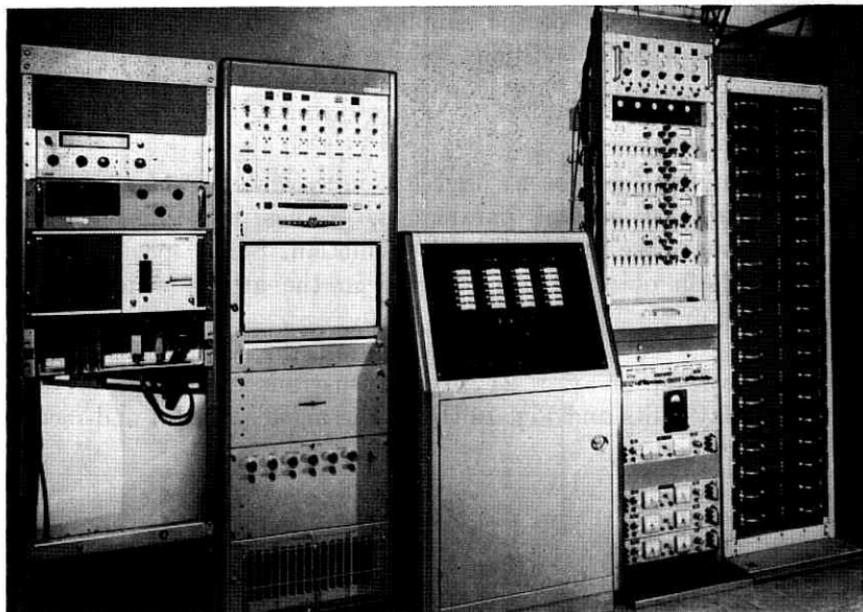


Fig. 15—Left to right: recording equipment, master control, and the fixed transmission delay simulators.

APPENDIX

Design of the Servodrive for the Buffer Stores.

The gain and phase properties required in the servodrive loop are described here. Let Δx be the departure of the shuttle from perfect tracking. Then using capitals to denote Laplace transforms, $X(s)/\Delta X(s)$ is the open loop gain, G , of the servo in Fig. 6 and the closed loop response will then be

$$\frac{X(s)}{X(s) + \Delta X(s)} = \frac{G(s)}{1 + G(s)}$$

where

$$G(s) = \frac{G_0 Q(s)}{s(1 + sT_1)}$$

where

$G_0 = ab$ is the gain constant

a is the volts per unit error in x out of the phase detector

b is the shuttle velocity per volt into the servo amplifier
 T_1 is the mechanical time constant of the motor and drive mechanism.

The stabilizing filter has the form

$$Q(s) = \frac{(1 + sT')}{(1 + sT_2)(1 + sT_3)}$$

Where the phase advancing time constant T' is necessary to stabilize the control and the two retarding time constants ensure a falling gain characteristic at high frequencies. This falling gain is needed to reduce noise and to limit below 200 cps the band of signals fed to the 400 cps motor drive.

The loop can be written

$$\frac{G}{1 + G} = \frac{(1 + sT')}{1 + s\left(\frac{1}{G_0} + T'\right) + \frac{s^2}{G_0}(T_1 + T_2 + T_3) + \frac{s^3}{G_0}(T_1T_2 + T_2T_3 + T_3T_1) + s^4 \frac{T_1T_2T_3}{G_0}}$$

and at low frequency as s tends to zero this resembles a simple delay G_0^{-1} . It is concluded that a gain G_0 greater than 500 is needed to meet the delay specification of 2 milliseconds. But the condition for stability as given by Routh is

$$\frac{1}{G_0} + \frac{T_1T_2T_3G_0\left(T' + \frac{1}{G_0}\right)^2}{(T_1T_2 + T_2T_3 + T_3T_1)(T_1 + T_2 + T_3)} > \frac{T_1T_2 + T_2T_3 + T_3T_1}{T_1 + T_2 + T_3} - T'$$

Given $T_1 = 5$ msec and selecting $T' = 5$ msec, $T_2 = 1$ msec, $T_3 = 0.8$ msec, the system is marginally stable with $G_0 = 2250$ and in practice was satisfactory for applications using gains up to 10^3 .

Experimental measurements confirmed that the effective low frequency delay approximates $1/G_0$, it was also noted that tracking errors decreased approximately as $1/G_0$ when the amplifier gain was varied. This tracking error is less than ± 0.003 centimeter on the magnetostrictive wire with a loop gain of 10^3 .

REFERENCES

1. Candy, J. C. and Karnaug, M., Organic Synchronization: Design of Controls and Some Simulation Results, B.S.T.J. (next article in this issue), pp. 227-259.
2. Byrne, C. J., Karafin, B. J., and Robinson, D. B., Jr., Systematic Jitter in a Chain of Digital Regenerators, B.S.T.J., 42, November 1963, pp. 2679-2714.
3. Candy, J. C., Generating Rectangular-Waves, Phase Locked to a Repetitive Signal, Proc. IEEE, 54, August 1966, p. 1131.

Organic Synchronization: Design of the Controls and Some Simulation Results

By J. C. CANDY and M. KARNAUGH

(Manuscript received October 24, 1967)

Organic synchronization is a method for phase-locking the signals of an extensive digital communication network. Each clock in the network is made to depend on the phase drift of signals arriving and departing its station. This work demonstrates the practicality of such schemes. A four station simulation operating in real time with realistic parameters for a transcontinental network is used to evaluate various types of linear and nonlinear controls and to study effects of changing clock frequency and transmission delays. Considerable attention is given to the analysis of linear organic systems in order to pave the way for reasonable choices of design parameters and to make the results more easily understandable. The experiments show that the systems are very stable and easy to implement. No difficulty was experienced in starting the systems or in modifying their structures and they were immune to large scale breakdown caused by local faults.

I. INTRODUCTION

A model for certain mutually-synchronized systems of clocks and transmission links has been described by M. Karnaugh.¹ He calls systems that conform to this model "organic systems." The work on organic synchronization has been motivated chiefly by a desire to synchronize the sampling and switching operations in a geographically widespread pulse code modulation communication network.

Broad sufficient conditions for the stability of nonlinear organic systems have not yet been mathematically established. Nevertheless, there is reason to believe that systems having readily achievable clock stabilities and transmission delays of terrestrial magnitude can be well behaved. Because of this, and because performance under a variety of starting conditions, parameter choices, and perturbations is of interest, an analog simulator for organic synchronization has been constructed and put to use.

Details of the simulator hardware are described in the preceding article.² The present discussion will center on what has been learned from experimenting with the simulator. Preliminary attention is given to the analysis of linear organic systems; this aids the selection of reasonable control parameters for the design and study of some representative systems.

The experiments have supported the conjecture that systems of continental, or even global, dimensions will be stable and easy to implement. No great difficulty in starting the systems or in modifying their structures has been encountered.

II. SYSTEM ORGANIZATION

2.1 Discussion

The simulator contains four oscillators. Each of these corresponds to the local clock at a geographically distinct switching station. These stations may be interconnected through selected delay lines in any or all of the twelve possible directed paths. The sinusoids transmitted through these paths correspond in period to the data frames in a pulse code modulation communication system.

Each transmission path includes a large, fixed delay and a small, continuously variable, delay. The latter is used to simulate the slow variations in transmission delay which may occur in cables. In addition, there is another continuously variable delay line at the receiving terminal of each path. This simulates the buffer store which is needed to retim all data arriving at a switching station. The buffer store must synchronize the incoming data frames with the local switching actions. The latter are timed by the local clock.

For example, consider the two stations illustrated in Fig. 1 which shows a signal transmitted from station j to station i . The arriving signal is held in close phase agreement with the i th clock by the servoloop that automatically adjusts the buffer delay. Intuitively, we see that a constant frequency difference between the two clocks would drive the buffer at constant speed to one end, where synchronism would be lost. To prevent such failure, the clock frequencies are controlled by voltages derived from the position of the buffer, as shown. Gain factors a and b are placed in the frequency control paths.

A complete network is composed of many links resembling the one we have described. When the frequency of the clock at each station depends only upon the states of the buffers in transmission links arriving at that station, $b = 0$, and the controls will be called "one-sided."

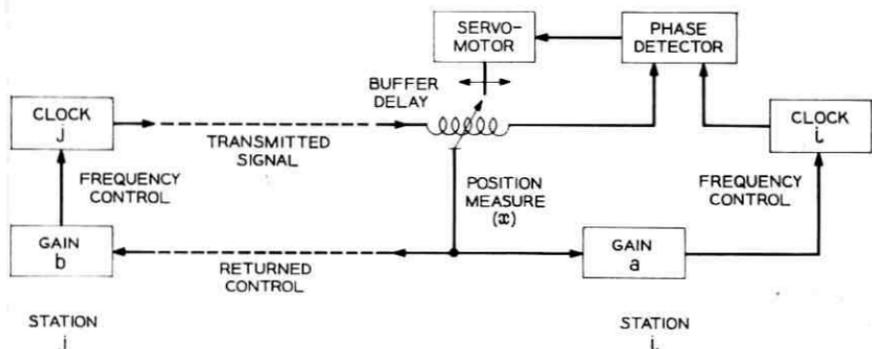


Fig. 1 — A transmission from station j to station i .

When the clocks are also controlled by the states of buffers in links leaving their stations, $b \neq 0$, the controls will be called "two-sided." Two-sided controls require the transmission of narrowband control signals between nearest neighbors in the system. In the simulator, separate baseband delays are used for this purpose.

The simulator incorporates filters, h , in the control paths to the clocks. They are used to explore the possibility of shaping the dynamic response of the network. Also included in the control paths is an amplitude limiter, ρ , which places a bound on the frequency deviations. Another nonlinearity, v , is placed in series with the buffer-position output, x . This causes the effective gain to vary with buffer position, so that the control influence of buffers near overflow may be made greater than those near their center position.

2.2 Analysis

Karnaugh has analyzed one-sided controls to determine the settling state after switching on from specified initial values. Now we shall examine changes in the settling conditions resulting from disturbances of transmission delays and clock frequencies, with two-sided controls. The model used by Karnaugh,¹ with a very slight change in notation, is illustrated in Fig. 2. Mathematically,

$$f_i(t) = F_i + \rho \left[h_i(t) * \left\{ \sum_{j=1}^N a_{ij} v[x_{ij}(t)] - \sum_{j=1}^N b_{ij} v[x_{ij}(t - \bar{\tau}_{ij})] \right\} \right] \quad (1)$$

That is, the frequency of the i th clock equals its natural frequency F_i plus a control function of x , the buffer states at the i th station and its nearest neighbors.

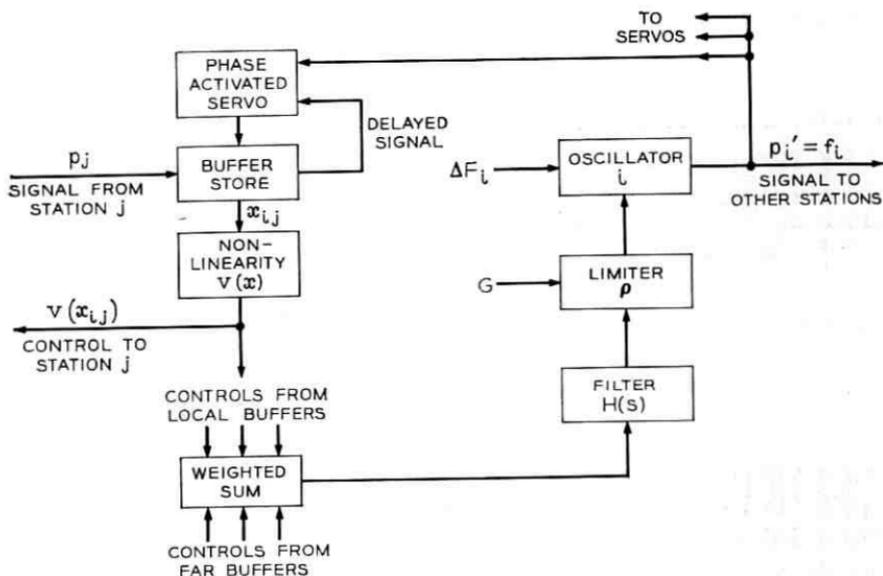


Fig. 2—Station i and its control with respect to station j .

