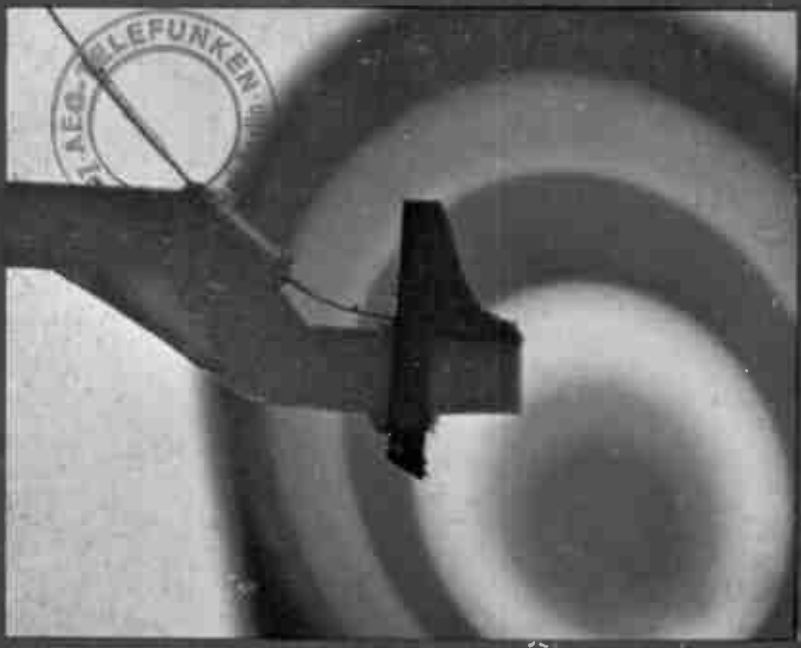
RCA Review



RCA VideoDisc Stylus

September 1981 Volume 42 No. 3

RCA Review, published quarterly in March, June, September and December by RCA Research and Engineering, RCA Corporation, Princeton, New Jersey 08540. Entered as second class matter July 3, 1950 under the Act of March 3, 1879. Second-class postage paid at Princeton, New Jersey, and at additional mailing offices. Effective January 1, 1978, subscription rates as follows: United States and Canada: one year $8.00, two years $14.00, three years $18.00; in other countries, one year $8.60, two years $15.20, three years $19.80. Single copies (except for special issues) up to five years old $3.00.

RCA Review

RCA Review (ISSN 0033-6831) is a technical journal published quarterly by RCA Research and Engineering in cooperation with the subsidiaries and divisions of RCA.

Contents

September 1981 Volume 42 Number 3 ISSN 0033-6831

The RCA CED VideoDisc System—An Overview

J. J. Brandinger

RCA "SelectaVision" VideoDisc Operations, Indianapolis, IN 46224

Abstract—This paper describes the various parts of the RCA CED VideoDisc system. First, the disc is described, including its manufacture and processing. This is followed by a description of the stylus and the player and its operation. Included is a brief discussion of the stereo CED system as well as a comparison of parameters for the 525 line/60 Hz and 625/50 Hz systems.

1. Introduction

Prerecorded video discs give to viewers the ability to see and hear what they want, when they want it. With it, we decide our own programming material—adventure, science fiction, comedy, mystery, drama, sports, information, and inspiration. The RCA CED (capacitance electronic disc) VideoDisc system was introduced to the American market on March 22, 1981, through more than 5,000 dealers. The players carried a suggested list price of $499.95, and discs, from a selection of 100 titles, were available at an average price of about $20. RCA's disc manufacturing plant has been expanded to support production of 3 million discs in 1981, with an ever increasing variety of program material, and will continue its expansion to produce 10 million discs in 1982.

Briefly, the RCA SelectaVision VideoDisc system comprises a disc on which television signals have been prerecorded and a player that converts these signals into a form suitable for input to television sets. Installation is simple and rapid with only three connections required, one to connect the player to the power line, the second to connect the player to the antenna terminals of the TV set; and the third to reconnect

the antenna to the VideoDisc player. When the player is off, antenna signals are fed through it to the television set. Turning the player on replaces the antenna signal with signals from the disc on either channel 3 or 4, at the discretion of the owner, to avoid possible interference from other television signals in the area.

The disc is housed in a protective sleeve, or caddy. Inserting the caddy into the player and then extracting it automatically places the disc on the player turntable, ready for play. After play, the empty caddy is reinserted into the player, retrieving the disc for storage. Each disc will play up to 1 hour on each side, depending on the length of the recorded program.

This paper briefly describes the RCA CED VideoDisc system, giving an overview of its various parts and indicating advances in the system planned for the future.

2. The Disc

Information is embossed on the disc as frequency-modulated vertical undulations in a 140° "V" shaped spiral groove. A small section of the disc is shown in Fig. 1. The signal pattern is initially recorded on a copper surface that is replicated by several successive nickel electro-forming

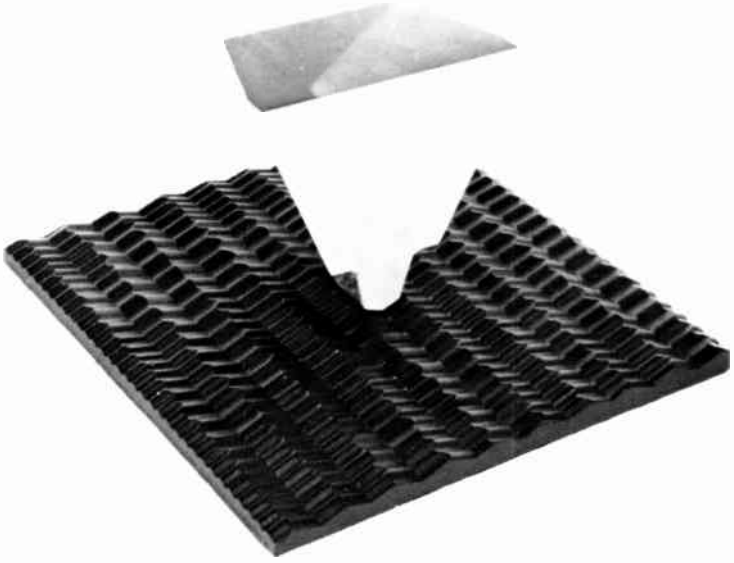


Fig. 1—Disc-stylus Model. A diamond stylus rides in a "V" shaped groove on the disc.

operations to produce stampers. These stampers are mounted in a multi-ton compression molding press on which discs of conductive carbon loaded polyvinyl chloride are molded.

The disc plays back at 450 revolutions per minute, and the signals from the disc vary from 4.3 MHz at sync tip to 6.3 MHz at peak white. The discs are 30.2 centimeters in diameter, and 1.9 millimeters in thickness, about the same dimensions as an audio LP record.

2.1 Disc Mastering and Replication

The various steps in making a master recording and producing a finished disc are shown in Fig. 2.

The major parts of the electromechanical recorder used in mastering are shown in Fig. 3. A smoothly turning, seismically isolated, precision turntable is accurately locked to the signal source. A piezoelectric cutterhead (Fig. 4) is mounted on a sturdy arm with a translation mechanism that radially moves the cutterhead a distance of one centimeter every 3793 rotations of the turntable. On the turntable is a flat aluminum substrate on which has been deposited a grainless electrolytic copper. Before recording, the copper surface is carefully machined flat on a jig borer.

The diamond stylus used to record on the master has a V shape that corresponds to the desired cross section of the finished groove. The stylus

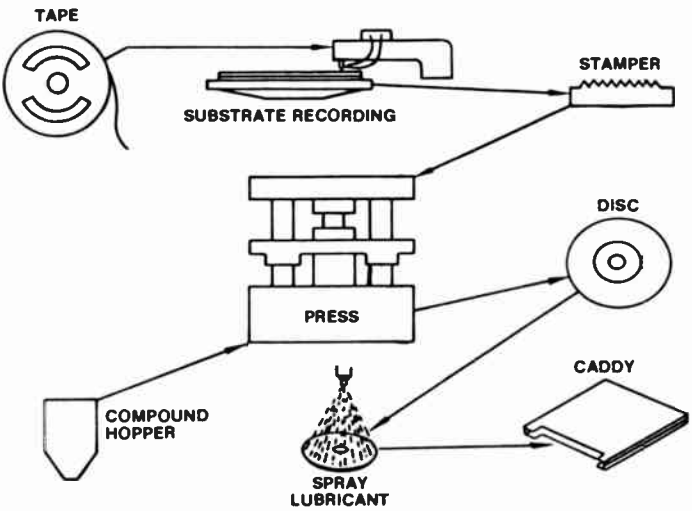


Fig. 2—Disc manufacturing overview.

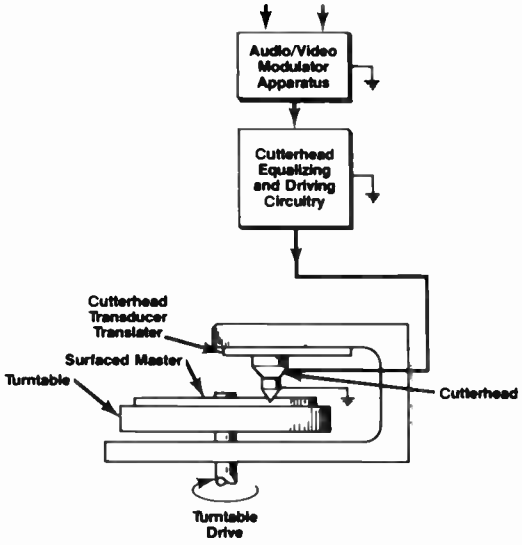


Fig. 3—Simplified schematic diagram of an electromechanical recorder.

cuts the groove (about 4000 Å deep) and the signal modulation (about 850 Å deep) at the same time. The cut is deeper than the groove, so that the groove depth is controlled by the shape of the recording stylus tip and the amount of translation between turns. The amplitude of the signal recorded is determined by the vertical motion of the stylus tip. Recording is done at half-real-time rate to achieve more consistent performance and longer life of the cutterhead. The dynamic stress on the bond to the

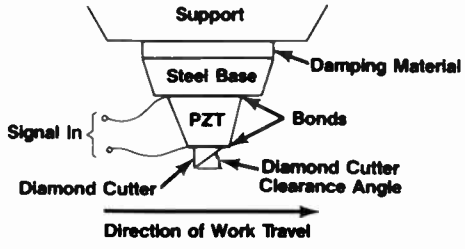


Fig. 4—Schematic of an electromechanical cutterhead.

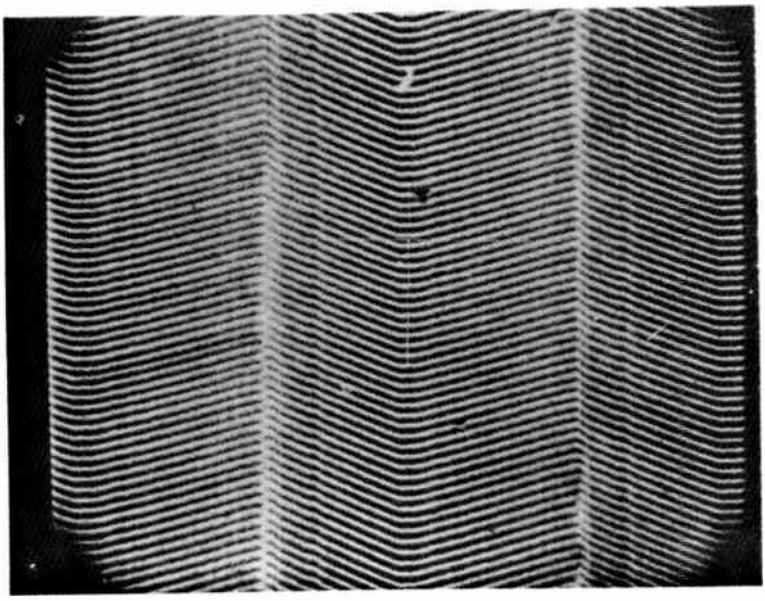


Fig. 5—Surface quality of a signal mechanically cut in copper.

piezoelectric element is in excess of 70 kilograms per square cm for 850 Å peak-to-peak motion at 2.5 MHz. Fig. 5 shows the resultant surface quality of a signal mechanically cut in copper.