The function $\rho(\cdot)$, which represents a limiter, is defined by

$$\begin{aligned} \rho(x) &= x, & \text{if } |x| \leq G \\ &= G, & \text{if } x > G \\ &= -G, & \text{if } x < -G \end{aligned} \quad (2)$$

It is stipulated that the control filters, h , have unity gain for dc:

$$H_i(0) \equiv \int_0^{\infty} h_i(t) dt = 1, \quad i = 1, 2, \dots, N \quad (3)$$

where $h_i(t)$ is the impulse response and $H_i(0)$ is the dc response. Mathematical convolution is indicated by “*.”

The gain coefficients a_{ij} , b_{ij} are nonnegative values. They are both equal to zero when there is no transmission link to station i from station j . When the link is present, $a_{ij} > 0$ and, if and only if the controls are two-sided, $b_{ij} > 0$.

The function $v(\cdot)$ is monotonic and of odd symmetry. It is also assumed that

$$v(\pm 1) = \pm 1. \quad (4)$$

Simple examples of such functions are

$$v_n(x) = x^{2n+1}, \quad n = 0, 1, 2, \dots \quad (5)$$

This is linear in the special case, $n = 0$.

The instantaneous state of the buffer delay in the link to center i from station j is denoted by $x_{ij}(t)$. Let the number of clock cycles stored in this buffer be $y_{ij}(t)$, and let its capacity be $2D_{ij}$. Then

$$x_{ij}(t) = [y_{ij}(t) - D_{ij}]D_{ij}^{-1}. \quad (6)$$

This represents the fractional deviation of the buffer delay from its half capacity. For synchronism to exist the buffers must lock the phase of the incoming signal to that of the local clock.

$$x_{ii}(t) = D_{ii}^{-1}[p_i(t - \tau_{ii}) - p_i(t)] + C_{ii}. \quad (7)$$

Where $p_i(t)$ is the phase of the i th clock, therefore

$$p'_i(t) = f_i(t). \quad (8)$$

The constant C_{ii} is determined by the initial switch-on conditions.¹

τ_{ij} is the transmission delay in the link to station i from station j . $\bar{\tau}_{ij}$ is the delay in sending control signals to station i from station j . It might or might not be approximately true that the two delays are equal when both are defined, but this has been true of the simulations that have been done.

Now assume that the system has settled down to a common frequency f and that the phase differences remain finite. Thereafter

$$p_i(t) = ft + r_i \quad (9)$$

where r_i is independent of time. Also, because the buffers are quiescent

$$x_{ii}(t + \tau) = x_{ii}(t) \quad \text{for all } \tau > 0$$

and

$$h_i(t)*x = x$$

so equation (1) becomes time independent,

$$f = F_i + \rho \sum_{i=1}^N \{a_{ij}v(x_{ij}) - b_{ij}v(x_{ij})\}, \quad (10)$$

and (7) becomes

$$x_{ii} = D_{ii}^{-1}[r_i - r_i - f\tau_{ii}] + C_{ii}, \quad (11)$$

which is also time independent.

For the purpose of investigating the changes in settling frequency of a system, F_i and τ_{ij} will be considered variables, and such variations are possible in the simulator. It is also possible to perturb the simulator by connecting or disconnecting transmission links, or by adding or removing an entire station.

III. LINEAR SYSTEMS

3.1 Analysis

The linear subclass of systems is of very special importance. Study of these systems has added much to our knowledge of organic synchronization.

One-sided linear systems have been very extensively studied. V. E. Beneš, in unpublished work, was the first to establish a very interesting sufficient condition for stability of these systems, and to study their settling frequency. A. Gersho and B. J. Karafin³ have recently simplified the derivation and proof of these results. M. B. Brilliant⁴ has shown that the requirement for network connectedness can be weakened, if master-slave relations between subsystems are permitted. He also has determined dynamic responses for some networks.⁵ Karnaug has formulated a model that makes explicit the influence of the starting conditions in the formulas for settling frequency.¹

Here, we shall extend the analysis to two-sided controls, considering the network to be switched on and in a quiescent state and then determining the change of quiescent conditions resulting from certain parameter changes. The result is a linear expression for the small changes in frequency that result from small changes in delay and in natural frequencies of the clocks.

The model defined in the previous section is linearized by removal of the limiter and by assuming $v(x)$ linear for the small changes. Total derivatives of equations (10) and (11) show how the settling state defined by the frequency f and the buffer states x is affected by changes. Thus for small changes, Δ , from a given quiescent state we have

$$\Delta f = \Delta F_i + \sum_{i=1}^N \{a_{ij}v'(x_{ij}) \Delta x_{ij} - b_{ij}v'(x_{ij}) \Delta x_{ij}\} \quad (12)$$

and

$$\Delta x_{ij} = D_{ij}^{-1}[\Delta r_i - \Delta r_i - \Delta(f\tau_{ij})]. \quad (13)$$

Now use

$$\alpha_{ij} = a_{ij} D_{ij}^{-1} v'(x_{ij})$$

and

$$\beta_{ji} = b_{ji} D_{ji}^{-1} v'(x_{ji})$$

which are assumed constants for a given settling state. Also use

$$g_i = \sum_{j=1}^N (\alpha_{ij} + \beta_{ji})$$

and

$$\tau_i = \sum_{j=1}^N (\alpha_{ij} \tau_{ij} - \beta_{ji} \tau_{ji})$$

to get

$$\Delta f = \Delta F_i + \sum_{j=1}^N (\alpha_{ij} + \beta_{ji}) \Delta r_j - g_i \Delta r_i - \Delta(f\tau_i).$$

Now measure phases with respect to the first station

$$q_i = p_i - p_1 = r_i - r_1, \quad \Delta q_i = \Delta r_i - \Delta r_1.$$

Then

$$\Delta f + \sum_{j=2}^N \{g_i \delta_{ij} - (\alpha_{ij} + \beta_{ji})\} \Delta q_j = \Delta F_i - \Delta(f\tau_i)$$

where δ_{ij} is the Kronecker delta

$$\delta_{ii} = 1, \quad \delta_{ij} = 0 \quad \text{for } i \neq j.$$

The solution of equation (19) by determinants, d , can be carried out to yield, after some simplification and index permutation,

$$\Delta f = \frac{\sum_{i=1}^N d_i \Delta F_i}{\sum_{i=1}^N d_i} - \frac{\sum_{i,j=1}^N (d_i \alpha_{ij} - d_j \beta_{ji}) \Delta(f\tau_{ij})}{\sum_{i=1}^N d_i}.$$

This equation is not solved for Δf explicitly because the term $\Delta(f\tau_{ij})$ depends on Δf , however this result is a convenient one for practical use. Let us investigate its meaning.

By definition, $f\tau_{ij}$ is the phase delay in the transmission link connecting center j to center i . Therefore equation (20) expresses the change of settling frequency as a weighted sum of the changes in

clock natural frequencies, ΔF , less a weighted sum of the change in phase delays during transmission, $\Delta(f\tau_{ij})$.

The usefulness of the result derives from the fact that in most practical networks the phase delay variations are determined by changes of delay alone, variations resulting from changes of frequency, $\tau\Delta f$, are usually negligible, that is

$$\Delta(f\tau_{ij}) \cong f \Delta\tau_{ij}$$

and this is true for the simulations.

Returning to the general result, for small changes we can write

$$\Delta(f\tau_{ij}) = f \Delta\tau_{ij} + \tau_{ij} \Delta f. \quad (21)$$

Then equation (20) becomes

$$\Delta f = \frac{\sum_{i=1}^N d_i (\Delta F_i - f \Delta\tau_i)}{\sum_{i=1}^N d_i (1 + \tau_i)}. \quad (22)$$

This is similar to the result given by Beneš. It demonstrates the surprising property that the sensitivity of the network to changes is small when the delays are large. However, the effect has only theoretical interest because in practice $\tau_i \ll 1$.

Each weighting coefficient, d_i , is the cofactor of the element in the i th row, first column, of the matrix $[\Gamma - \lambda]$, defined by

$$\lambda_{ij} = \alpha_{ij} + \beta_{ji}$$

$$\Gamma_{ij} = g_i \delta_{ij}.$$

It can be shown, following Brilliant,⁴ that $d_i > 0$ when

$$\left[\sum_{n=1}^N \lambda^n \right]_{ji} > 0 \quad \text{for all } j \neq i$$

and $d_i = 0$ otherwise. The criterion for positive d_i is, heuristically speaking, that a chain of transmission links of λ shall run from station i to every other station. If the i th station provides a master clock for the system, $d_i > 0$ and $d_j = 0$, $j \neq i$. The strongly connected systems of Beneš, on the other hand, have $d_i > 0$ for all i . The system cannot be synchronized unless at least one of these coefficients is positive.

The solution, (22), of the linear equations depends upon their being nonsingular. This is equivalent to the nonvanishing of the denominator. This condition will be seen to hold in systems of practical interest.

3.2 The Expression for Settling Frequency

To give direction to the simulation experiments, we summarize the salient properties of equation (20) in a set of rules that can be readily confirmed experimentally without need for solving determinants. They are also important because they describe conditions that might be used in a real network.

Rule 1

In the case of two-sided controls that are "proportional" so that $K_i\alpha_{ji} = K_j\beta_{ji}$ for all $i \neq j$, where $(K_1K_2 \cdots K_N)$ is a set of N positive numbers, then it can be shown that

$$\frac{K_1}{d_1} = \frac{K_2}{d_2} = \frac{K_N}{d_N}$$

and that

$$\Delta f = \frac{\sum_{i=1}^N K_i \Delta F_i}{\sum_{i=1}^N K_i} \quad (23)$$

The settled frequency is a weighted average of the oscillator center frequencies and is independent of delay.

Rule 2

In the particular case of "balanced" two-sided control defined by $\alpha_{ji} = \beta_{ji}$ for all $i \neq j$,

$$\Delta f = \frac{1}{N} \sum_{i=1}^N \Delta F_i \quad (24)$$

The settled frequency of a balanced two-sided organic system is always the unweighted average of the clock center frequencies.

Rule 3

In the case of a reciprocal control defined by $K_i(\alpha_{ji} + \beta_{ji}) = K_j(\alpha_{ji} + \beta_{ji})$ for all $i \neq j$, it can be shown that

$$\frac{K_1}{d_1} = \frac{K_2}{d_2} = \frac{K_N}{d_N}$$

and that

$$\Delta f = \frac{\sum_{i=1}^N K_i \Delta F_i - \sum_{i,j=1}^N \{K_i \alpha_{ij} - K_j \beta_{ji}\} \Delta(\tau_{ij})}{\sum_{i=1}^N K_i} \quad (25)$$

Having reciprocal control requires that the controls also be linear, $v(x) = x$. One sided controls, $\beta = 0$, can be reciprocal if $K_i \alpha_{ij} = K_j \alpha_{ji}$.

Rule 4

In the case of proportional control, as in rule 1, if a transmission link is symmetrical, $\tau_{ij} = \tau_{ji}$ then, in response to a change $\Delta\tau_{ij} = \Delta\tau_{ji}$

$$\Delta x_{ij} = D_{ij}^{-1} \Delta\tau_{ij} f \quad \text{and} \quad \Delta x_{ji} = D_{ji}^{-1} \Delta\tau_{ij} f$$

and there is no change in the position of any other buffer in the network.

Rule 5

The net phase delay around a closed loop in a synchronized network is a constant integer.

To prove this we notice that the net phase delay in link ij is

$$\phi_{ij} = D_{ij}(1 + x_{ij}) + \tau_{ij} f. \quad (26)$$

Therefore

$$\Delta\phi_{ij} = D_{ij} \Delta x_{ij} + \Delta(\tau_{ij})$$

using (13)

$$\Delta\phi_{ij} = \Delta r_j - \Delta r_i. \quad (27)$$

Then around a closed loop in the network the net phase delay will be

$$\begin{aligned} \Delta\phi_{ij} + \Delta\phi_{jk} + \cdots, \Delta\phi_{ki} &= \Delta r_j - \Delta r_i + \Delta r_k - \Delta r_j, \cdots, + \Delta r_i \\ &= 0. \end{aligned}$$

The definition of synchronization requires that the net phase delay be an integer. Notice that changes in the phase delays must be given a direction. That is, changes in delays introduced into signals flowing in one direction around the loop have the opposite sign from changes in delays introduced in signals flowing in the other direction.

IV. DYNAMIC RESPONSE AND STABILITY

In the foregoing analysis, stability was tacitly assumed and the actual form of the dynamic response to the parameter changes was ignored. Here we shall explore conditions for stability of two-sided controls for a linear network having no limiter. The clock frequencies will be considered variables but for convenience we shall assume transmission delays are fixed.

The transient response and stability of linear organic systems may be studied by means of the Laplace transforms of equations (1) and (7). Combining these equations and putting

$$r_i(t) = p_i(t) - p_i(0) \quad (28)$$

we obtain the result

$$sR_i(s) = H_i(s) \sum_{j=1}^N (\hat{\alpha}_{ij} + \hat{\beta}_{ji}) R_j(s) - H_i(s) R_i(s) \sum_{j=1}^N (\alpha_{ij} + \bar{\beta}_{ji}) + V_i(s) \quad (29)$$

where capitals denote Laplace transforms, and

$$\hat{\alpha}_{ij} = \alpha_{ij} e^{-s\tau_{ij}} \quad (30)$$

$$\hat{\beta}_{ji} = \beta_{ji} e^{-s\tau_{ji}} \quad (31)$$

$$\bar{\beta}_{ji} = \beta_{ji} e^{-s(\tau_{ji} + \tau_{ii})} \quad (32)$$

$$V_i(s) = \frac{F_i(s)}{s} + C_i(s) \quad (33)$$

and $C_i(s)$ represents initial conditions of the network. Following Gersho and Karafin,³ we notice that equation (29) may be put in the form

$$R_i(s) = B_i(s) \sum_{j=1}^N \frac{\hat{\alpha}_{ij} + \hat{\beta}_{ji}}{g_i} R_j(s) + \frac{B_i(s)}{g_i H_i(s)} V_i(s), \quad (34)$$

where

$$B_i(s) \equiv \frac{g_i H_i(s)}{s + H_i(s) \sum_{j=1}^N (\alpha_{ij} + \bar{\beta}_{ji})} \quad (35)$$

Let M be the matrix whose (ij) th element is

$$M_{ij} \equiv B_i(s) \frac{\hat{\alpha}_{ij} + \hat{\beta}_{ji}}{g_i}, \quad (36)$$

and let Q be the diagonal matrix with

$$Q_{ii} \equiv \frac{B_i(s)}{g_i H_i(s)}. \quad (37)$$

The system is now seen to obey the vector equation

$$R(s) = [I - M]^{-1} QV(s). \quad (38)$$

When $V(s)$ is specified, we can compute the solution $R(s)$ from this relation.

Arguments which differ from those of Gersho and Karafin only in minor detail can now be used to establish a sufficient condition for the stability of connected systems. The proof will not be repeated here. The resulting condition is that

$$|B_i(s)| < 1 \quad \text{for } s = \omega\sqrt{-1} \neq 0 \quad i = 1, 2, \dots, N, \quad (39)$$

where ω is real.

$B_i(s)$ is independent of the system delays when the controls are one-sided. This is not the case for two-sided controls, as one may see from equations (32) and (35). However, the sufficient condition may be checked for any particular network and is easily satisfied in practice.

M. B. Brilliant has simplified (39) for a simple but revealing case:

$$H(s) = 1, \quad g_i = g, \quad \tau_{ij} + \bar{\tau}_{ji} = \tau \quad \text{for all } i \neq j.$$

The sufficient condition for stability is then

$$g\tau < 0.5. \quad (40)$$

Thus, the largest product of the delay and the gain can, if it is less than 0.5, guarantee stability of the unfiltered two-sided linear network.

V. GAIN AND BUFFER DELAY

It will be shown later that, for one-sided controls, limiting of the frequencies can cause synchronization failures when the transmission delays are changing. Therefore, we are motivated to use small enough values of gains, g_i , to avoid limiting.

When there are no filters, or only single-pole filters, equations (1), (2), (4), and (6) show that limiting will not occur if

$$\sum_{i=1}^N (a_{ii} + b_{ii}) < G. \quad (41)$$

Notice that G would be made equal to the maximum tolerable frequency deviation. For simplicity, suppose that

$$g_i = g \quad \text{for } i = 1, 2, \dots, N$$

and

$$D_{ij} = D \quad \text{for all } (i, j).$$

Then the inequality, (41), becomes

$$gD < G. \quad (42)$$

Now consider the buffer size. The buffers deflect to compensate for delay changes and to correct oscillator drift. Let the greatest transmission delay variation from midrange be denoted by $\hat{\Delta}\tau$ and let the maximum error magnitude of the oscillator center frequency be denoted by $\hat{\Delta}F$ so that

$$\hat{\Delta}F = \max_i |\Delta F_i|.$$

Then we require

$$D > F \hat{\Delta}\tau + \frac{\hat{\Delta}F}{g} \quad (43)$$

and we wish to keep the buffer sizes small by making g large. Then most of the buffer delay capacity is used to compensate for transmission delay variations

$$F \hat{\Delta}\tau > \frac{\hat{\Delta}F}{g}. \quad (44)$$

Inequalities (44), (43), and (42) lead to

$$\frac{\hat{\Delta}F}{F \hat{\Delta}\tau} < g < \frac{G}{D} < \frac{G}{F \hat{\Delta}\tau + \frac{\hat{\Delta}F}{g}}. \quad (45)$$

Values of D that satisfy the last two inequalities in (45) will exist, provided that

$$g < \frac{G}{F \hat{\Delta}\tau + \frac{\hat{\Delta}F}{g}}$$

whence,

$$\frac{\hat{\Delta}F}{F \hat{\Delta}\tau} < g < \frac{G - \hat{\Delta}F}{F \hat{\Delta}\tau}. \quad (46)$$

Inequalities (46) are satisfied over a positive interval of values of g whenever

$$G > 2 \hat{\Delta}F. \quad (47)$$

Finally, D must be chosen in the interval,

$$F \hat{\Delta}\tau + \frac{\hat{\Delta}F}{g} < D < \frac{G}{g}. \quad (48)$$

For some specific examples, let us make the following choices:

$$G = 4 \hat{\Delta}F$$

$$g = \frac{2 \hat{\Delta}F}{F \hat{\Delta}\tau}$$

$$D = \frac{7}{4} F \hat{\Delta}\tau.$$

Let us further assume the largest single transmission delay to be 2×10^{-2} seconds, which is of transcontinental magnitude, and let $\hat{\Delta}\tau = 2 \times 10^{-5}$ seconds. The latter is almost surely an overestimate for underground coaxial cable, but is reasonable if about 10 per cent of the cable is above ground.

Two values of F will be used, corresponding roughly to voice and video sampling rates. For each of these, two values of $\hat{\Delta}F$ will be used, corresponding roughly to the accuracies of simple crystal oscillators and atomic oscillators. Table I shows the resulting parameters.

The greatest value of g encountered in Table I is 3×10^{-2} . Let us examine the consequences of this with respect to the stability condition, equation (40). The product τg is less than 12×10^{-4} for all cases. Therefore in satisfying the stability condition $\tau g < 0.5$ we have a factor of 400 to spare.

From (35), we see that $B_i(s)$ is the transfer function for the i th clock with respect to equal perturbations of phase in all arriving

TABLE I—Parameters for Voice and Video Sampling

F	$\hat{\Delta}F$	$\hat{\Delta}\tau$	G	g	D
(Hz)	(Hz)	(seconds)	(seconds ⁻¹)		(cycles)
10^4	3×10^{-3}	2×10^{-5}	12×10^{-3}	3×10^{-2}	3.5×10^{-1}
10^4	10^{-5}	2×10^{-5}	4×10^{-5}	10^{-4}	3.5×10^{-1}
6×10^6	18×10^{-1}	2×10^{-5}	7.2×10^0	3×10^{-2}	2.1×10^2
6×10^6	6×10^{-3}	2×10^{-5}	2.4×10^{-2}	10^{-4}	2.1×10^2

signals. In the absence of delays,

$$B_i(s) = \frac{g_i H_i(s)}{s + g_i H_i(s)}$$

Therefore, if $H_i(s)$ is either constant or is monotone decreasing with little phase shift in

$$s = \omega \sqrt{-1}, \quad 0 \leq \omega \leq g_i,$$

the response radian bandwidth is no more than g_i . When $H_i \equiv 1$, the response has the simple time constant, g_i^{-1} .

This enables us to interpret τg as the ratio of the delay between stations to their time constants of response. When this ratio is very small, the transmission delays have negligible effect on the dynamic response of the network. This is another strong reason for using relatively low gains.

Another useful approximation that applies to most practical networks concerns the phase delay during transmission $f\tau_{ij}$. The change of phase delay is

$$\Delta(f\tau_{ij}) = f \Delta\tau_{ij} + \tau_{ij} \Delta f. \quad (21)$$

We have seen that delays change about 0.1 per cent* while clock frequencies change about 10^{-7} , thus the second term in the expression is negligible in most cases and this is true for the simulations. Equation (20) then becomes

$$\Delta f \cong \frac{\sum_{i=1}^N d_i (\Delta F_i - f \Delta\tau_i)}{\sum_{i=1}^N d_i}. \quad (49)$$

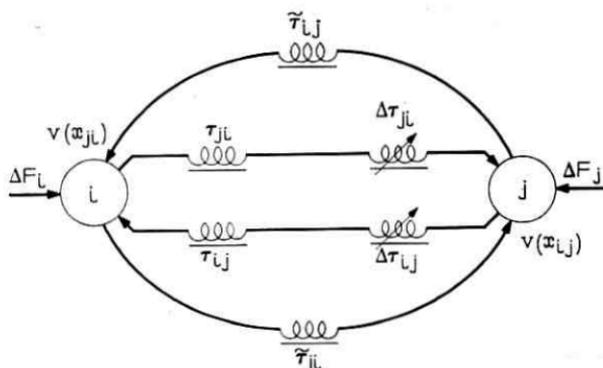
VI. A TWO STATION NETWORK

6.1 Response of the System to Small Change

Many revealing properties of the organic synchronizing scheme can be illustrated with two stations connected in a loop as in Fig. 3. Such a system is characterized by the following equations

$$\Delta f \cong \frac{g_i \Delta F_i + g_j \Delta F_j}{g_i + g_j} - \frac{(\alpha_{ij}\alpha_{ji} - \beta_{ij}\beta_{ji})f \Delta(\tau_{ij} + \tau_{ji})}{g_i + g_j} \quad (50)$$

* For short delays the fractional change can be much larger.