Fig. 6 shows the mastering process flow including the cutterhead assembly and substrate preparation. One of the advantages of cutting in copper is that optical laser read-out can be made as soon as the copper is cut; no development time is required as in laser-photoresist mastering.

Recording tests have shown that copper can be cut more smoothly and with less loss of high frequencies than lacquer. Further, evaluation of electron beam-resist and laser-resist mastering generally have shown these systems to produce orders of magnitude higher defect levels than those found in copper mastering. For a groove spacing of about 5 μm, defect levels in a copper master were measured to be about 0.01/cm^2, while with baked and developed photo-resist they were 1/cm^2. This difference was confirmed by evaluating defect levels achieved in integrated circuit photolithography. Yields in IC photo-resist processing decrease exponentially with line spacing due to the increase in defects. This is similar to the results obtained in caddy dust studies (see Fig. 17), which show that the density of particles increases exponentially with

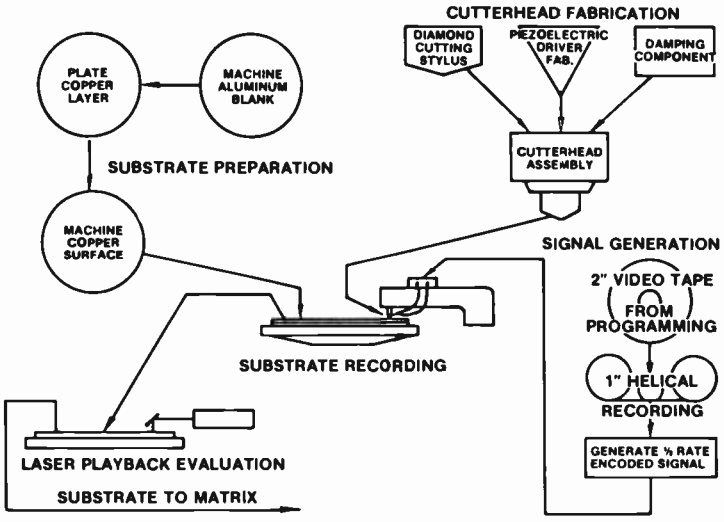


Fig. 6—Mastering process flow diagram.

decreasing particle size. Thus, the expected number of video disc recording defects follow a similar exponential increase with decreasing groove spacing. Actual copper recordings achieve yields well in excess of 50% for all failure causes.

Signal generation starts with input video tape that has been carefully prepared to match the recording standards of the CED system. Imperfections in editing, gamma correction, dynamic range, timing and many other parameters are not acceptable for mastering. Source signals are NTSC compatible, complying with EIA standard RS-170A, and chrominance is limited to peak levels represented by 75% of color bars. (CED system parameters and standards are given in Appendix 1.) If the recorded material is supplied in 2-inch quadruplex formats, it is fed into a 1-inch helical scan video tape machine to achieve a two-to-one slow-down. The head wheel of the 1-inch machine plays back at full speed while the capstan advances the tape at half speed. This means the output of the tape machine plays each TV field twice. One-inch tapes in C-format are acceptable to the system and are played directly on the one-inch helical scan recorder.

The composite video information is passed from the tape recorder through time-base correction, pre-emphasis, and encoding into buried subcarrier format. This analog signal is then converted to digital form (i.e., 14.3-MHz sampling rate, 8-bit amplitude quantitization), stored

in two field storage units, and read out at half rate. The digital signal is converted to analog form and the field code added in the vertical blanking interval on lines 17 and 280. The slowed down video is then fed to a video FM modulator. The half-rate audio from the tape recorder is FM modulated and added to the video FM to produce the composite cutterhead drive signal.

The luminance signal is pre-emphasized (see Appendix 1) to improve the luminance signal-to-noise ratio. This means that high frequency luminance components are recorded at higher than normal amplitude. The pre-emphasis has two parts: (1) *RC* pre-emphasis, which boosts 6-dB per octave between 249 kHz and 995 kHz, and (2) linear phase pre-emphasis, which boosts high frequencies by a dB number equal to the square of the frequency in MHz to a maximum limit of 12 dB.

Replication of the copper substrate is done by conventional nickel electro-forming successive masters, mothers, and stampers as shown in Fig. 7. The original copper substrate has the surface relief pattern that will appear on the VideoDisc. Successive electro-forming steps can provide a fan-out of several hundred metal parts, sufficient to produce disc quantities in the hundred-thousand range from one recording operation.

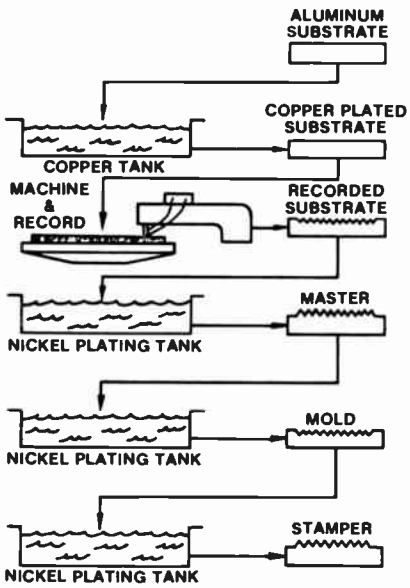


Fig. 7—Matrix process.

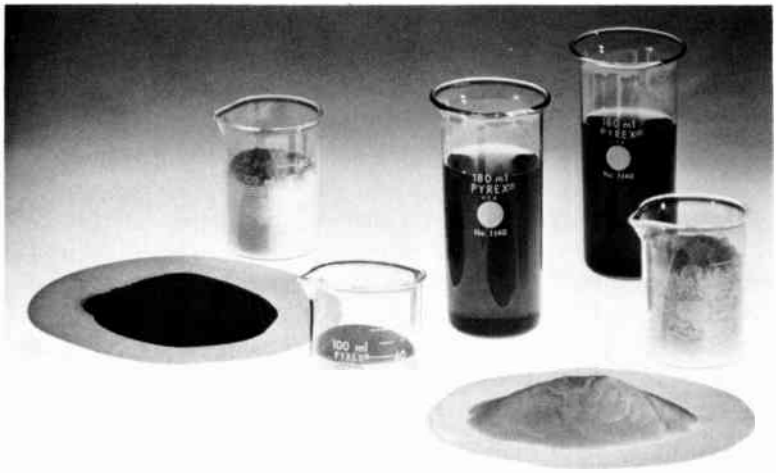


Fig. 8—Materials used in making compound.

2.2 Disc Material

Fig. 8 shows the various ingredients used in making the special compound material for the disc. The base material is a polyvinyl chloride (PVC) homopolymer resin. Conductivity is achieved by the addition of

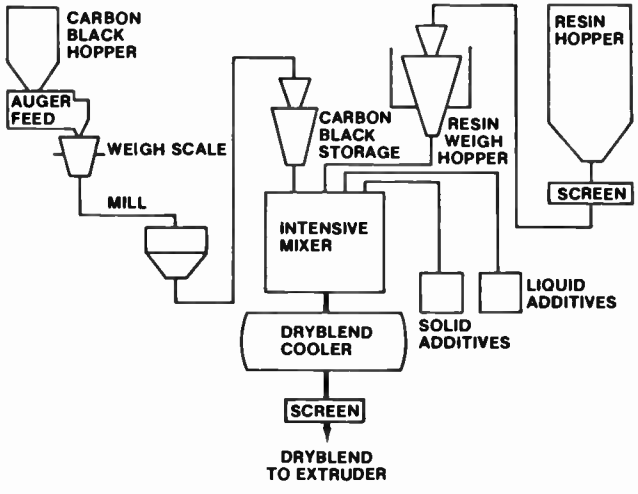


Fig. 9—Material flow sequence.

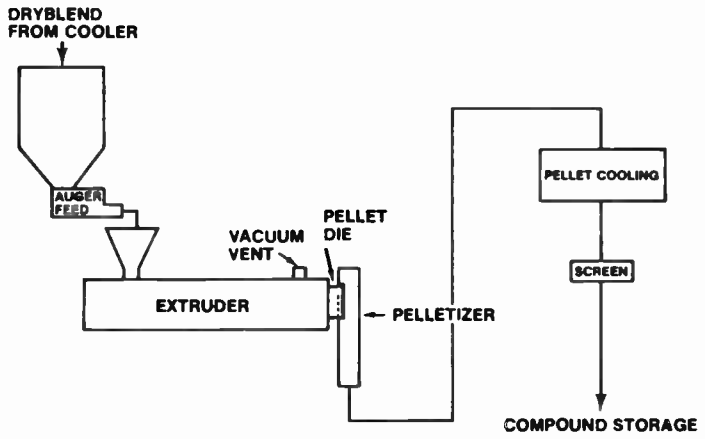


Fig. 10—Dryblend flow sequence.

carbon whose basic particle size is smaller than the 5000-Å minimum size information elements and with conductive properties sufficient to achieve a resistivity of less than 5 ohm centimeters with 15% loading of the base PVC material.

The basic compounding system is shown in Figs. 9 and 10. In Fig. 9, the carbon black is weighed, centrimilled to break up agglomerates and

Fig. 11—Pellets of dryblend.

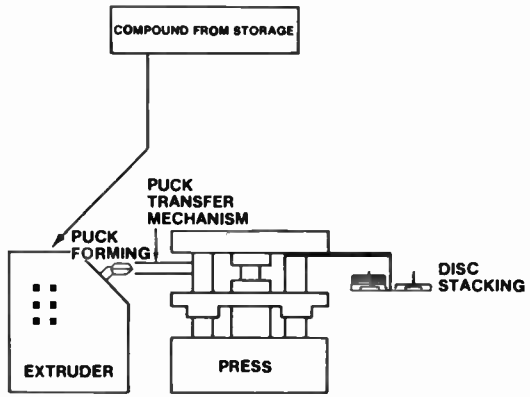


Fig. 12—VideoDisc molding.

mixed with the PVC and processing additives to produce the dryblend. In Fig. 10, the dryblend is fed to a twin screw extruder, pelletized and screened for size. These pellets are shown in Fig. 11.

The basic materials used in the compound have purities similar to those used in pharmaceuticals, i.e., impurities less than 50 parts per billion.

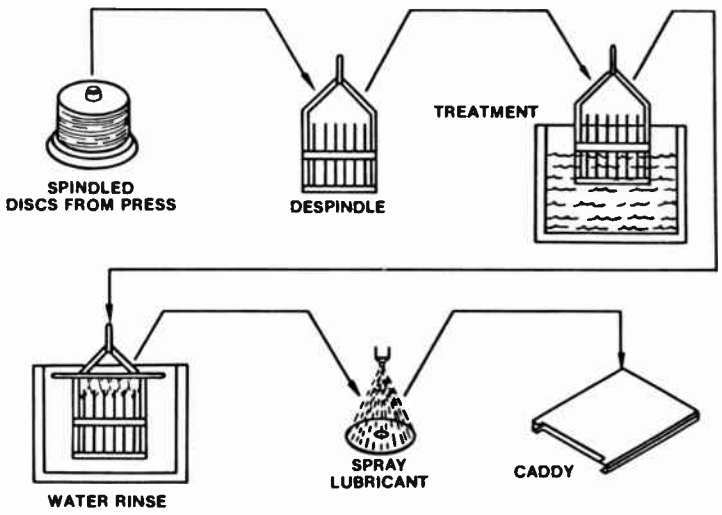


Fig. 13—Lubrication.

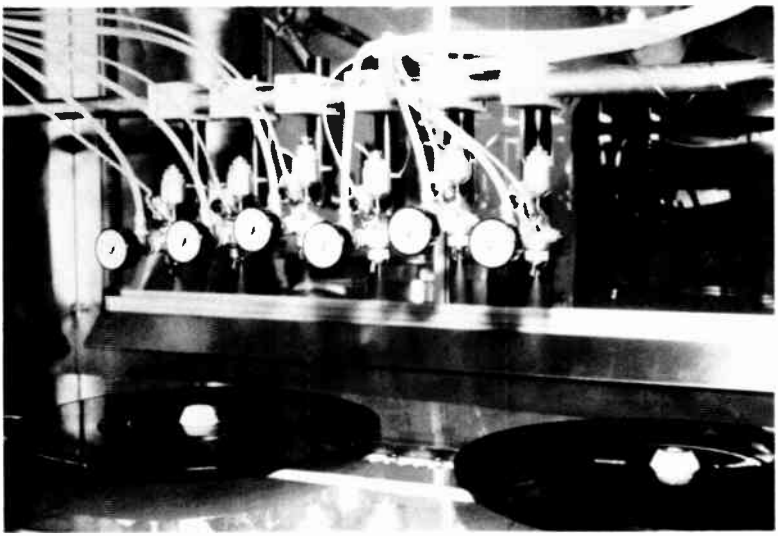


Fig. 14—Special spray configuration.

The pellets are fed to the press extruder (Fig. 12) and formed into a puck. The puck is automatically fed into a 100-ton hydraulic compression molding press, and the disc is formed in less than 40 seconds, trimmed of its flash, and spindled.

The spindled discs from the press are visually inspected for defects, sampled for performance and geometric integrity, and despindled into special carriers for post pressing treatment shown in Fig. 13. The discs are cleaned in special baths, water rinsed, dried, lubricated, and inserted into caddies. The lubricating oil is uniformly applied to the disc in a controlled 300 Å thick layer. Fig. 14 shows the special lubrication spray configuration.

2.3 Caddy Assembly and Packaging

The special caddy container assembly for the disc is shown in Fig. 15. The caddy halves first have the felt-like lip seals applied and are then sonically welded together. Spines, which act as the disc carrier for player insertion and extraction, are inserted into the caddy and the composite unit is provided to the packaging line.

Fig. 16 shows the disc packaging sequence. The assembled caddies have the label applied, discs are assembled into the caddy and the unit is shrinkwrapped. These final units are bulk packed for shipment. In

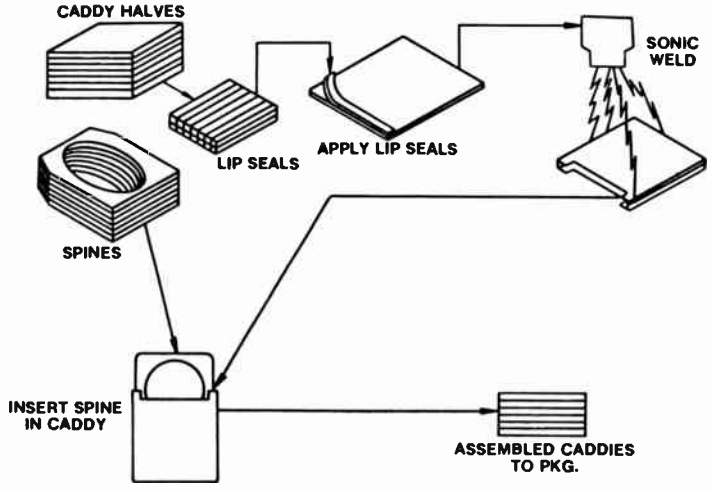


Fig. 15—Caddy assembly.

addition to protecting the disc from dust and from damage due to handling, the caddy package has the added advantages of providing space for a label, convenient storage and shipment. The disc never leaves its caddy, except during play when it is automatically extracted by the

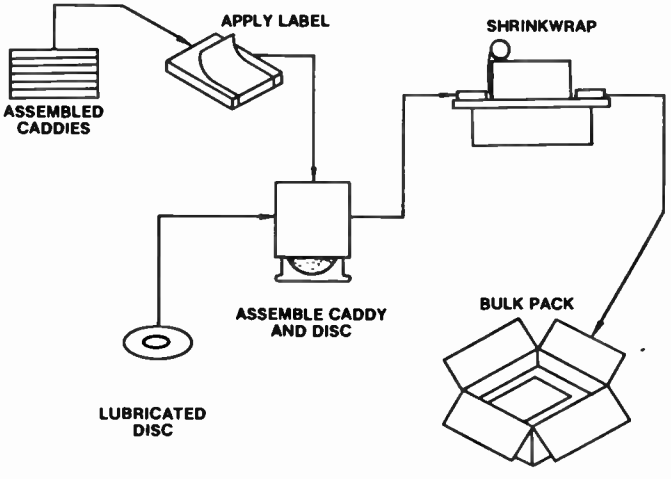


Fig. 16—Disc packing sequence.

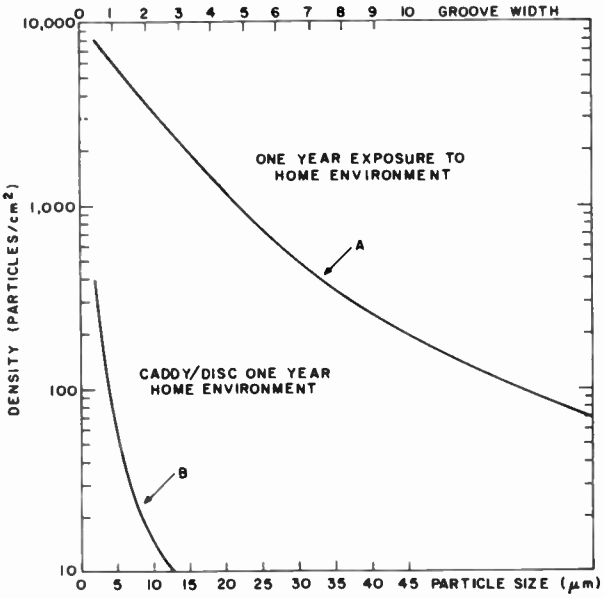


Fig. 17—Analysis of dust accumulated in 1-year simulated exposure to typical home environment. Curve A shows dust on outer cover; curve B shows dust on disc protected by caddy.

player mechanism. After play, the caddy is reinserted into the player and the disc is automatically inserted into the caddy through felt-like lips at the caddy mouth entrance. These lips act as a dust seal and remove gross debris from the disc. Fig. 17 shows results of studies of defects due to dust with and without the caddy.

2.4 Disc Testing

Extensive testing is provided for the product. Fig. 18 illustrates the play tests, physical tests, and environmental stress tests that are routinely performed on the product. In the early stages of production, large samples were taken to build the statistical data base. Today, sampling of one or two per cent per press run is more than adequate to monitor product quality.

The process outlined in the preceding discussion has produced more than 2 million discs in over 170 titles.

3. The Stylus

During playback, a metal electrode attached to a diamond stylus (Fig. 19), reads signals from the disc. The diamond foot is shaped to fit the

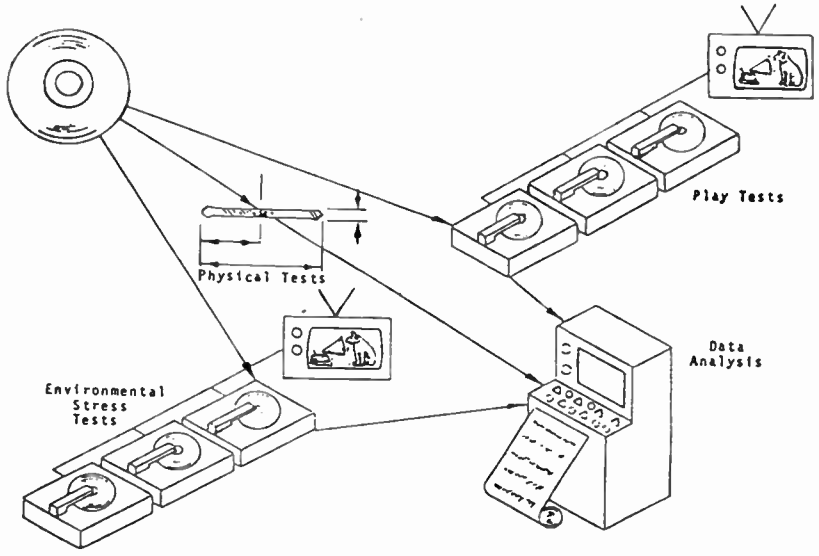


Fig. 18—Disc performance analysis.

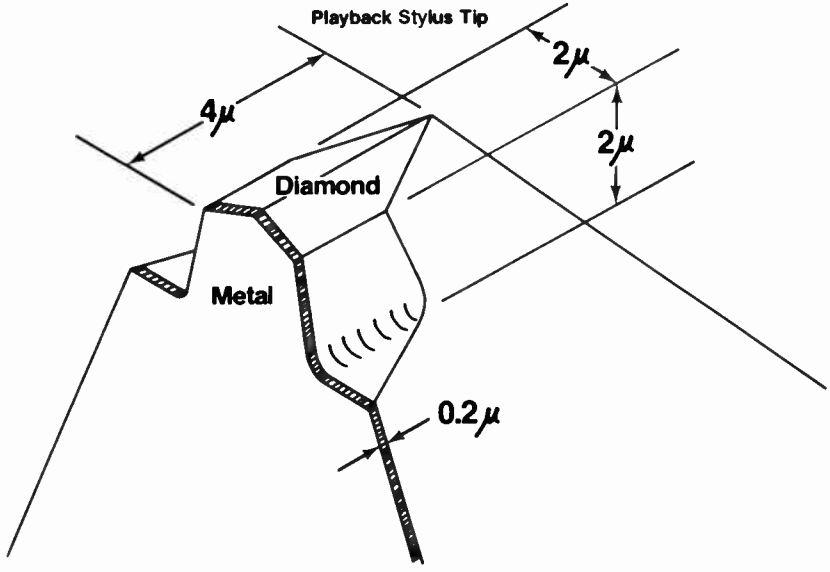


Fig. 19—Stylus tip. This inverted view of the stylus tip shows the relative sizes of the diamond tip and the readout electrode.

groove cross section and is sufficiently long to cover several of the longest recorded wavelengths (1.5 μm maximum). Thus, the diamond rides smoothly on the crests of the recorded waves much as a skier rides over small depressions in the snow. The undulations of the recorded signal rise and fall under the metalized end of the stylus, causing variations in the electrical capacitance between the electrode and the conductive disc surface.

The end of the stylus electrode is about 2 μm wide by 0.2 μm thick. The sides of the stylus tip are tapered as shown, providing a keel shape that maintains the electrode width as it is worn slowly during play. The 2 μm height permits extended useful play life of the stylus tip. In normal use, diamond wear is less than 1.0 μm per 1000 hours. The video FM waves pressed into the disc have a peak-to-peak amplitude of about 850 Å. For this height change, the variation in electrical capacitance experienced by the stylus electrode is proportional to the change in disc-electrode separation and is in the order of 1×10^{-16} farads.

The stylus-disc capacitance is part of a resonant circuit which has a resonance peak at about 910 MHz, shown in Fig. 20. Changes in the disc-stylus capacitance change the tuned circuit resonant frequency. By

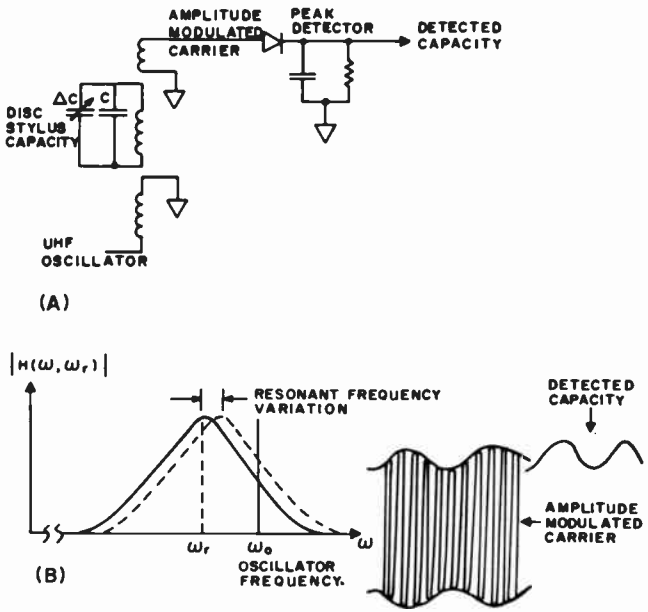


Fig. 20—Capacitance detection (signal pickup): (A) detection circuit and (B) circuit signal response.