Fig. 3 — Two coupled stations, i and j .

$$\Delta x_{ii} \cong \frac{D_{ij}^{-1}}{g_i + g_j} [\Delta[F_i - F_i] + (\alpha_{ii} + \beta_{ii})f \Delta(\tau_{ii} + \tau_{ii})] \quad (51)$$

$$\Delta x_{ji} \cong \frac{D_{ji}^{-1}}{g_i + g_j} [\Delta[F_i - F_i] + (\alpha_{ii} + \beta_{ii})f \Delta(\tau_{ii} + \tau_{ii})].$$

Equation (50) was derived from (49), and equations (51) follow from (13), (18), (19), and (49).

Our simulation of this network has a nominal 1 megahertz center frequency with ± 1 hertz control range. The gains ($g_i + g_j$) are in the range 0.1 to 0.001 sec^{-1} . Transmission delay around the loop can be preset in the range 0 to 0.1 second and be varied continuously by ± 100 microseconds in each link.

In all setups the settling states agree well with prediction. They confirm that fixed delay has negligible effect (\ll one percent) on buffer and frequency deflections. Therefore the approximation in deriving (49) from (20) is justified.

For illustration, Table II gives some typical results. The data have been normalized to represent unit amplitude disturbances. The incremental delay changes, $\Delta\tau$ are expressed as a fraction z of the associated buffer capacity such that

$$z_{ij} = \frac{\Delta\tau_{ij}F}{D_{ij}}. \quad (52)$$

In the simulations the gain delay product τg is less than 10^{-2} . This should guarantee stability with reasonable valued filters, and indicates that fixed delays are too small to have appreciable effect on transients. Observations of responses after switch-on and of subse-

TABLE II—Settling States of Two Stations

Control Type	Gain Setting				Disturbance		Result		
	α_{ij}	β_{ij}	α_{ji}	β_{ji}	ΔF_i	Δz_{ji}	Δx_{ij}	Δx_{ji}	Δf
Balanced	1	1	1	1	1	0	-1/4	+1/4	1/2
	1	1	1	1	0	-1	1/2	1/2	0
	2	2	2	2	1	0	-1/8	+1/8	1/2
	2	2	2	2	0	-1	1/2	1/2	0
	2	2	1	1	1	0	-1/6	-1/6	1/2
	2	2	1	1	0	-1	1/3	2/3	0
Proportioned	2	1	1	2	1	0	-1/6	1/6	1/3
	2	1	1	2	0	-1	1/2	1/2	0
	1	2	2	1	1	0	-1/6	1/6	2/3
	1	2	2	1	0	-1	1/2	1/2	0
One Sided	1	0	1	0	0	-1	1/2	1/2	1/2
	2	0	2	0	1	0	-1/4	1/4	1/2
	2	0	2	0	0	-1	1/2	1/2	1
	2	0	1	0	0	-1	1/3	2/3	2/3

Unit gain = 10^{-2} sec^{-1}

Unit frequency = 1 Hz

Unit delay = $10^2 \mu\text{sec}$

Nominal clock frequency = 1 MHz

quent disturbances confirm this conclusion. Indeed, delay had insignificant effect when various one- and two-pole low pass filters were included in the control loops. The filters investigated had cut-off frequencies from 0.01 to 50 hertz and q -factors up to 10.

When fixed delays are neglected, the response of the linear system to small change can be expressed as

$$D_{ij}X_{ij}(s) \cong \frac{F_i(s) - F_j(s) - [\alpha_{ji}(s) + \beta_{ji}(s)][\tau_{ij}(s) + \tau_{ji}(s)]F}{[s + g_i(s) + g_j(s)]} \quad (53)$$

$$f(s) \cong \alpha_{ij}(s)X_{ij}(s) - \beta_{ji}(s)X_{ji}(s) \quad (54)$$

where

$$\alpha_{ii}(s) = \alpha_{ij}H_i(s), \quad \beta_{ij}(s) = \beta_{ij}H_j(s)$$

and

$$g_i(s) = g_iH_i(s).$$

When no filters are used these responses have a simple time constant $1/(g_i + g_j)$. Fig. 4 gives some typical response curves for a system with no filters and using various gain values. Including delays up to 0.1 second had no noticeable effect on these curves.

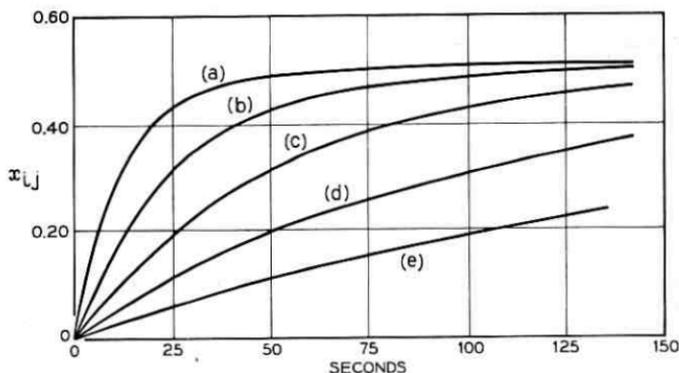


Fig. 4—Response of buffer x_{ij} to a step decrease in delay $\Delta\tau_{ij}$. The controls are balanced, and the net gain is such, that $(g_i+g_j)^{-1}$ is 12 seconds for curve a, 24 for b, 48 for c, 108 for d, and 216 for e.

Fig. 5 shows responses with similar one-pole low-pass filters in both oscillator circuits. Notice that use of filter time constants near $1/g$ speeds the response. We do not anticipate that filters will play an important role in the operation of the networks. We have seen that they are not needed to stabilize the controls, nor will they be needed to speed responses because most changes will be thermally induced, and will therefore occur very slowly. Some filtering may be needed

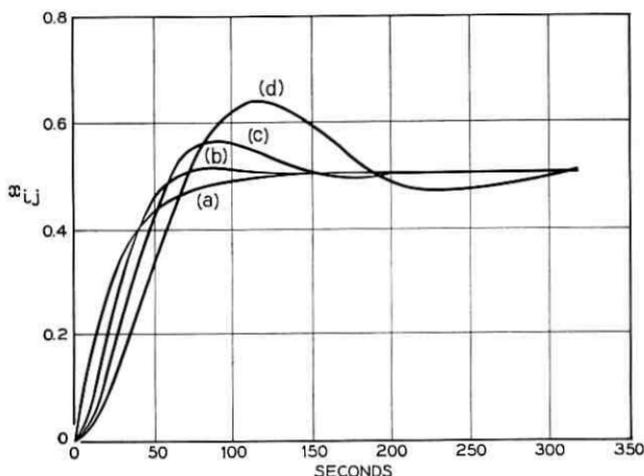


Fig. 5—Response of buffer x_{ij} to a step decrease in τ_{ij} with various filters $1/(1+s\tau)$. The controls are balanced and the net gain $(g_i+g_j) = 1/24 \text{ sec}^{-1}$. The filter time constants τ are 0 seconds for curve a, 10 for b, 20 for c and 40 for d.

to smooth the output of phase detectors and to reduce noise. For this, time constants not greater than $1/g$ should be adequate.

6.2 When Control Signals Limit

A saturating limiter ρ is placed in the control path to the oscillators in order to place a bound on the frequency deviations. To demonstrate its effect, Fig. 6 shows a buffer response to a ramp change of delay using various gain values in a symmetrical one-sided control system with no filters

$$\alpha_{ij}(s) = \alpha_{ji}(s) = g \quad \beta_{ij}(s) = \beta_{ji}(s) = 0.$$

At the start of the experiment the network is quiescent with the two clock natural-frequencies offset from one another and the two buffers deflected by amounts that provide sufficient control voltage to align the running frequencies.

$$\Delta x_{ij} = \frac{\Delta F_{ii}}{g_i} D_{ii}.$$

This initial buffer-displacement decreases proportionately with increased gain. For this purpose a large gain is attractive. However, when the transmission delay varies, limiting occurs if the gain is large, as shown in curve *c* of Fig. 6, where there is loss of control and ultimately loss of synchronization. Clearly, saturation must be avoided, that is, equation (41) satisfied, when one-sided controls are used.

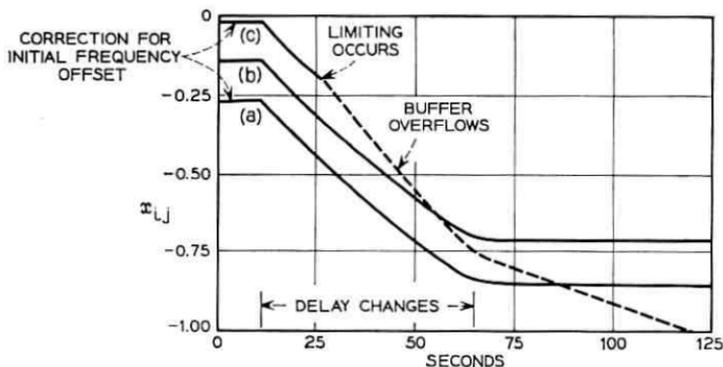


Fig. 6—Response of x_{ij} to similar ramp delay increases $\Delta\tau_{ij}$ and $\Delta\tau_{ji}$ in a system having initial frequency discrepancy. The controls are one-sided and the gains are such that g_i^{-1} and g_j^{-1} are both equal to 50 seconds for curve a, 25 for b, and 5 for c.

With two-sided "proportioned" control, changes of delay do not cause change of settling frequency. See rule 1. Any frequency changes that might occur during the dynamic response will usually be small because large parameter changes will occur slowly in a practical network. Thus with "proportioned" control, delay changes will not cause limiting, the signal in the limiters being determined only by the differences of oscillator center frequencies.

Larger gains can be used with two-sided control systems than with one-sided. However, in practice equation (41) should always be obeyed so that systems continue to operate reliably if faults interrupt the distant control paths. The reasons for avoiding limiting are equally valid for more complex networks, and this is not a hard restriction on design. It has already been seen in Table I how limiting can be avoided while using gains that satisfy the main requirements of a practical network.

6.3 Nonlinear Control

Use of nonlinearity in the control loop has been proposed¹ as a means for exaggerating the correcting influence of buffer stores that are near overflow, at the expense of those nearer center. For this purpose the nonlinear circuit $v(x)$ is included in Fig. 2. It makes the incremental gains a function of buffer position, as illustrated by equation (14). We shall examine separately the nonlinear response to incremental delay and incremental frequency change as given by equations (50) and (51).

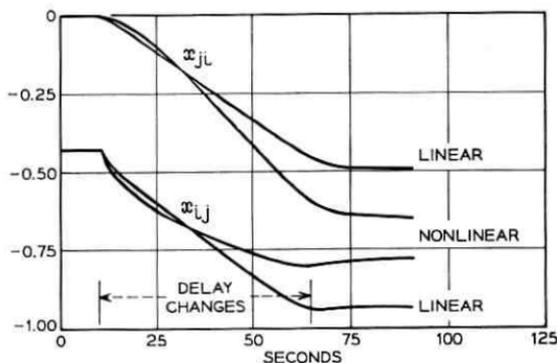


Fig. 7—The response of both buffers to a ramp increase in delay $\Delta\tau_{iJ}$ with linear and nonlinear controls. Controls are balanced with net $g = 0.04 \text{ second}^{-1}$ and $\mu = 1.2$.

When x_{ji} is nearer zero than x_{ij} the nonlinearity can be selected to make $|v'(x_{ji})|$ very small compared with $|v'(x_{ij})|$. Then $(\alpha_{ji} + \beta_{ji}) \ll (\alpha_{ij} + \beta_{ij})$. Therefore x_{ij} will approximately track the delay change and x_{ji} will have little change.

To illustrate this, Fig. 7 shows the response of both buffers to a steadily increasing delay commencing with $x_{ij} = 0$ and $x_{ji} \cong 0.5$. The nonlinearity successfully makes the buffer with the most reserve compensate for most delay change. In this and subsequent experiments

$$v(x) = \frac{x(\mu - 1)}{(\mu - x^2)}, \quad \mu > 1. \quad (55)$$

A useful method for demonstrating the response of nonlinear control is a locus on a graph of x_{ji} plotted against x_{ij} . Fig. 8 shows such loci for various starting conditions, and Fig. 9 is an enlargement of a section of Fig. 8 repeated for different values of nonlinearity. These results show that the nonlinearity can prevent overflow in some circumstances.

Next, consider the response to frequency change with constant delays. Equation (51) shows that when the deflections are expressed in cycles

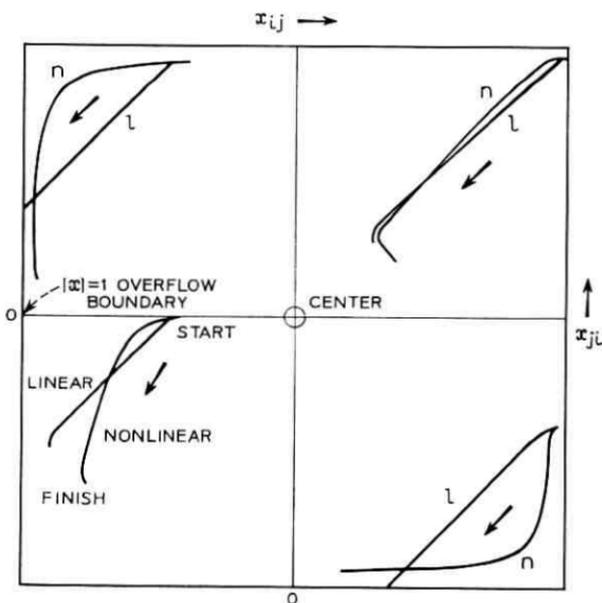


Fig. 8—Loci of x_{ji} against x_{ij} for a slow ramp change of delay $\Delta\tau_{ij}$ in a two-station network; l : balanced linear controls $g_i = g_j = 0.04 \text{ second}^{-1}$; n : balanced nonlinear controls $\mu = 1.2$ $g_i = g_j = 0.04 \text{ second}^{-1}$.

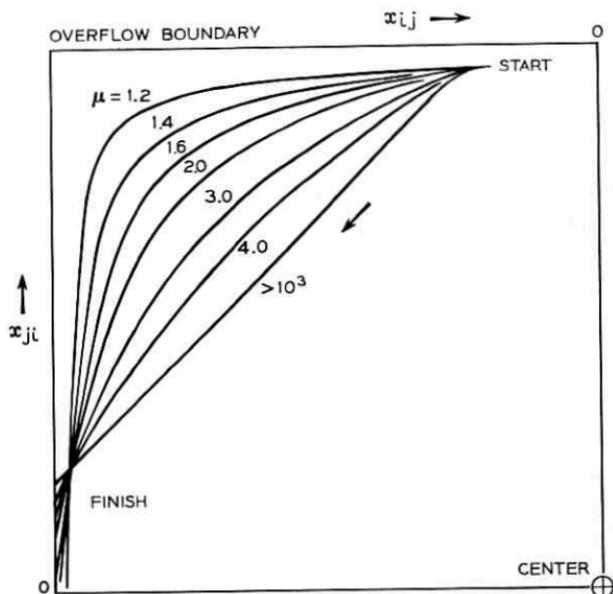


Fig. 9—An enlargement of the upper left quadrant of Fig. 8 for various values of nonlinearity.

they are the same at both stations. We need to minimize their value when a buffer is near overflow. Consider a worst case with $x_{ji} = 0$ and $x_{ij} \cong 1$. Then there is improvement over the linear control if the net gain $(g_i + g_j)$ exceeds the linear gain. That is, if

$$[v'(0) - 1](\alpha_{ji} + \beta_{ji}) + [v'(1) - 1](\alpha_{ij} + \beta_{ij}) > 0.$$

For example, when $(\alpha_{ji} + \beta_{ji}) \cong (\alpha_{ij} + \beta_{ij})$ there is improvement if $1 < \mu < 3$ using nonlinearities described by equation (48) or if $n > 2$ using nonlinearities described by equation (7). An illustration is given in Fig. 10.

The degree of nonlinearity used will be limited by the need to maintain some minimum gain for controls near center.

VII. A CHAIN OF FOUR STATIONS

The simulation is extended by adding two more stations to form the chain shown in Fig. 11. In designing controls for this network it seems sensible to make controls associated with each link reciprocal and symmetric, that is,

$$(\alpha_{ij} + \beta_{ji}) = (\alpha_{ji} + \beta_{ij}) \quad \text{for all } i \neq j.$$

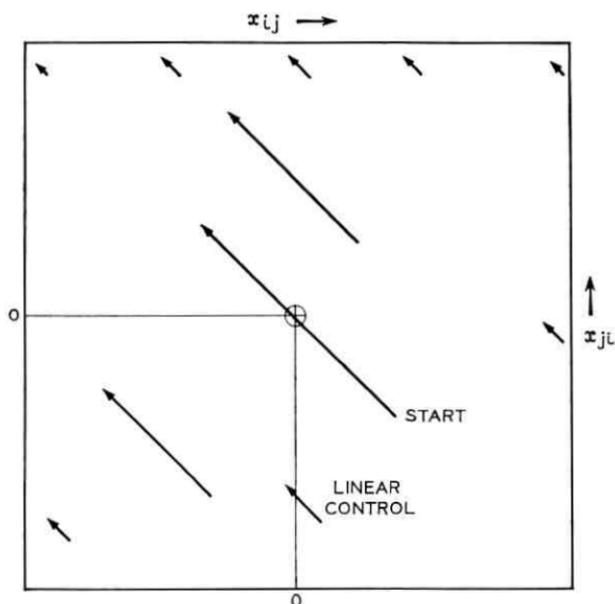


Fig. 10 — Loci of x_{ji} against x_{ij} for a step change in frequency ΔF_i in a two-station network with balanced nonlinear control $\mu = 1.2$.

Then with the buffers at center the settling frequency is equally sensitive to frequency drift in each oscillator. With such control the net gain, g , of the relay stations j and k exceeds that of the end stations i and l . The relay stations therefore have greater risk of limiting, but this risk is lessened by the larger possibility of averaging and canceling of disturbances.

Some typical settling states for this chain are given in Table III. The top of Fig. 12 shows how a sudden change of station i frequency causes transients through the system. The bottom shows the corresponding response with nonlinearities added. Notice the slowing down of the response and the increased deflections particularly at remote stations. These effects are a consequence of the low control gains when buffers are near center.

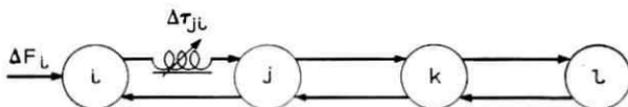


Fig. 11 — The four station chain.

TABLE III—Settling States of a Four Station Chain

Control Type	Gain Setting						Disturbance						Result					
	α_{ij}	β_{ji}	α_{ik}	β_{kl}	α_j	β_k	ΔF_i	ΔF_j	Δz_{ij}	Δx_{ij}	Δx_{jk}	Δx_{kl}	Δz_{jk}	Δz_{kl}	Δx_{jk}	Δx_{kl}	Δz_{kl}	Δf
Balanced	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	1	1	1	1	1	1	0	1	0	-1/4	-1/4	1/4	1/4	-1/8	-1/8	1/8	1/4	1/4
	1	1	1	1	1	1	0	0	-1	0	0	0	0	0	0	0	0	0
Proportioned	1	1	1	1	1/2	1/2	1	0	0	-1/4	-1/4	1/4	1/4	-1/12	-1/12	1/12	1/6	1/6
	1	1	1	1	1/2	1/2	0	1	0	-1/6	-1/6	1/2	1/2	-1/6	-1/6	1/6	1/3	1/3
	1	1	1	1	1/2	1/2	0	0	-1	0	0	0	0	0	0	0	0	0
One-sided Reciprocal	2	0	2	0	2	0	1	0	0	-1/4	-1/4	1/4	1/4	-1/8	-1/8	1/8	1/4	1/4
	2	0	2	0	2	0	0	1	0	-1/4	-1/4	1/4	1/4	-1/8	-1/8	1/8	1/4	1/4
	2	0	2	0	2	0	0	0	-1	0	0	0	0	-1/4	-1/4	1/4	1/2	1/2

Code: $\alpha_j = \alpha_{jl} = \alpha_{jk}$ $\beta_j = \beta_{ij} = \beta_{kj}$
 $\alpha_k = \alpha_{kl} = \alpha_{kj}$ $\beta_k = \beta_{ik} = \beta_{jk}$

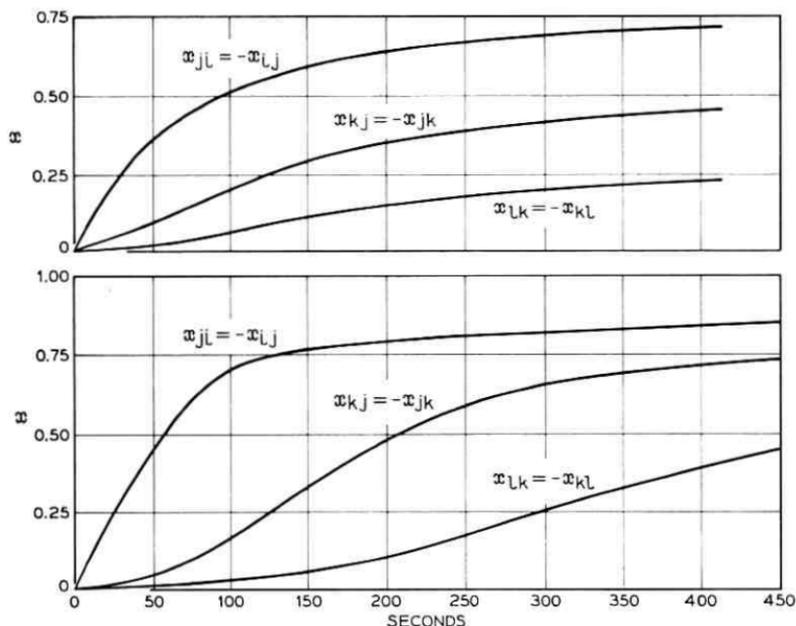


Fig. 12— Buffer responses to a step change of frequency ΔF_i in a four station chain. Top: linear balanced controls $\alpha = \beta = 0.02 \text{ second}^{-1}$. Bottom: nonlinear balanced controls $\alpha = \beta = 0.02 \text{ second}^{-1}$ $\mu = 1.2$.

However, we have said that frequency drift is small in practice, more significant are effects of delay change. These are demonstrated in Fig. 13 which shows loci of x_{ji} against x_{ik} for a delay change $\Delta \tau_{ji}$. For simplicity, buffer x_{ij} starts with the same content as x_{ji} . The content of x_{ji} and x_{ik} are apparent from the graph; all other buffers start at center. Curve *a* is the response with two-sided balanced control. For slow change, it is a vertical line whose shape is independent of position on the graph or of nonlinearity value. Curve *b* is the response with one-sided linear control. The others are for nonlinear one-sided controls.