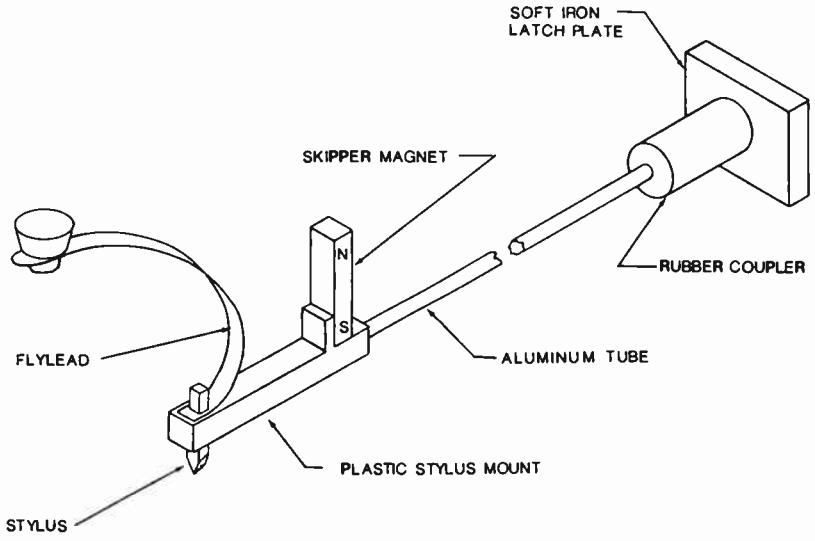


Fig. 21—Stylus arm.

coupling a 915-MHz oscillator signal to the tuned circuit, the 915-MHz oscillator signal is amplitude modulated by changes in the stylus-disc capacitance. The amplitude modulated 915-MHz signal is demodulated by a diode peak detector to extract the frequency modulated signal that corresponds to the signal recorded on the disc.

The diamond stylus is mounted (Fig. 21) on the end of a 8.4-cm long stylus arm made from thin-wall aluminum tubing. A flexible rubber mounting support provides the stylus arm with enough compliance so that the stylus follows irregularities in the disc both vertically and laterally. A small permanent skipper magnet mounted on the stylus arm near the stylus is used to provide limited lateral motions to the stylus when activated by a magnetic field from the stylus kicker coils (discussed later). The rear of the stylus arm is fitted with a soft iron plate that is held by a cup magnet on the arm stretcher transducer. The electrode on the diamond stylus is connected to the UHF circuitry by a flexible flylead, which also serves to hold the stylus against the disc with about 65 milligrams of force. The stylus arm, flylead, and compliant support are mounted in a replaceable cartridge (Fig. 22). The replacement of a stylus cartridge requires no tools or adjustments. The stylus is designed to provide years of service in normal use.

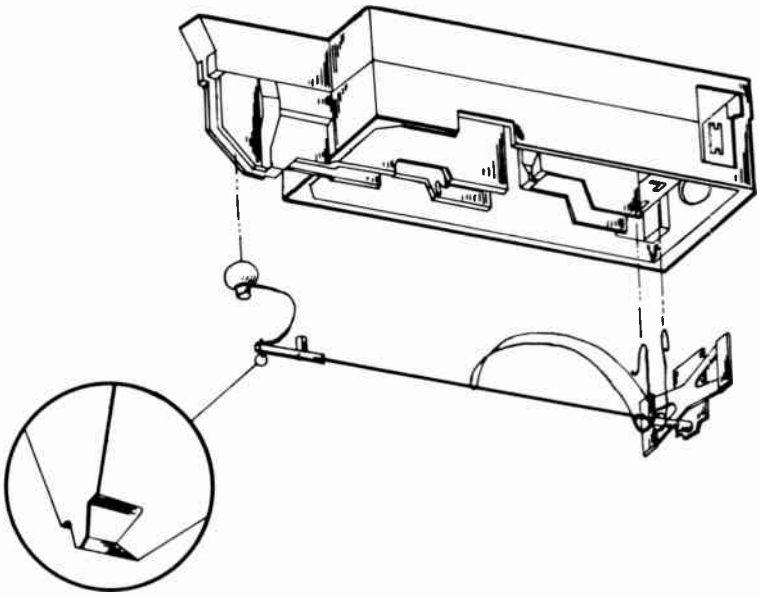


Fig. 22—Stylus cartridge.

4. The Player

4.1 Description

The VideoDisc player consists of a turntable driven at 450 revolutions per minute, synchronized with the 60 Hz power line, a mechanism for inserting and retrieving a disc, a stylus-cartridge pickup arm, signal processing circuitry and system control electronics. A simplified block diagram of the player is shown in Fig. 23.

The output of the 915-MHz oscillator is amplitude modulated by variations in stylus-disc capacitance. This AM modulation is diode peak detected so as to provide a reproduction of the FM signal that was originally impressed on the disc surface. This signal is fed to both audio and video demodulators. Whenever the audio or video FM carrier drops below a given threshold, defect detectors operate appropriate circuitry to prevent noticeable disturbances in the resultant audio or video output. When no audio disturbance is detected, the audio feeds through a track-and-hold circuit to the 4.5 MHz frequency modulator to generate the audio portion of NTSC signal. When a defect is detected, the audio is held at its last valid amplitude until the defect is corrected.

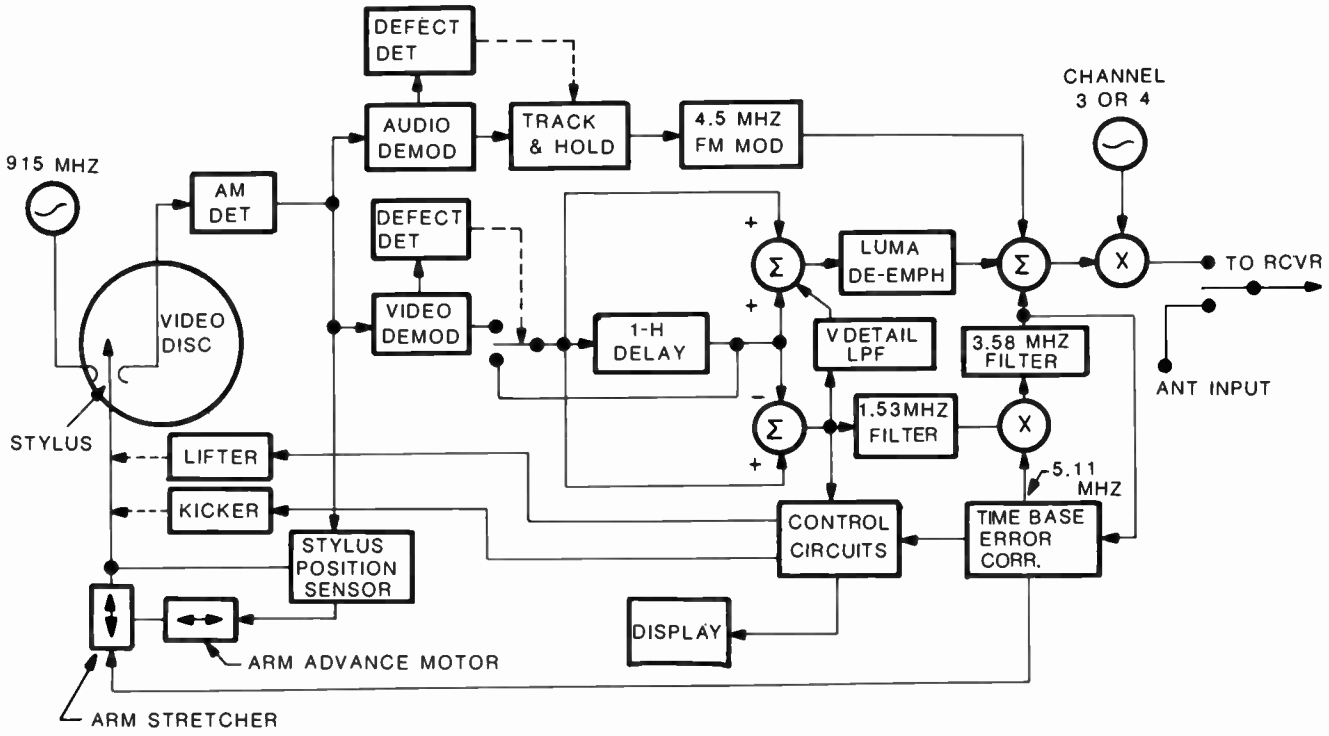


Fig. 23—Player block diagram.

When no defect is detected in the video signal, it is passed directly to a 1-H delay comb filter which separates the luminance and chrominance components as described in Appendix 2. (A detailed discussion of the comb-filter techniques employed in RCA VideoDisc system is given in an accompanying paper in this issue of RCA Review by Pritchard, Clemens, and Ross.) After deemphasis, the luminance signal is combined with other signals to make up the composite NTSC signal.

The output of the delay line is subtracted from its input to extract the color signal. The output of this filter containing the 1.53-MHz subcarrier and color sidebands is mixed with a 5.11-MHz oscillator to convert the color subcarrier to 3.58 MHz, the normal NTSC color subcarrier. The resultant 3.58-MHz color signal is added to the luminance and the 4.5-MHz audio FM to provide the composite NTSC signal which is then translated to either channel 3 or 4. This composite rf signal is available at an output jack at the rear of the player. The output jack of the player is connected directly to the antenna input of a consumer television receiver. When the player is turned off, the antenna input signal is fed directly to the output jack and passed unattenuated to the television receiver.

In traversing the comb filter, low-frequency vertical detail information is removed from the luminance signal. For example, the video signal necessary to produce alternate black and white lines is completely canceled when the input and output of the delay line are added together. To regenerate this low-frequency information, it is necessary for vertical detail to be added back into the picture. This information is extracted from the low-frequency portion of the combed chrominance signal by a low-pass filter and added to the combed luminance signal to provide an uncombed signal below 500 kHz.

When a defect is detected in the video FM carrier, the defect detector automatically switches the input of the delay line to the output of the delay line so that the signal is recirculated. When a defect is detected on any one horizontal line in the picture, the automatic throwing of the delay-line switch to permit recirculation results in a signal from the previous line being substituted during the disturbed video input signal. This type of substitution is similar to that used as "drop-out compensation" in VCR's. Since information on a line-to-line basis is highly correlated, such substitutions are nearly invisible. The result is that defects on the disc, which would otherwise be objectionable, are made inconspicuous.

Tolerance buildup as a result of commercial manufacture of the disc and player and the placement of the disc on the turntable is such that the information on the disc can be slightly off-center during playback. This non-centered condition would result in a speed variation along the

groove that can be translated into horizontal picture jitter at the 7.5-Hz once-around rotation rate of the turntable. Such off-centering causes groove speed to be higher than average on one side of the rotation and lower on the other side.

The arm stretcher shown in the lower left-hand corner of Fig. 23 corrects for the effects of off-centering. This device is similar to a moving-coil loudspeaker mechanism which is fastened by a flexible coupling to the rear end of the stylus arm and moves the stylus arm tangentially along the groove. When the groove speed is low, the stylus arm is moved in a direction opposite to the groove velocity to increase its relative speed. When the groove speed is too high, the stylus arm is moved in the same direction as the groove velocity to reduce the relative speed. The signal for driving the arm stretcher is obtained by comparing the 3.58-MHz color burst from the disc with a fixed oscillator at the same frequency. The net result is that off-centered conditions as large as 254 micrometers do not produce detectable picture jitter in playback.