VIII. VARIOUS CONNECTIONS OF FOUR STATIONS

Figs. 14, 15, and 16 show four stations connected as a loop, star, and complete network, respectively. The settling states of these three networks are illustrated in Tables IV, V, and VI. These results were virtually independent of fixed delay values and confirm the prediction of equation (49).

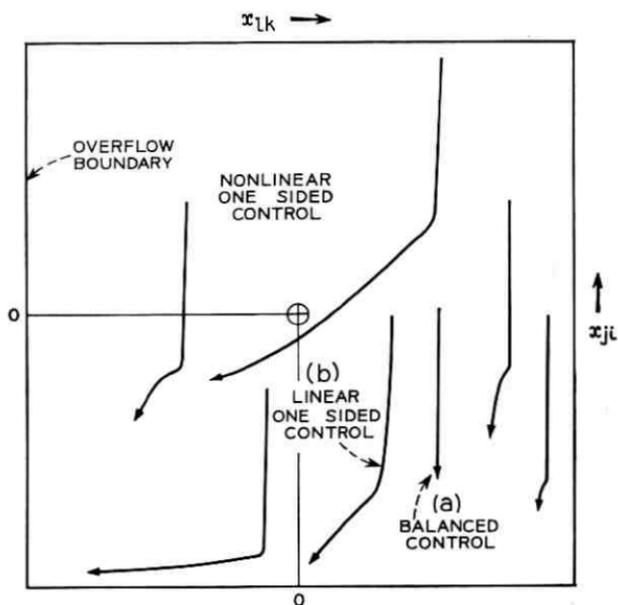


Fig. 13 — Loci of x_{jt} against x_{lk} for a slow ramp change of delay $\Delta\tau_{jt}$ in a four station chain.

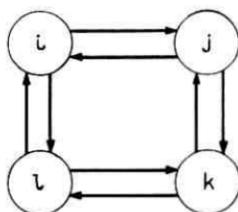


Fig. 14 — Ring network.

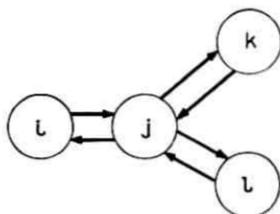


Fig. 15 — Star network.

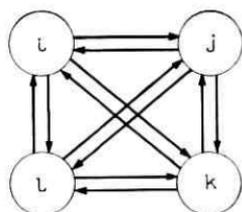


Fig. 16 — Complete network.

Fig. 17 gives some typical transient responses of the loop connection for both linear and nonlinear controls. Fig. 18 demonstrates the effect of nonlinearity on loci of x_{ji} to x_{kj} for various starting positions. A corresponding graph for the fully connected network is shown in Fig. 19.

In studying the responses of all these networks no noticeable effects were observed when delays up to 0.1 second were included in signal and control paths. The responses were always stable even with single pole filters in the control loop.

Obtaining proper statistical data on the response of the synchron-

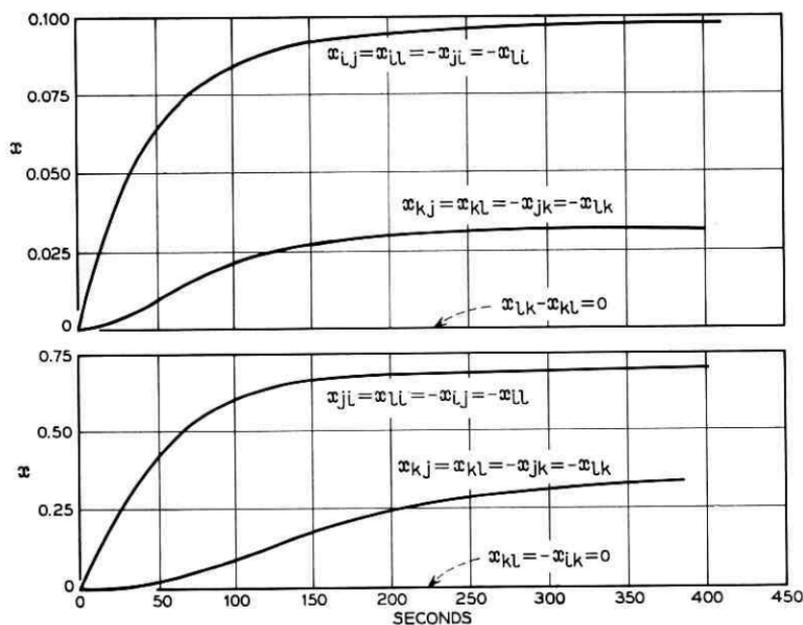


Fig. 17 — Buffer responses to a step change of frequency ΔF_t in a four station ring. Top: balanced linear controls $g = 0.012 \text{ second}^{-1}$. Bottom: nonlinear controls $\mu = 1.2g = 0.012 \text{ second}^{-1}$.

TABLE IV — Settling States of a Four Station Ring

Control Type	Gains		Disturbance			Result						
	α	β	F_i	z_{ji}	z_{ij}	Δx_{ij}	Δx_{ji}	$-\Delta x_{kj}$	$-\Delta x_{kl}$	$-\Delta x_{li}$	$-\Delta x_{li}$	Δf
Balanced	1	1	1	0	0	-3/16	3/16	1/16	-1/16	-3/16	1/4	
	1	1	0	-1	0	3/8	5/8	1/8	1/8	1/8	0	
	1	1	0	-1	-1	1	1	0	0	0	0	
One Sided Reciprocal	2	0	1	0	0	-3/16	3/16	1/16	-1/16	-3/16	1/4	
	2	0	0	-1	0	3/8	5/8	3/8	1/8	-1/8	1/2	
	2	0	0	-1	-1	1	1	1/2	0	-1/4	1/2	

ized network to disturbances is an extensive project because of the large number of interacting parameters available and because of the slow response of the system to change. To short cut this work we have examined tendencies in the response in some extreme conditions.

The fully connected network was disturbed by driving the delays and the oscillators individually with triangular waves having unrelated frequencies in the range 0.01 to 0.001 Hz. The maximum deflection of the twelve buffers and their likelihood of overflow under various control settings were observed.

The following conclusions were made from these experiments.

(i) The chance of buffers overflowing is decreased by use of increased control gain, provided limiting is avoided. Increasing the gain past the limiting point increased the chance of overflow.

(ii) The chance of overflow increases when limiting levels are decreased.

(iii) The chance of either limiting or overflow occurring is less with balanced control than with one-sided control.

(iv) Use of low pass filters with long time constants increases the chance of overflow.

(v) Use of nonlinearities ($\mu \cong 1.2$) in a balanced control system reduces the chance of overflow.

(vi) Use of nonlinearity with single ended control increases the chance of overflow.

(vii) Use of nonlinearity increases the chance of overflow when disturbances change at rates comparable with the control bandwidth.

IX. RESPONSES IN FAILURE

If a single transmission link is severed or a signal is lost for any reason the servo in Fig. 1 normally drives its buffer to an extremity and

TABLE V — Settling States of a Four Station Star

Control Type	Gain Setting			Disturbance			Result				
	α_j	β_j	$\frac{\alpha_{ij}}{\alpha_{kj}} \quad \frac{\beta_{ji}}{\beta_{kj}} \quad \frac{\alpha_{ij}}{\alpha_{lj}}$	ΔF_i	ΔF_j	Δx_{ji}	Δx_{ij}	Δx_{ji}	Δx_{jk}	Δx_{ij}	ΔF
Balanced	1	1	1	1	0	0	$-\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
	1	1	1	0	1	0	$\frac{1}{8}$	$-\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
	1	1	1	0	0	-1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
Proportion	$\frac{1}{3}$	$\frac{1}{3}$	1	1	0	0	$-\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
	$\frac{1}{3}$	$\frac{1}{3}$	1	0	1	0	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
	$\frac{1}{3}$	$\frac{1}{3}$	1	0	0	-1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
One Sided	2	0	2	1	0	0	$-\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
	2	0	2	0	1	0	$-\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
	2	0	2	0	0	-1	$\frac{1}{4}$	$\frac{3}{4}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{2}$
	1	0	3	0	0	-1	$\frac{1}{6}$	$\frac{5}{6}$	$-\frac{1}{6}$	$-\frac{1}{6}$	$\frac{1}{2}$

Code:

$$\alpha_j = \alpha_{ji} = \alpha_{jk} = \alpha_{jl}$$

$$\beta_j = \beta_{ij} = \beta_{kj} = \beta_{lk}$$

TABLE VI — Settling State of Four Stations Fully Connected

Control Type	Gains		Disturbance				Result							
	α	β	ΔF_i	Δs_{ij}	Δz_{ji}	Δx_{ij}	Δx_{ji}	$\frac{\Delta x_{ik}}{-\Delta x_{ki}}$	$\frac{\Delta x_{il}}{-\Delta x_{li}}$	$\frac{\Delta x_{jk}}{-\Delta x_{kj}}$	$\frac{\Delta x_{jl}}{-\Delta x_{lj}}$	$\frac{\Delta x_{ik}}{\Delta x_{ki}}$	$\frac{\Delta x_{il}}{\Delta x_{li}}$	Δf
Balanced	1	1	1	0	0	$-1/8$	$1/8$	$-1/8$	$-1/8$	0	0	0	0	$1/4$
	1	1	0	-1	0	$1/4$	$3/4$	$1/8$	$1/8$	$-1/8$	$1/8$	0	0	0
	1	1	0	-1	-1	1	1	0	0	0	0	0	0	0
One Sided Reciprocal	2	0	1	0	0	$-1/8$	$1/8$	$-1/8$	$-1/8$	0	0	0	0	$1/4$
	2	0	0	-1	0	$1/4$	$3/4$	0	0	$-1/4$	$-1/4$	0	0	$1/2$
	2	0	0	-1	-1	1	1	$-1/4$	$-1/4$	$-1/4$	$-1/4$	0	0	1

thus deflects the frequency of the whole network. In the simulator this is prevented by auxiliary equipment that takes the buffer to center in case of signal loss. Meanwhile, the remainder of the network can function correctly.

Another interesting fault occurs when a buffer is driven to overflow by the normal controls. In the simulator it rests at full deflection until the phase drift brings another frame within half a cycle of synchronism with the local clock, the system locks to that frame.

Little difficulty has been experienced in switching stations in and out of an active network. The switching transient can be reduced by adjusting the station frequency so that its buffers are near center.

X. CONCLUSION AND RECOMMENDATION

There seems to be no good reason for using nonreciprocal controls, moreover using linear-symmetric controls $(\alpha_{ij} + \beta_{ji}) = (\alpha_{ji} + \beta_{ij})$ makes the network equally sensitive to frequency drift in each oscillator (rule 3). Then if the oscillators have similar properties this minimizes the frequency displacement caused by oscillator drift.

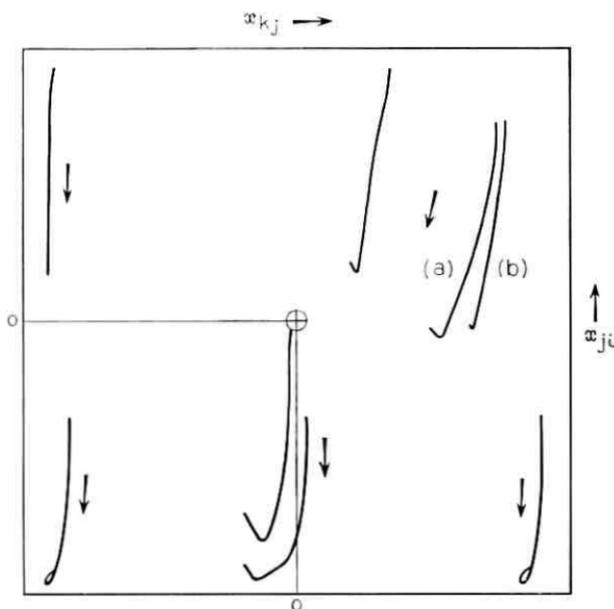


Fig. 18— Responses of buffers to a ramp change in delay $\Delta\tau_{ji}$ in a four station ring: (a) linear balanced control; (b) linear one-sided control. All other curves for nonlinear balanced control $\mu = 1.2$ starting with $x_{jk} = x_{kj}$, $x_{ij} = x_{ji}$ and all other buffers at center.