During the play of a disc, the stylus is maintained in the center of the groove. This is accomplished by sensing the lateral position of the stylus relative to the cartridge housing by capacitive coupling of the stylus flylead to varactor diodes driven out-of-phase by a 260-kHz oscillator. As the capacitance of one diode increases, the other decreases and vice-versa. These diodes are located on each side of the stylus flylead, so that a fraction of their capacitance is added to the stylus capacitance as a function of how close the flylead is to the diodes. When the stylus is centered, the resultant varactor-diode outputs are canceled. In the off-centered condition, the output of one diode is greater than the other causing a change in the stylus resonant circuit and giving rise to 260-kHz components in the output of the 915-MHz amplitude detector. The amplitude and phase of the 260-kHz signal indicate the amount and direction of stylus off-centering. A dc arm-advance motor is driven in response to this 260-kHz error signal to center the stylus arm and return the error signal to zero.

The stylus kicker shown in the lower left of Fig. 23 permits small, rapid lateral movements of the stylus during play. A tiny, permanent magnet mounted on the stylus arm near the stylus is "kicked" sideways by magnetic fields from small coils mounted in the stylus-cartridge arm housing. When movement of the stylus is desired, an appropriate pulse of current through the coils kicks the stylus sideways one or more grooves in either the forward or reverse direction. This operation is activated by pressing a visual search button on the player. A pulse is applied to the kicker coils just prior to each vertical blanking interval such that the stylus moves about two grooves. Since there are 8 fields per rotation, the picture shown on the TV screen advances (or reverses) at 16-times

normal speed. Since the stylus movement takes place during the vertical blanking interval, no picture breakup occurs.

The stylus kicker is also used to correct for locked groove defects on the disc. Recorded on the disc is a unique field number in each vertical blanking interval. The numbers increase monotonically from the start of play to the end of play on the disc. Circuits are built into the player to decode and keep track of the field numbers. During normal play, these numbers progress regularly and after conversion to minutes are displayed on an LED 2-digit minute indicator. When a locked groove occurs, the numbers do not progress normally but instead decrease. The player recognizes this fact and causes the application of pulses to the kicker coils to move the stylus ahead by two grooves. These pulses continue until field numbers read from the disc equal or exceed the numbers predicted in the player microprocessor. In most cases, locked-groove defects are corrected so rapidly as not to be objectionable to a viewer. No corrective action is taken when forward groove skips occur. In general, forward skips cause little disturbance to the viewer.

The stylus lifter is activated to raise and lower the stylus as required for proper player operation. The stylus is lowered during normal play and visual search, and it is lowered momentarily onto a stylus cleaner each time a disc is removed from the player. In all other conditions, including power off, the stylus is lifted off the disc.

4.2 Player Operation

Fig. 24 is a photograph of the SFT100 player introduced in the United States on March 22, 1981. It measures 43.2-cm wide by 39.6-cm deep by 14.7-cm high, weighs 9.1 kg, and consumes 35 watts of power. Additional player parameters are given in Appendix 1. Jacks for antenna input and rf output, a channel selection switch, and power-line cord are on the back of the player. As shown in Fig. 24, a function lever switch on the right side of the front panel has positions for Load/Unload, Play and Off. In the Load/Unload position, a caddy-entry door is opened to permit the loading or retrieving of a disc by means of caddy insertion. In the Play position, the caddy-entry door is closed, the turntable is energized, the stylus is dropped onto the disc, and the disc is played.

Player operation is controlled by five push buttons. The Pause push button causes the stylus to lift from the disc and the arm advance to stop. A second push of this button causes the stylus to drop down onto the disc and the arm advance to engage. Thus, normal play is resumed. The Pause button allows one to interrupt the program for as long as desired without missing any of the program content and without harm to the disc. The two Visual Search buttons cause the stylus to be kicked nominally two

Fig. 24—Video Disc player—Model SFT100.

grooves either forward or reverse during the vertical-blanking interval so that the program action speeds up approximately 16 times normal rate without picture breakup. These Visual Search buttons conveniently permit locating a particular section of the program. The two Rapid Access push buttons cause the stylus to be lifted and move the stylus arm housing at about 150 times normal speed in either the forward or reverse direction. Both video and audio signals are muted during this operation.

The Play Time indicators give an indication of the stylus position measured in minutes of play starting at the beginning of the disc. These indicators also show an "L" for the Load/Unload position of the function lever, "P" for Pause, and an "E" to indicate the program has ended.

After a disc has been inserted into the player, the Side indicator shows which side of the disc is actually being played or is ready to be played.

The stylus cartridge access door on the top of the player permits simple and easy changing of the stylus cartridge when required. The stylus is designed to operate for years in normal home operation. Discs can be played many hundreds of times without deterioration. All of the parts of the system are designed for extended life.

4.3 Special Playing Modes

The design of the CED player-disc system has many options for unique

playing modes. The stylus-disc interface permits radial disc scanning, repeat groove playing (millions of times) and random access by time or field index number, all without damage to the disc.

Typical of the features that have been demonstrated in the laboratory are:

- Fast Motion—Forward/Reverse
- Random Access—Forward/Reverse and Loop Back or Forward
- Slow Motion
- Stop Action

} With One-Frame Storage

The addition of these features to products requires appropriate software, development time, and market acceptability in terms of value versus cost.

5. Stereo/Bilingual

RCA plans to introduce a stereo version of its CED player in mid 1982 along with appropriate stereo discs. The form of the signal on the disc is similar to the monaural version with the addition of a second audio modulated FM carrier at 905 ± 50 kHz deviation, as shown in Fig. 25.

The left and right stereo channels are matrixed to produce L + R for modulation of the 716 kHz channel and L-R for modulation of the 905 kHz channel. This provides compatibility for the monaural players, which use 716 kHz as their normal audio decoding channel.

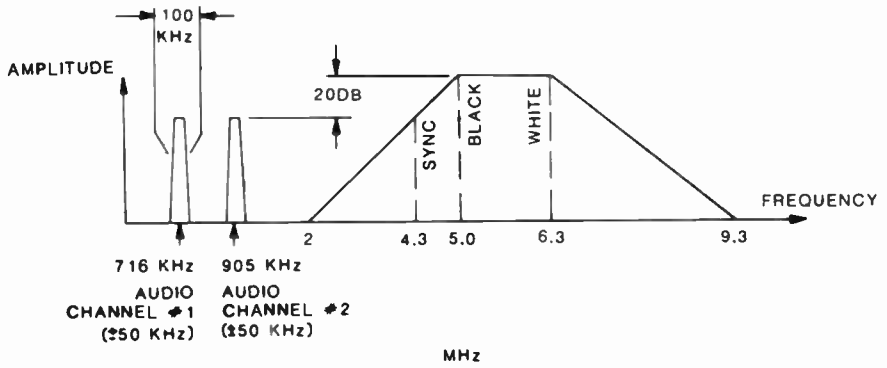


Fig. 25—Frequency spectrum. In stereo or two-channel recording, the audio carriers are 20 dB below the video carrier.

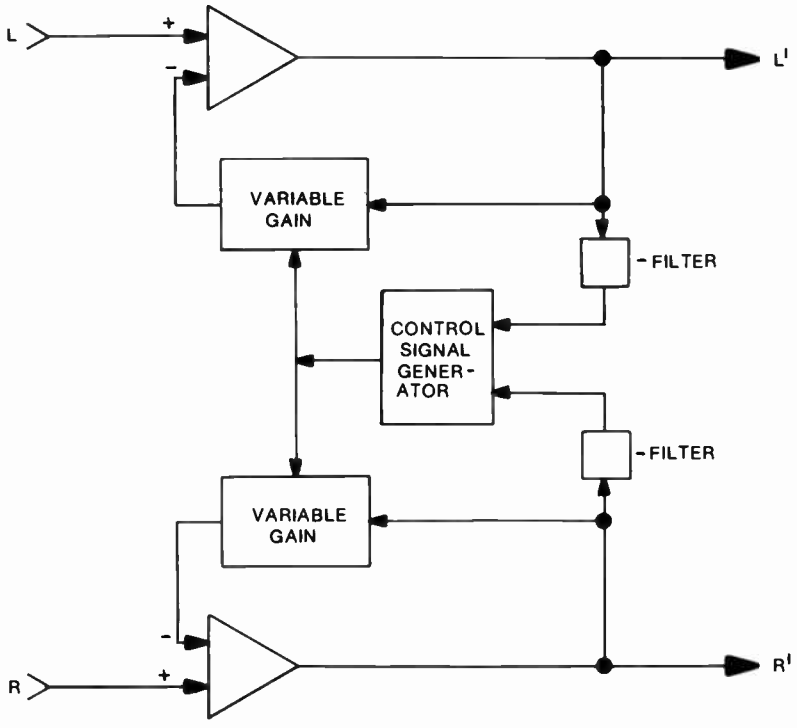


Fig. 26—CX Stereo signal compression block diagram (CX is a trademark of CBS Laboratories).

Each stereo channel prior to matrixing is encoded in the CX* system for noise reduction and extension of dynamic range. The CX compression system has the general form shown in Fig. 26. Voltage controlled amplifiers in the feedback loop of the left and right channels are adjusted by a control signal derived from the left and right channel outputs. Reference operating level is defined as that which produces half the maximum deviation of the FM modulated sum channel with 1 kHz applied to both the left and right channels (0 dB reference).

Steady state compression of a 1 kHz sine wave is 2:1 on a log-log plot at input signal levels greater than −40 dB and 1:1 for signals less than −40 dB. The filter, control-signal generator, and variable gain amplifiers produce a resultant steady state frequency characteristic for a constant input amplitude sine wave as shown in Fig. 27.

The control signal generator, shown in block diagram form in Fig. 28,

* CX is a trademark of CBS Laboratories.

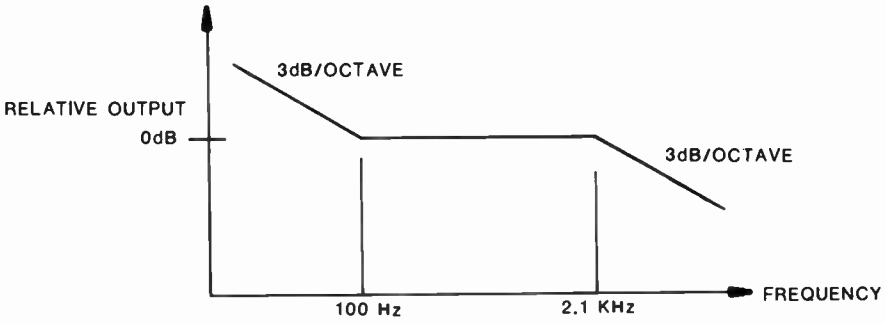


Fig. 27—CX signal compression amplitude versus frequency characteristic.

has a non-linear transient response. The details of the control signal generator are beyond the scope of this paper except to note that low amplitude transients do not modify the steady-state response whereas larger transients result in control signals that reduce the noise pumping associated with noise reduction systems. To take full advantage of the CX system the player must perform an inverse expansion of the audio signals that matches the original encoding. There is up to 20-dB additional noise suppression, resulting in 70-dB dynamic range with the full use of the CX system.

The use of the two independent audio channels for bilingual application requires a separate input to each of the available audio channels. CX encoding/decoding is not part of this type of system. Separation of the two channels exceeds 55 dB.

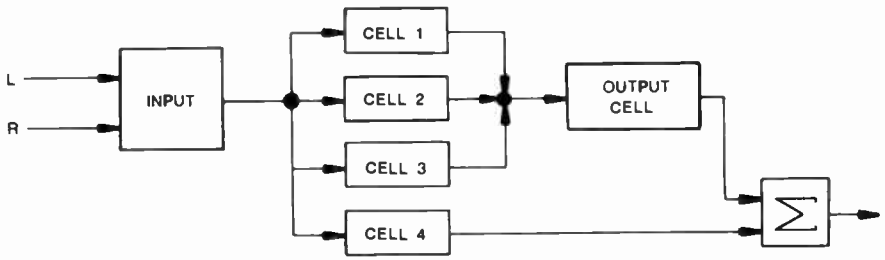


Fig. 28—CX system control signal generator.

6. 625 Line/50 Hz Field System

The developoment of a 625 line/50 Hz system has been based on maintaining similarity in disc manufacture while achieving improved signal performance. A summary of CED parameters for this system is given in Appendix 1 along with details of the signal encoding.

The following items are identical to those used in the 525 line/60 Hz field system:

- Disc Geometry
- Groove Dimensions
- 8 Fields/Revolution
- 1 Hour/Side
- Master Recording Disc
- Recording Cutterhead
- Matrix—Metal Parts
- Pressing
- Surface Treatment
- Caddy
- Stereo Audio Compansion
- Video FM Standards
- Cartridge Pickup System
- Player-Caddy Handling System

System parameters that are slightly different are as follows:

	525 Line/60 Hz	*625 Line/50 Hz*
Audio FM Carriers	715.9 kHz	710.9 kHz
Chrominance Subcarrier	1.534 MHz	1.523 MHz
Turntable Speed	450 RPM	375 RPM
Horizontal Scan Frequency	15.734 kHz	15.625 kHz

Major differences relate to (1) the input tape software standards, which for the 625 line/50 Hz system require PAL standards compatible with CCIR recommendation 472-1, (2) player video signal processing, which requires comb filter delay line decoding in accordance with 625 line/50 Hz standards, and (3) modulation of baseband signals to form PAL and/or SECAM compatible rf/video signals for television receiver input.

Another major difference is in the level of radio frequency emission the players may have in the 902–928 MHz band. Permissible radiation in Region 2* is up to 50,000 μv/meter at 30 meters whereas in Region 1

* ITU Regions as defined in Table of Frequency Allocation.

only 63.1 μv/meter is permitted (VDE 0871, ISM, Class B). This creates special problems for player design.

No other important difference exists in the CED standards for the two systems.

References:

H. N. Crooks, "The RCA SelectaVision VideoDisc System," *RCA Engineer*, **26**, No. 5, 1981, pp. 4–10.

J. K. Clemens, "Capacitive Pickup and the Buried Subcarrier Encoding System for the RCA VideoDisc," *RCA Review*, **39**, March 1978, pp. 45–50.

RCA Review Special Issue, "VideoDisc," Vol. 39, No. 1, March 1978.

D. H. Pritchard, J. K. Clemens and Ross, M. D., "The Principles and Quality of the Buried Subcarrier Encoding and Decoding System," *IEEE Trans. Consumer Electronics*, **CE-27**, No. 3 (ISSN0098-3063), Aug. 1981, pp. 352–360.

T. J. Christopher, F. R. Stave, and W. M. Workman, "The SelectaVision Player," *IEEE Trans. Consumer Electronics*, **CE-27**, No. 3 (ISSN0098-3063), Aug. 1981, pp. 340–351.

W. M. Workman, "The VideoDisc Player," *Electro 81*, Professional Program Session Record, April 7–9, 1981.

Appendix 1—CED System Parameters

A1.1 525/60 Hz and 625/50 Hz Systems

Disc Physical Dimensions

Diameter	30.2 cm
Thickness	1.9 mm
Center-Hole Diameter	3.3 cm
Rotation Rate (525/60 Hz)	450 RPM
(625/50 Hz)	375 RPM
Recorded Band	7.3 cm wide
Starting Diameter Preprogram Band	29.3 cm
Starting Diameter Program	29.1 cm
End Diameter Program (60 Minutes)	
(525/60 Hz)	14.38 cm
(625/50 Hz)	16.81 cm
End Diameter Post Program Band	
(525/60 Hz)	13.56 cm
(625/50 Hz)	16.00 cm
Groove Density	3793 grooves/cm
Groove Depth (p-p)	0.48 μm
Signal Amplitude (p-p)	0.085 μm

Disc Performance Parameters

Play Time (4 Frames per Revolution)	60 minutes per side
Recorded FM Signal	4.3 to 6.3 MHz
Luminance Bandwidth	3.0 MHz

Chrominance Bandwidth	0.5 MHz
Video Signal-to-Noise Ratio	>46 dB (CCIR)
Chrominance Signal-to-Noise Ratio	>40 dB
Audio Carriers (525/60 Hz)	715.9 kHz
(625/50 Hz)	710.9 kHz
Audio Bandwidth	15 kHz
Audio FM Signal Deviation	±50 kHz
Audio Signal-to-Noise Ratio	>50 dBA(USASI)
Minimum Stereo Separation at 1 kHz	26 dB
Dynamic Range Mono (without compansion)	50 dB
Stereo (with compansion)	70 dB

A1.2 525 Line/60 Hz System

Source Signal

Source signals are NTSC compatible television signals, complying with EIA Standard RS-170A. Chrominance is limited to peak levels represented by 75% color bars.

Standard Predistortion

Composite signals to be encoded for the CED System are delay equalized before baseband video processing as follows:
$f < 3.0$ MHz: $t = 0$
$f > 3.0$ MHz: $t = (-170)\ (f - 3.00\ \text{MHz})/(0.58\ \text{MHz})$ nanoseconds.

Baseband Video Processing

The luminance signal is combed such that information contained at frequencies* $(2n + 1)\ f_H/2$ is attenuated by at least 40 dB with respect to frequencies nf_H (where n is any integer), in the band from 1.0 to 2.0 MHz. Vertical detail is added back such that at frequencies 0.9 MHz and below, vertical detail is not attenuated more than 3 dB. The vertical detail is attenuated at least 17 dB at 2.0 MHz.

Pre-emphasis

6 dB/Octave pre-emphasis is applied to the luminance signal between t_1 and t_2. The time constants are:
$t_1 = 640$ nanoseconds
$t_2 = 160$ nanoseconds
Maximum pre-emphasis 12 dB

* $f_H \doteq 15.734264$ kHz.

Additional linear phase pre-emphasis is added to luminance as follows:
$P = Kf^2$, ($f \leq 3.0$ MHz); P is pre-emphasis in dB; $K \approx 1$;
f is frequency in MHz.
Maximum $P = 12$ dB.

Chrominance

The chrominance signal is combed at all frequencies such that information contained at frequencies nf_H is attenuated by at least 40 dB.

Chrominance carrier frequency is $195\, f_H/2$, approximately 1.534 MHz

Frequency Modulation—Video

Carrier Frequency	IRE Units	Frequency, MHz
Sync tip	−40	4.3
Blanking	0	4.871
White Level	+100	6.3
Maximum Frequency	+144	6.929
Minimum Frequency	−66	3.929

Frequency Modulation—Audio

L + R; Monaural; Independent Channel 1 $91 f_H/2$, approximately 715.9 kHz
L − R; Independent Channel 2 $115\, f_H/2$, approximately 904.7 kHz
Peak deviation: ±50 kHz
Pre-emphasis: 6 dB/Octave, time constant = 75 μsec.

A1.3 625 Line/50 Hz System

Source Signal

Source signals are 625 lines/frame, 50 fields/sec, PAL—compatible television signals complying with CCIR Recommendation 472-1 (Study Group II). Chrominance is limited to peak levels represented by 75% color bars.

Baseband Video Processing

The luminance signal is comb-filtered to remove information at frequencies* $\simeq (2n + 1)\, f_H/4$. Attenuation of information at $(1135\, f_H/4) + 25$ Hz should be ≥30 dB. The information removed (which includes some vertical detail) is reinserted through a modified $2.5T$ $\sin^2$ pulse filter (having its first zero at 4.0 MHz and a zero added at 4.43 MHz).

* $f_H \doteq 15.625$ kHz.

The resultant signal is comb-filtered to remove information at frequencies $\simeq (2n + 1) f_H/2$. The information so removed (which includes vertical detail) is reinserted through a band-rejection filter whose characteristics are such that, in the combined signal, information has been attenuated $\geq$20 dB at $(195f_H/2)$ and $\approx$3 dB at $(195f_H/2) \pm 48f_H$.

Pre-emphasis

6-dB/octave high frequency pre-emphasis is applied to the luminance signal between t_1 and t_2. The time constants t_1 = 640 nanoseconds; t_2 = 160 nanoseconds.

Maximum pre-emphasis 12 dB.

Additional linear phase pre-emphasis is added to luminance as follows:

$P = Kf^2$, $(f \leq 3.0$ MHz); P is pre-emphasis in dB; $K \simeq 1$; f is frequency in MHz.

Maximum P = 12 dB.

Chrominance

The chrominance signal is comb-filtered to remove information at frequencies $\simeq nf_H/2$. Attenuation of information at $568 f_H/2$ should be $\geq$30 dB.

The resultant signal should be combed to separate PAL U and $\pm V$ components. The $\pm V$ component polarity is switched so that $+V$ is always recombined with U.

Chrominance carrier frequency is $(195 f_H/2) - 25$ Hz, approximately 1.523 MHz.

Frequency Modulation—Video

Carrier Frequency	IRE Units	Frequency, MHz
Sync tip	−40	4.3
Blanking	0	4.871
White Level	+100	6.3
Maximum Frequency	+144	6.929
Minimum Frequency	−66	3.929

Frequency Modulation—Audio

L + R; Monaural; Independent Channel 1

$91 f_H/2$, approximately 710.9 kHz

L − R; Independent Channel 2

$115 f_H/2$, approximately 898.4 kHz

Peak deviation: ±50 kHz.

Pre-emphasis: 6 dB/octave, time constant = 75 μsec.

A1.4 CED System Parameters

Player (525/60 Hz) – SFT100

Electrical

Readout	Capacitance
RF Modulator Frequency	61.25 MHz (CH.3) 67.25 MHz (CH.4)
RF Output Level	2.0 mV/75OHMS
Luminance Bandwidth	3.0 MHz
Luminance Signal-to-Noise Ratio	>46 dB (CCIR)
Chrominance Bandwidth	0.5 MHz
Audio Bandwidth	15 kHz
Audio Signal-to-Noise Ratio	>50 dBA (USASI)
Power Consumption (115V AC, 60 Hz)	35 Watts

Mechanical

Cabinet: Width	43.2 cm
Height	14.7 cm
Depth	39.6 cm
Weight	9.1 kg
Stylus Set-Down Diameter	29.2 cm
Stylus End of Travel Diameter	14.0 cm
Cartridge Contact—Height from Disc	1.4 cm
Maximum TIR	102 μm

Playing Mode

Normal Forward
Visual Search, 16 ×, Forward/Reverse
Rapid Access, 150 ×, 30 Seconds/Side, Forward/Reverse
Pause

Appendix 2—Buried Subcarrier System—1-H Delay Comb Filter

An NTSC television signal is repetitively sampled at rates of f_H = 15.734264 kHz, 59.94 Hz and 29.97 Hz, which results from the horizontal and vertical scanning rates. A fourier analysis of the signal shows an energy spectrum concentrated at interval spacings of f_H, with subsidebands grouped around each f_H interval at multiples of 59.94 and 29.97 Hz spacing as shown in Fig. 29.

The addition of a color subcarrier at 1.53 MHz has its energy spectrum interleaved with the baseband video by synchronizing the subcarrier with horizontal sync at an odd multiple of one-half line rate (195 $f_H/2$ or 1.5340907 MHz).