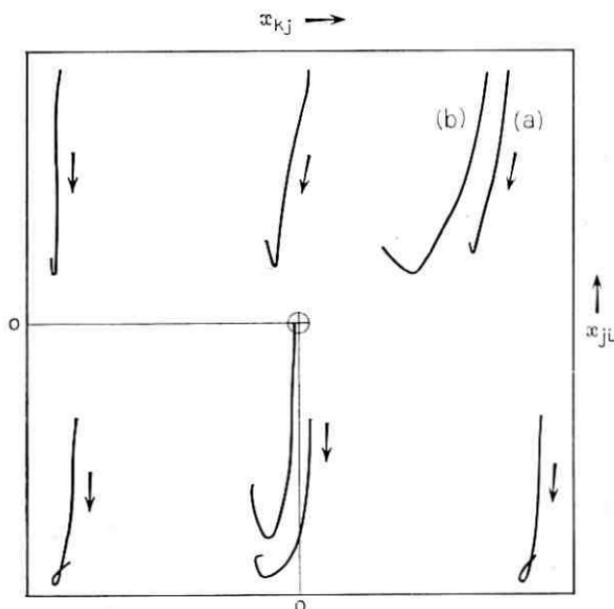


Fig. 19—Responses of buffers to a ramp change of delay $\Delta\tau_{ji}$ for the complete network with the same controls as in Fig. 18.

There is additional advantage in also balancing controls ($\alpha_{ij} = \beta_{ij}$). Then slow delay change will not affect the system frequency and the chance of buffers overflowing is reduced. The disadvantage of using two-sided, and hence also of balanced controls, is the need for control paths between centers. These paths require little bandwidth (< 1 Hz) but must pass dc; they could be included in the framing and signaling codes of a PCM system.

A commercially useful network will almost certainly contain both sending and receiving links in each path used, and these links will be similar in opposite directions. For example, they will expand and contract together so that $\Delta\tau_{ij} = \Delta\tau_{ji}$. Rule 4 shows that if the controls in such a network are proportional then each delay change is corrected directly by the buffer in its link and no disturbance propagates through the network.

If the buffers in such a symmetric path start at corresponding positions they will approximately track one another, deviating only to correct small frequency drift and the misbalance of the links. There will be little use for nonlinear shaping to equalize their positions, besides, nonlinearities lower sensitivity when buffers are near center

and increases the practical difficulty of keeping controls proportional.

Use of symmetric links and linear balanced controls provide for near optimum design of buffers. Each one corrects only the delay change of its own link and the small frequency drifts. They can be designed independently of the network.

XI. ACKNOWLEDGMENTS

The simulator used for the experiments described here was designed and built in cooperation with the authors of the preceding article that describes the hardware.² We are especially grateful to M. B. Brilliant for an extremely painstaking review and many valuable and poignant comments.

REFERENCES

1. Karnaugh, M., A Model for the Organic Synchronization of Communications Systems, *B.S.T.J.*, *45*, December 1966, pp. 1705-1735.
2. Bosworth, R. H., Kammerer, F. W., Rowlinson, D. E., and Scattaglia, J. V., Design of a Simulator for Investigating Organic Synchronization Systems, *B.S.T.J.*, this issue, pp. 209-226.
3. Gersho, A. and Karafin, B. J., Mutual Synchronization of Geographically Separated Oscillators, *B.S.T.J.*, *45*, December 1966, pp. 1689-1704.
4. Brilliant, M. B., The Determination of Frequency in Systems of Mutually Synchronized Oscillators, *B.S.T.J.*, *45*, December 1966, pp. 1737-1748.
5. Brilliant, M. B., Dynamic Response of Systems of Mutually Synchronized Oscillators, *B.S.T.J.*, *46*, February 1967, pp. 319-356.

Contributors to This Issue

RAYMOND H. BOSWORTH, graduate, Air Force Radio Technical School, 1950; attended Union Junior College, 1952-56; R.C.A. Institutes 1962-64; Bell Telephone Laboratories, 1952—. Mr. Bosworth has worked on negative impedance repeaters, repertory dial telephones, and PCM coders, and now is investigating color television. He holds a patent on a universal printed circuit card.

J. C. CANDY, C.Sc., 1951, Ph.D., 1954, University of Wales Bangor; Bell Telephone Laboratories, 1960—. Mr. Candy has worked on digital circuits and pulse transmission schemes. He is concerned with video signal processing methods.

O. E. DELANGE, B.S. in E.E., 1930, University of Utah; M.A. (Physics), 1937, Columbia University; Bell Telephone Laboratories, 1930—. He was involved in studies of frequency modulation and ultrahigh frequency research up to the start of World War II. The war years were spent on development and design of naval radar. The following period was devoted to studies of broadband pulse systems with emphasis on pulse code modulation. He was responsible for the satellite tracking radar used at the Holmdel, N. J. Bell Laboratories for the Echo I experiment. Recent years have been devoted to studies of light propagation and light transmission systems. Fellow, IEEE.

A. F. DIETRICH, Monmouth College, 1942-44; Bell Telephone Laboratories, 1942—. Mr. Dietrich has been engaged in experimental studies on broadband, baseband, and microwave systems. This has included experimental research work on both FM and PCM ultrashort pulse terminals and repeaters for microwave radio and waveguide applications. He is engaged in experimental studies of light transmission systems.

PETER KAISER, Diplom Ingenieur, 1963, Technische Hochschule, Munich, Germany; M.S., 1965, and Ph.D., 1966, University of California, Berkeley; Bell Telephone Laboratories, 1966—. At Berkeley, Mr. Kaiser was working on frequency independent antennas. He now is engaged in optical transmission research with emphasis on gas lens beam waveguides. Member, IEEE.

FRED W. KAMMERER, Bell Telephone Laboratories, 1925—. His early work included the development of electron beam devices for television systems and multiplexed telephony. During World War II he worked on advanced radar systems. Since 1944 Mr. Kammerer has worked on pulse-code modulation systems, electron beam coders, several special-purpose cathode ray tubes for oscillographs and radar systems. His work on the *Telstar*[®] project was concerned chiefly with equipment design and with the video system at the Andover, Maine, ground station. He is now working on color studies for the *Picturephone*[®] visual telephone. Member, IEEE.

MAURICE KARNAUGH, B.S., 1948, The City College of New York; M.S., 1950, and Ph.D., 1952, Yale University; Bell Telephone Laboratories, 1952–1966. Mr. Karnough has engaged in research on the synthesis of digital data systems, logic circuits, telephone switching networks, and pulse code modulation techniques. As Head, Systems Research Department, he was concerned with special problems in analog-to-digital conversion, system simulation, and nonlinear signal processing. Member, IEEE, Sigma Xi, Phi Beta Kappa.

MARTIN P. LEPSALTER, B.M.E., 1951, College of the City of New York; 1951–53, Ordnance Engineer, U. S. Navy Bureau of Ordnance; 1953–55, Principal Engineer, Hudson Fixtures, Inc.; 1955–57, Lecturer, M.E., College of the City of New York; Bell Telephone Laboratories, 1957—. Mr. Lepselter has worked on semiconductor device development since joining Bell Telephone Laboratories. He holds patents on beam lead integrated circuit structure, metallurgy, and process, as well as the shallow-junction contacts for the *Telstar*[®] solar cells currently in use on many satellites. He is supervisor of the New Device Technology Group. Member, IEEE.

DONALD E. ROWLINSON, Bell Telephone Laboratories, 1951–57; 1960—. He took part in studies of transmission systems for color television and systems for facsimile and medical electronics. Mr. Rowlinson took part in the S.O.S. project described in this issue. He is now with the Optoelectronics Research group, engaged in research on color systems for the *Picturephone*[®] visual telephone.

J. V. SCATTAGLIA, RCA Institute's advanced technology course, 1949; Bell Telephone Laboratories, 1952—. Mr. Scattaglia has done research

on pulse code modulation systems and associated synchronization, and is now engaged in research on color pictures for the *Picturephone*[®] visual telephone.

SIMON M. SZE, B.S., 1957, National Taiwan University, Taiwan, China; M.S., 1960, University of Washington; Ph.D., 1963, Stanford University; Bell Telephone Laboratories, 1963—. Mr. Sze has been concerned with the study of semiconductor device physics. At present he is engaged in studies of metal-insulator-semiconductor devices and metal-semiconductor devices. Member, Sigma Xi, IEEE.

B. S. T. J. BRIEFS

AC Focusing of an Electron Microscope Objective Lens

By R. D. HEIDENREICH

(Manuscript received November 29, 1967)

Control of phase contrast in electron microscope images is presently achieved by fine vernier control of the dc objective lens current. An incremental change δJ in lens current produces a corresponding increment δf in focal length and a focal separation in object space of δL_o . L_o is the distance from lens to conjugate object plane so that ΔL_o is the axial separation of conjugate object plane and scattering specimen. The defocus phase at the image plane of a ray scattered at angle β relative to the axial ray is

$$\chi = \frac{|K|}{2} \Delta L_o \beta^2 \quad (1)$$

with $|K| = 2\pi/\lambda$. The phase contrast G in the image goes as

$$G \sim \sin \chi = \sin \frac{|K|}{2} \Delta L_o \beta^2 \quad (2)$$

ignoring spherical aberration, astigmatism, etc., for this purpose. Precise control of phase contrast for a spacing " a " = $\lambda\beta^{-1}$ requires that the focal separation increments δL_o be as small as 100 Å and preferably less.^{1, 2}

For unsaturated magnetic lenses the change in focal length with lens current is taken as³

$$\delta f = -2C_{ca} \frac{\delta J}{J} \quad (3)$$

with C_{ca} the chromatic coefficient. The focal increment $\delta L_o \approx \delta f$ in this case. For saturated lenses δL_o and δf may not be equal so that it is important to calibrate δL_o in terms of the objective vernier control using a Bragg beam from a crystal.

As a result of re-examining the effect of drift and ripple in the high tension and in the objective lens current, the possibility of achieving a selective or "tuned" focus for certain desired object spacings suggested

itself. If the lens current has a harmonic ripple $\sin \omega t$, then the focal separation ΔL_0 is likewise a periodic function of time and so is the phase contrast (2). The response of the regulated dc supply to an ac voltage applied across the lens through a condenser apparently results in a time-dependent focal separation that behaves like

$$\Delta L_0(t) = -\varphi_0(1 + \sin \omega t) \quad (4)$$

with φ_0 the amplitude in angstroms proportional to the applied ac voltage. The time average phase contrast obtained by using (4) in (2) is found to be

$$\langle G \rangle_t \sim J_0\left(\frac{K}{2} \varphi_0 \beta^2\right) \sin \frac{K}{2} (\Delta L_0 - \varphi_0) \beta^2. \quad (5)$$

If φ_0 is zero, the Bessel function is unity and there is just the usual phase contrast. J_0 has appreciable values at its subsidiary maxima or side bands being -0.4 , $+0.3$, -0.23 , etc., occurring at $(K/2) \varphi_0 \beta^2 = 3.8$, 6.9 , 10.1 , etc., respectively. Since it is very difficult to reduce the ripple in a power supply below about 1 mv it is seen that for certain values of residual ripple it is feasible that the Bessel function in (5) could be zero for certain object distances. The injection of an amplitude φ_0 of the right amount could give "side band contrast" in J_0 for a particular distance " a " in the object.

Fig. 1 is an example of the effectiveness of this type of focusing. These photographs are from a focal series taken of a graphitized carbon black particle and exhibit the 3.4 \AA (002) spacing of graphite. Growth defects in the form of dislocations and curvature of the planes are evident. The ac injection increments in Fig. 1 are 1.5 mv, 2.5 mv and 5 mv, with the corresponding amplitudes φ_0 roughly 300, 500, and 1000 \AA , respectively. Both the Bessel function and the dc effect of φ_0 are operative in the contrast. The micrographs were taken with a Siemens Eleniskop I modified for short focal length (2.4 mv at 100 KV).

Since the ripple in the high tension is an uncertainty the absolute values of φ_0 to achieve maximum side band contrast are uncertain. A continuously variable ac potential is thus desirable with the lens voltage being monitored by an oscilloscope.

The possible advantage of "side band" or "microphase" focusing lies in obtaining useful contrast at spacings $a \leq 2 \text{ \AA}$. To fall well within the principle maximum of J_0 , the residual ripple for a 1.5 \AA spacing must be below 1 part in 10^6 . The injection of $\varphi_0 \approx 75 \text{ \AA}$, 125 \AA , or

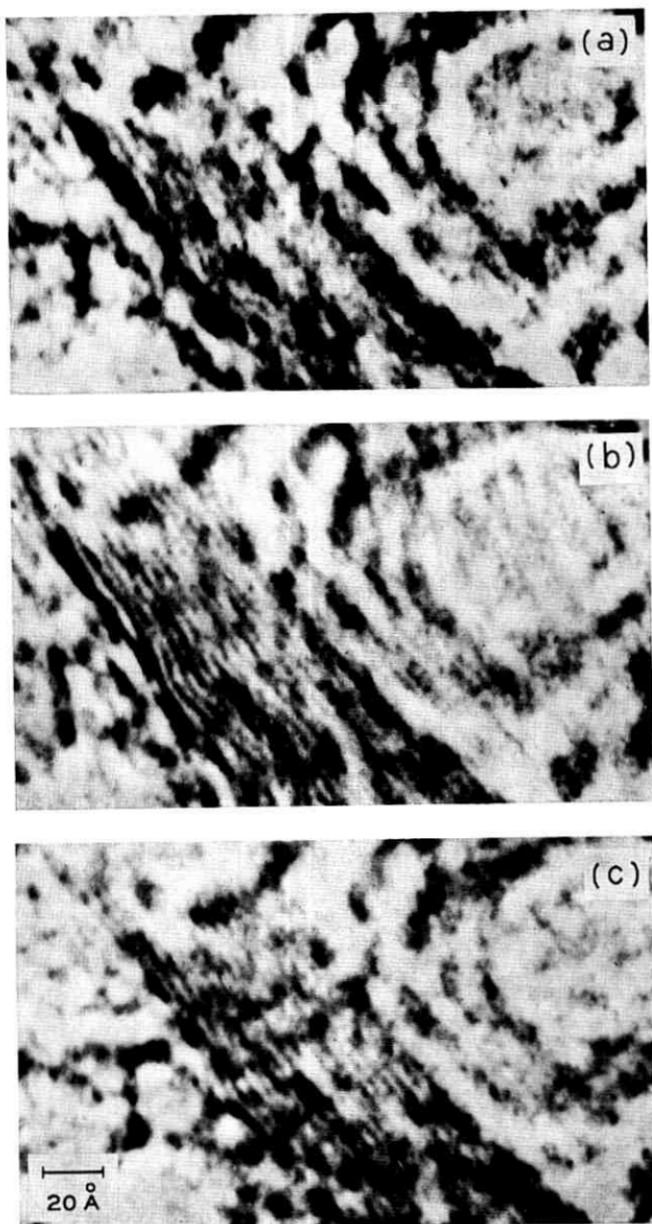


Fig. 1—Electron micrographs showing the 3.4 Å (002) spacing in graphitized carbon black. By injecting 1.5, 2.5, and 5 mv of 60 cycle ac into the lens coil, the focal series shown by a, b, and c, respectively, was obtained. The center picture is very near optimum and shows growth defects in the lattice, whereas a is underfocused and c slightly overfocused for the 3.4 Å spacing.

225 Å ac should "tune" the respective side bands for the 1.5 Å spacing. The (004) period of 1.7 Å in graphite has been imaged in this way using carefully aligned, *axial* illuminations.

REFERENCES

1. Heidenreich, R. D., "Electron Phase Contrast Images of Molecular Detail," *Jour. Electron Microscopy*, 16, No. 1 (1967), pp. 23-38.
2. Thon, F., "Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung," *Z. Naturforschung*, 21 (1966), p. 476.
3. Haine, M., *The Electron Microscope*, New York: Interscience Publishers, (1961), p. 13.

Air-Insulated Beam-Lead Crossovers for Integrated Circuits

By M. P. LEPSALTER

(Manuscript received December 4, 1967)

Air-insulated crossovers for integrated circuit interconnections have been developed which feature low capacitance and dual dielectrics for high yield and reliability. This type of crossover is fabricated using beam lead technology and is applicable to integrated circuit chips as well as substrates.

The cross section drawing in Fig. 1 shows a crossover with dual dielectric insulation consisting of a 5 micron air gap in series with a 2000 Å layer of ZrO_2 . Fig. 2 is a plan view photograph of a typical pattern. The dual dielectric crossover will not fail under mechanical load even if the top conductor is deflected until it touches the bottom level, since the ZrO_2 has a breakdown strength by itself of approximately 100 volts. (See Fig. 3.) Any pinholes in the solid dielectric will not lead to short circuits either, as they would if the top level metallizing were applied by deposition. When the external force is released, the top beam lead will elastically return to its original position, and the air gap will once again sustain over 200 volts.

The fabrication procedure is as follows. A Ti-Pt lower level contact pattern is formed according to methods previously described.¹ A layer of Zr-Cu is then deposited, and feed-through holes etched where contact to the lower level is desired. The top layer, a gold beam lead, is electroformed using the copper as a base, and the copper is etched away. Thermal oxidation of the Zr layer to ZrO_2 completes the structure: two levels of metallization separated by a ZrO_2 -air dielectric.

The use of a thick air gap as one of two dielectrics between two

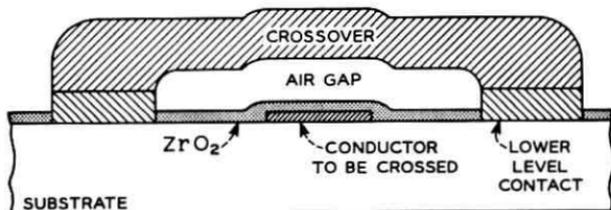


Fig. 1 — Cross section of crossover.

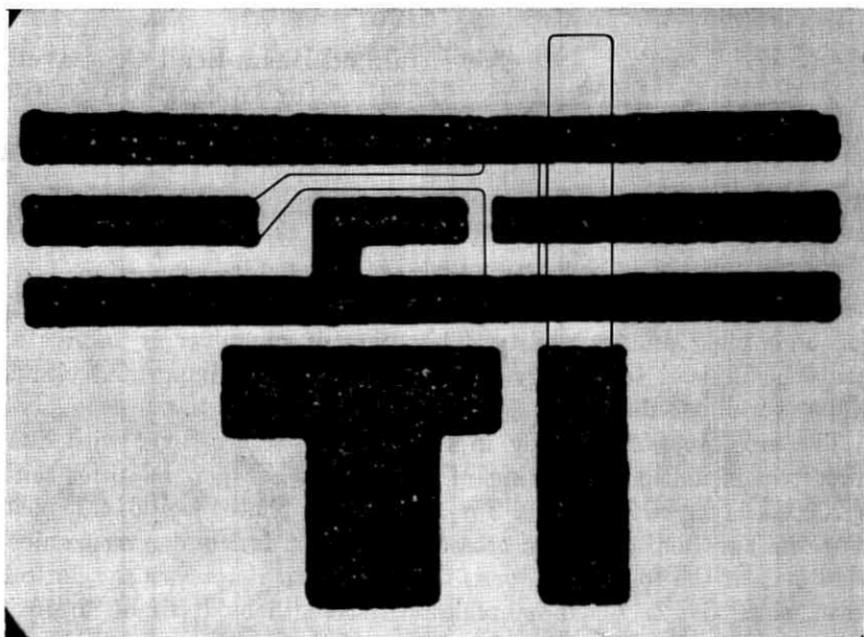


Fig. 2—Plan view of crossover pattern on integrated circuit chip (narrowest line is 0.4 mil).

crossing conductors leads to a system where pinhole shorts are nearly impossible, thermal stresses are minimal, and parasitic capacitance is low (~ 0.001 pf/mil²). Conventional crossovers consisting of evaporated metals over and under a deposited dielectric (for example, SiO₂) are prone to pinhole shorts, conductor opens, and high thermal stresses owing to the mismatches involved. Diffused crossunders are limited to use in chips only, and use large areas of silicon because of their high sheet resistances.

The beam lead crossover system has been successfully applied to integrated circuit chips as well as complex substrates (see Fig. 4).

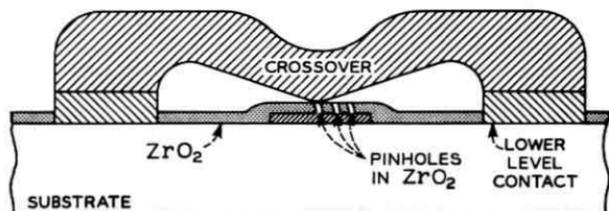


Fig. 3—Cross section of loaded crossover.

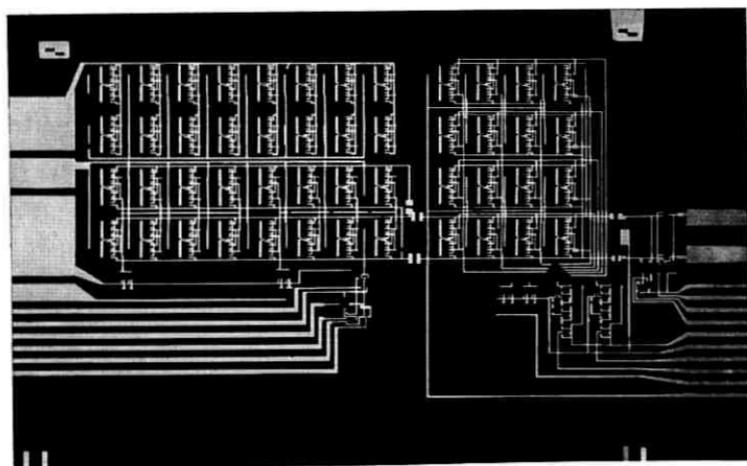


Fig. 4 — Photograph of integrated circuit memory substrate.

In addition to the high initial yields, aging of hundreds of crossovers in room air at 350°C for over 1000 hours has shown no apparent change, nor is any envisaged. The metallurgical system has been stringently tested on beam lead transistors and found to be very rugged. It is therefore believed that these crossovers will be generally useful in many integrated circuit applications.

REFERENCE

1. M. P. Lepselter, "Beam-Lead Technology," *B.S.T.J.*, 45, No. 2 (February 1966), pp. 233-253.