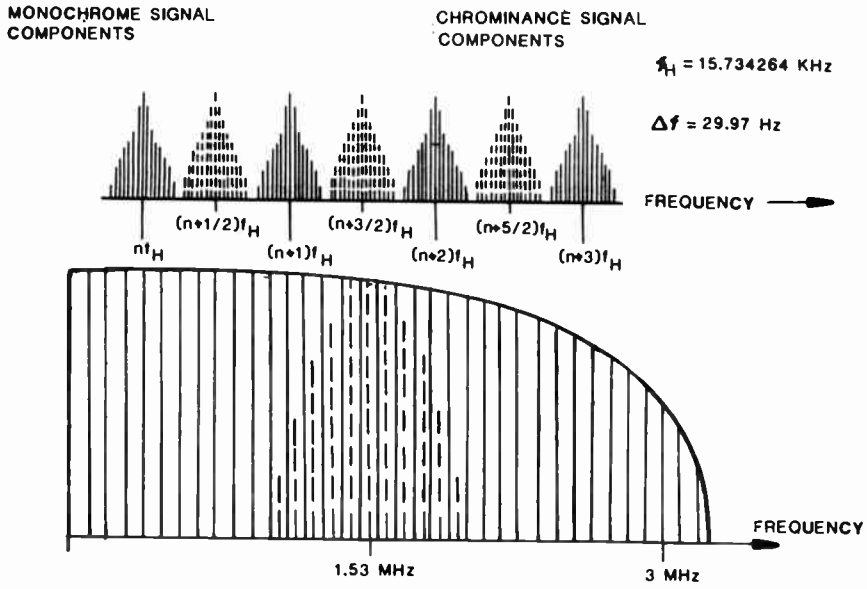


Fig. 29—Video signal spectrum.

The baseband luminance signal is limited in bandwidth to 3 MHz consistent with the resolution reproduced in home television receivers. The color carrier sidebands are similarly limited to 500 kHz, thus restricting all picture information within a 3-MHz band. This efficient coding system permits excellent picture reproduction with minimum spurious beats from the limited recording bandwidth of current video discs.

Separation of the color and luminance information is accomplished using a comb filter in the player. Fig. 30 shows the basic block diagram of a 1-H delay comb filter along with resultant amplitude versus frequency responses of the circuit operations. The video input is delayed by one horizontal scanning line in 585 stage CCD line with a 9.2 MHz clock frequency. The delayed and undelayed signals are added in the summation block such that all frequency components which are spaced in even multiples of $f_H/2$ are passed unattenuated to the output as shown. All frequency components spaced at odd multiples of $f_H/2$ are canceled. The frequency spectrum as shown follows a $|\cos(\pi f/f_H)|$ form. Thus, all luminance video components are passed through this comb filter while chroma components are rejected. Conversely, in the sub-

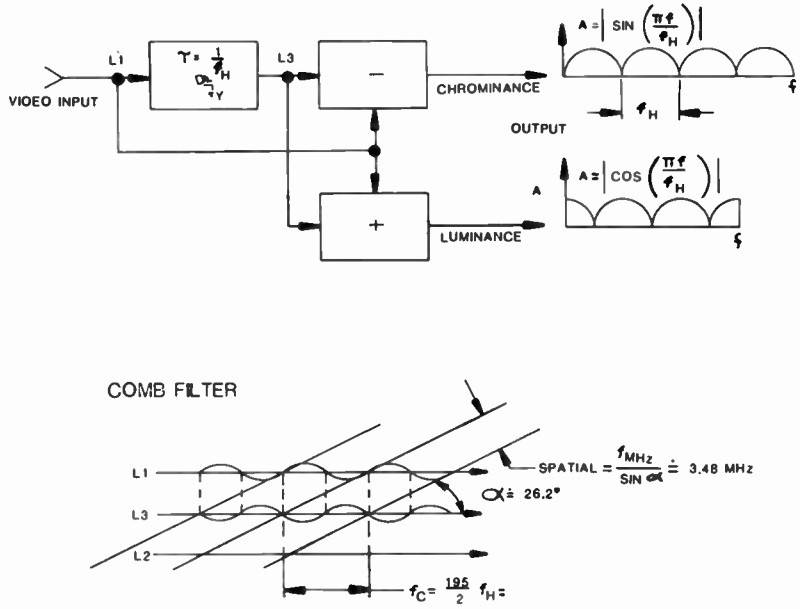


Fig. 30—Comb filter.

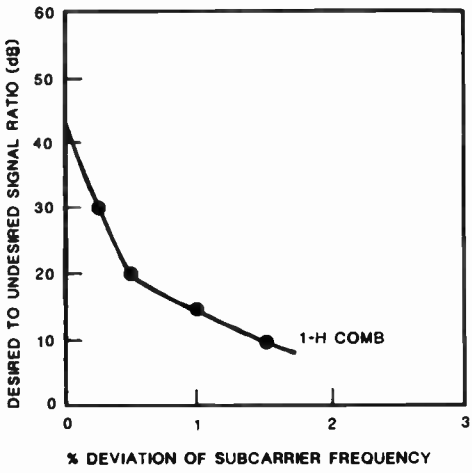


Fig. 31—Comb filter rejection ratio versus percent subcarrier frequency stability.

traction network, the luminance components are rejected and the chroma components are passed to the output.

In the player CCD 1-H delay line comb filter, the desired-to-undesired signal ratio is normally 28 dB at 1.53 MHz and more than 23 dB at the passband edges. The clock frequency of the delay line (6×1.53 MHz) is automatically adjusted to compensate for time base errors in all but extreme cases. Time base errors in excess of 8–10 nanoseconds causes 7.5 kHz horizontal striations to become visible due to comb filter imbalance. A plot of rejection ratio versus per cent frequency variation is shown in Fig. 31. As can be seen, the comb filter is reasonably insensitive to small changes in carrier frequency, i.e., less than 1%.

The Principles and Quality of the Buried-Subcarrier Encoding and Decoding System and Its Application to the RCA VideoDisc System

D. H. Pritchard, J. K. Clemens, and M. D. Ross

RCA Laboratories, Princeton, NJ 08540

Abstract—The first portions of this paper present the fundamental principles involved in an innovative color-television signal encoding and decoding technique developed specifically for the RCA VideoDisc System. The term "buried-subcarrier" refers to a method involving comb-filter techniques by which, in the composite video signal domain, the chrominance information is placed at a relatively low frequency in the luminance channel spectrum. In this way, chrominance and luminance may be easily and uniquely separated and converted to standard color-signal format at the decoder. This approach provides specific performance advantages in relation to the capacitive pickup and FM video signal processing in the areas of bandwidth, playing time, beat and interference effects, linearity, and signal-to-noise ratio.

Succeeding portions of the paper discuss these unique advantages in detail and describe typical implementation approaches to both the encoding and decoding functions that meet the objective of providing a high-quality color-television picture recording and reproduction system.

1. Introduction

The buried-subcarrier color signal encoding and decoding system was developed specifically for the RCA VideoDisc. However, the fundamental principles involved provide unique advantages in other applications having similar overall system requirements.

The term "buried-subcarrier" refers to a technique by which the chrominance subcarrier and its modulation sidebands are placed at a

relatively low frequency range within the wider-band luminance channel in such a way as to cause a minimum of interference and such that the chrominance and luminance may be easily separated at the reproducer and converted to standard NTSC format. This systems approach involves innovative comb filter techniques that may be used in any color television system where both the encoding and decoding functions can be controlled by the overall system designer.

The specific techniques by which the buried-subcarrier approach is accomplished in the RCA VideoDisc system are discussed in this paper. The discussion includes the advantages of the buried-subcarrier encoding and decoding process in conserving bandwidth, in eliminating beats and interference effects, and in meeting the constraints of the system in areas such as nonlinearities and noise. To achieve the objective of a consumer-acceptable color TV picture requires a system that coordinates the buried-subcarrier principles with the FM recording and playback technology used in the RCA VideoDisc System.

2. System Background

The buried-subcarrier encoding and decoding system adapts the principles of comb-filtering to the needs of a specific recording and playback system. Therefore, an understanding of the basic fundamentals of the comb-filtering function as applied to video signal processing in general is essential to an understanding of the reasons for the choice of particular parameters as applied to the VideoDisc system.

2.1 Basic Comb-Filter Principles

A composite television signal is periodic in nature in that it is sampled at repetitive rates of approximately 15 kHz, 60 Hz, and 30 Hz, as a consequence of the horizontal and vertical scanning process. A Fourier analysis of the signal in the video domain indicates an energy spectrum concentrated in discrete clusters with interval spacings of 15 kHz and with sub-sidebands grouped around each 15 kHz interval at multiples of 60 Hz and 30 Hz.[1] A second signal (such as the color subcarrier and its modulation sidebands in a standard NTSC format) having its own periodic energy spectrum due to sampling can be interleaved with the baseband spectrum by synchronizing the subcarrier with horizontal sync as an odd-multiple of one-half line rate (3.579 545 MHz in the case of NTSC). An example of this interleaved relationship in the case of a signal consisting of an average picture content is shown in Fig. 1.

It becomes apparent that a filter whose pass-band and stop-band characteristics resemble the teeth of a comb with appropriate spacing

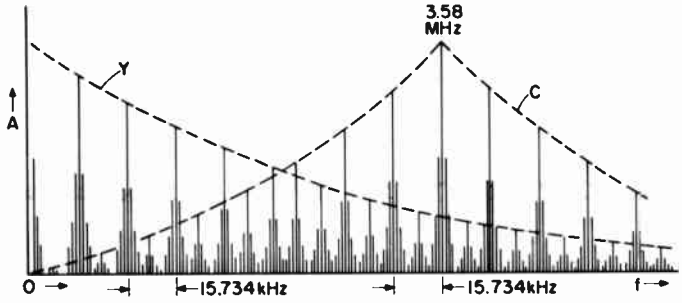


Fig. 1—Basic NTSC interleaved signal relationship for an average scene.

would be effective in separating the two signal components that make up the NTSC signal, namely, chrominance and luminance. Such a filter and its complementary form would overcome some of the performance limitations inherent in conventional separation techniques involving over-lapping low-pass (luminance) and band-pass (chrominance) filters. In the conventional approach, the luminance signal is contaminated by the chrominance subcarrier and its modulation sidebands that appear as "dot-crawl" in large color areas and at color transitions. A rejection trap is normally employed, with its attendent bandwidth and/or resolution limitation, to eliminate the dot visibility in large areas. However, dot-crawl due to chrominance sideband energy still appears at edges. Conversely, the high-frequency luminance detail information is included within the chrominance channel and converted to "cross-color" contamination by the color demodulation process which manifests itself as low-frequency beat patterns on edges of objects in the scene, patterns that are subjectively undesirable. A comb filter in the luminance channel will therefore eliminate the dot-crawl without imposing the usual 3.58-MHz bandwidth restriction. The complementary comb filter in the chrominance channel will exclude the luminance information that produces the "cross-color" beats except in the case of scene content having spatial directions of about 45° (for the case of 3.58 MHz) where line-to-line cancellation of the chrominance components cannot occur.

A two-terminal transversal filter configured as shown in Fig. 2 provides the desired comb-filter response and results in separation of interlaced luminance and chrominance signal components as follows. Non-interlaced frequency components are those that have the same phase on succeeding scan lines and are integral multiples of the horizontal line frequency. Interlaced frequencies are those whose components are exactly 180° out of phase on succeeding scan lines and are caused to be odd multiples of one-half horizontal line frequency. In the NTSC television

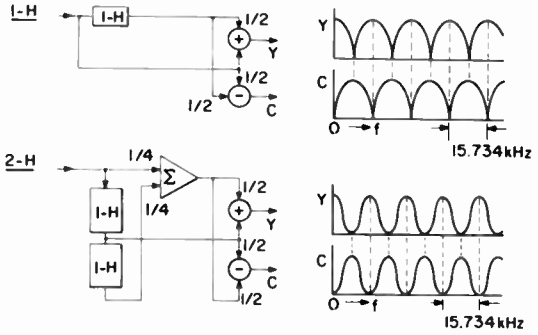


Fig. 2—Basic 1-*H* and 2-*H* comb-filter circuit configuration and pass-band characteristics.

system, the interlaced frequencies are those containing the chrominance subcarrier and its modulation sidebands.

The fundamental comb-filter circuit consists of a delayed path and a direct path, with appropriate linear summation and subtraction processing circuitry. In the case of the NTSC system (as well as the buried-subcarrier VideoDisc system) the delay element is made to be exactly one horizontal line time (1-*H*). If the amplitudes of the signals from the delayed path are made equal to the amplitude of the same signal through the direct path, while maintaining a time delay difference of 1-*H* across the frequency band of interest, a summation process results in cancellation of interlaced components while the noninterlaced components are reinforced. Conversely, subtraction (obtained by inverting one signal path followed by summation) results in cancellation of the noninterlaced components and reinforcement of the interlaced components. Thus, the adder circuit output contains luminance information free of contamination effects (dot-crawl) from the color subcarrier. Similarly, the subtractor circuit output consists of chrominance signals free of cross-color beat contamination resulting from high-frequency luminance detail entering the chrominance signal demodulation process.

If the assumption is made, as in Fig. 2, that the combing process is allowed to extend from zero to an arbitrarily high frequency, the integration of two successive line periods results in (1) a 3-dB improvement in signal-to-noise ratio (4.25 dB in the case of 2*H*, 3-line integration, comb systems) and (2) the loss of one half of the vertical resolution of scene content, since the first null in the luminance channel occurs at one-half horizontal period (7.8 kHz). It should be apparent that to successfully separate chrominance signal information from the luminance signal information, the combed frequency range could simply be limited to approximately 500 kHz on either side of the color subcarrier, thereby leaving the low-frequency vertical details untouched.

If, however, combing is accomplished down to zero frequency, the low-frequency combed output of the chrominance channel consists of exactly the vertical detail content that was removed from the luminance signal. An advantage may be realized by re-inserting the vertical detail content into the luminance channel to an amount just sufficient to restore the vertical detail to normal. Then, simply by increasing the level of the restoration signal beyond normal, vertical detail enhancement may be provided, resulting in subjective enhancement of picture sharpness. This feature may be utilized to advantage in the buried subcarrier system as will be pointed out later.

Fig. 2 indicates the basic comb-filter circuit configurations for both a 1-*H* and 2-*H* comb filter. It is important to note the difference in the filter pass-band and stop-band shapes. The 1-*H* produces essentially a full-wave rectified sinusoidal shape with relatively sharp cusp-like null points, while the 2-*H* system results in an offset sinusoidal shape. Other than the obvious implications as to circuit and component complexity, there is a tradeoff between timing stability for a given degree of unwanted signal rejection and the extent of signal-energy rejection for sideband signal components. In developing both the encoding and decoding systems, these tradeoffs must be considered from performance and cost-effectiveness points of view.

2.2 Buried-Subcarrier Application to Encoding and Decoding Systems

In the development of recording and playback approaches, the systems designer may take advantage of a unique situation in that the recording format and the playback technique may be designed to complement each other to up-grade the performance of the overall system.

In the VideoDisc case, the system parameters, such as playing time, capacitive pick-up constraints, FM system linearity, and spurious beats, are considered together to provide the optimum trade-offs for best picture-quality results. The details of these parameters will be discussed later in this paper.

The judicious use of comb-filtering principles both in the encoding process as well as in the subsequent playback decoding process forms the basis for the buried-subcarrier system concept.[2–4] Buried-subcarrier refers to composite video signal format in which the chrominance carrier is placed at a relatively low frequency within the luminance channel using comb filter techniques at the encoder to minimize interference effects—thus the term "buried" subcarrier. In addition, comb-filter techniques are advantageously employed at the decoder to uniquely separate chrominance and luminance with the least degradation effects. The advantages of such a buried-subcarrier system are that it conserves

bandwidth, increases available playing time (eases the shortest reproducible wavelength requirements), and meets channel constraints such as noise, linearity, and beat-frequency content. It is, therefore, attractive for a variety of signal encoding systems. Channel distortions affect the design decisions of video encoding, and the buried-subcarrier approach is particularly useful in the VideoDisc approach to provide high quality reproduction of color TV signals.

3. Application to VideoDisc System

In the VideoDisc System, the luminance bandwidth was chosen to be 3 MHz to be consistent with acceptable resolution in a home TV receiver. The video signal is modulated on an FM carrier and the resultant is in turn recorded on the disc.

It is possible, if desired, to separate luminance and chrominance and to record them simultaneously in separate channels in the same track. This separation causes difficulties, however, since the signals must ultimately be matrixed together, and the amplitudes, frequency responses, and even the noise characteristics of the two channels should be matched.

It was decided, therefore, to include the chrominance on a subcarrier within the luminance signal. The next choice was to determine the subcarrier frequency. There are two reasons for choosing the subcarrier frequency as low as possible. The first is that the subcarrier frequency is directly related to the FM carrier frequency and therefore to playing time on a disc, since there is a limit on the shortest recorded and reproduced wavelength. The second reason is that in an FM system, lowering the subcarrier frequency increases the chrominance SNR. The details of these considerations will be discussed in Sections 3.5, 3.6, and 3.7. Since buried-subcarrier encoding allows a low subcarrier frequency, it is very advantageous for the VideoDisc system.

3.1 Choice of Subcarrier Frequency

As previously noted, the color subcarrier in the basic NTSC system is interlaced in that it is an odd multiple of one-half horizontal line rate. It was chosen to have a value high within the luminance band (3.579545 MHz) since the spatial dimensions of one-half cycle are then relatively small, thus minimizing the visibility of the interfering "dot" pattern. With a subcarrier value at a much lower frequency, e.g., in the order of 1.5 MHz, the spatial dimensions of the interference pattern would, without special processing, be large enough to be objectionable.

Since the first requirement in the choice of the buried-subcarrier

frequency is that it be interlaced, i.e., an odd multiple of one-half horizontal line rate, comb filtering techniques are applicable in developing a satisfactory signal-processing method. In the VideoDisc system, a chrominance subcarrier frequency of 1.534091 MHz was chosen, i.e., 195 $f_H/2$ (relative to U.S. NTSC Standards). This value also meets the desire to move the color subcarrier and its ±500 kHz modulation sideband components low enough within the 3.0 MHz luminance channel, i.e., there will be no undesired color information within 3 MHz of the 5-MHz video FM carrier. The potential for possible sideband and harmonic interference beats is thereby minimized. Other advantages of lowering the chrominance subcarrier within the available bandwidth are realized in signal-to-noise improvement and in the increase in potential playing time because the shortest reproducible wavelength limitation is reduced.

For proper implementation of the buried-subcarrier system, the encoded signal prior to recording must be properly processed. First, the luminance signal is comb filtered over at least the band of frequencies that is to be shared with the subcarrier and its sidebands in order to avoid annoying crosstalk effects. A series of nulls, or troughs, are formed in the frequency spectrum of the luminance signal in which the appropriate chrominance subcarrier components may be "buried." In addition, it is desirable that the modulated color-subcarrier signal components are also comb filtered in a complementary fashion to that employed for the luminance signal. Effectively, the chrominance signal components are confined to those that will fall into the troughs of the equivalent frequency spectrum of the comb-filtered luminance channel. The appropriate luminance comb-filtered signal (for NTSC) is one with recurring peaks at multiples of half-line frequency, assuming one-half-line offset for the subcarrier frequency. Conversely, the chrominance signal has recurring peaks at odd multiples of half the line frequency and recurring nulls at multiples of line frequency.

When a subcarrier frequency of 1.53 MHz with a reasonable bandwidth of ±500 kHz for color sidebands is used, a moderately wide band (0–1 MHz) of luminance information is retained at the low end of the spectrum that is occupied solely by luminance signal components. If the system designer chooses to comb filter this region, it is possible to reinsert certain luminance-signal components in such a manner as to bring about vertical detail picture enhancement, as previously mentioned. This enhancement may be used to subjectively restore the loss of vertical detail components due to comb filtering of the spectrum containing the interlaced chrominance signal components (1 to 2 MHz), or it can be used to provide overall picture quality sharpness enhancement.

Fig. 3 indicates typical pass-band characteristics for the composite

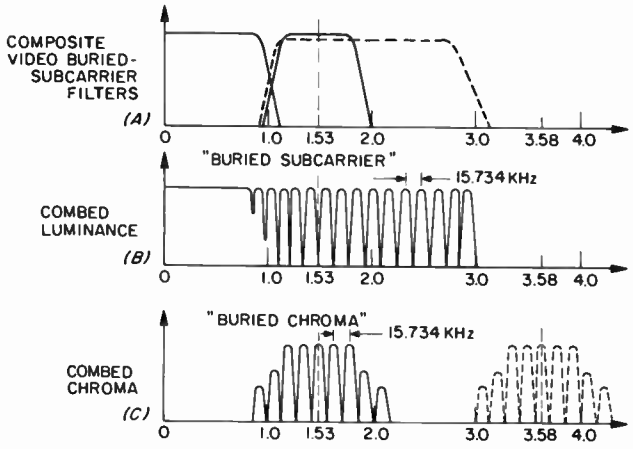


Fig. 3—Typical buried-subcarrier filter characteristics.

buried-subcarrier video channel with the buried color subcarrier at 1.53 MHz with ±500 kHz sidebands. In such a system, chrominance information may be conveyed and subsequently recovered, while sharing a midband location with luminance signal components, with substantial freedom from spurious chrominance/luminance interference effects.

The overall quality of the reproduced picture formed from a signal that has undergone a comb-filtering process is excellent. However, certain deteriorations do occur with regard to loss of resolution in the luminance channel at specific spatial angles. The angle is determined by the time period, as measured in the horizontal direction from one line to the succeeding line, for the signal content to reach a 180° line-to-line relationship. In the case of standard NTSC with a subcarrier at 3.58 MHz this angle is approximately ±45°.

As the frequency of interest is lowered, the time period along the horizontal axis for line-to-succeeding-line correspondence increases and the equivalent spatial angle for loss of signal content becomes lower (0° being the horizontal axis).

Figs. 4 and 5 show a nomograph that pictorially describes this effect. The following list describes the formation rules for such a diagram:

(1) Fig. 4 represents one quadrant of a polar diagram indicating the relationship between spatial frequencies and the corresponding electrical frequencies for a variety of angles of scene content.
(2) The origin approaches $f = \infty$ as a limit.
(3) The outer edges of the diagram approach $f = 0$ as a limit.
(4) Spatial frequencies are plotted as circles concentric with the origin linearly displaced as a function of frequency.
(5) Corresponding electrical frequencies (arising due to the rate of

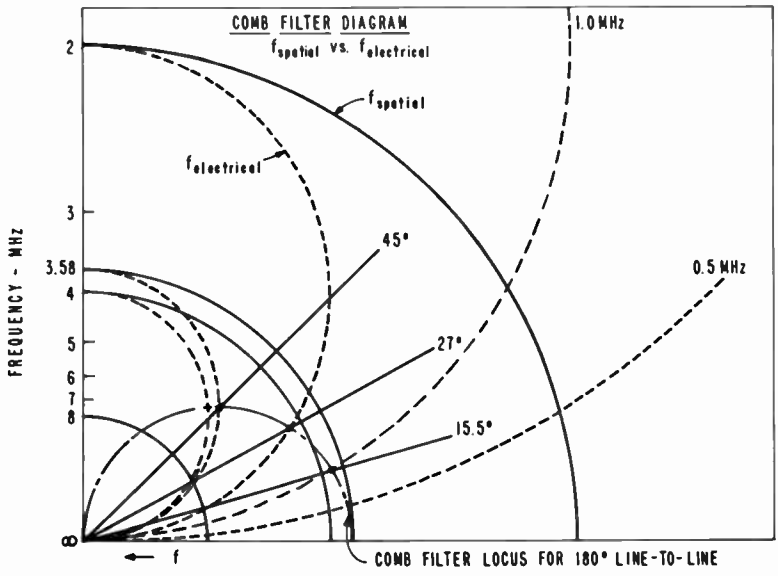


Fig. 4—Buried-subcarrier nomograph #I indicating the relationship between subcarrier frequency, spatial frequency, and spatial angles for comb nulls.

horizontal scan) are plotted as circles tangent to the origin with a diameter equal to the spatial frequency intersection with the vertical axis.

(6) Radial lines are constructed at angles corresponding to that required for 180° relationship to exist on a line-to-line basis for a given electrical frequency. (For example, 45° for 3.58 MHz, 27° for 2 MHz, and 15.5° for 1.0 MHz.)

(7) The locus for comb filter 180° relationship (rejection or acceptance points) are determined by the intersection of the radial lines and the corresponding electrical frequency circles.

Thus, maximum comb-filter rejection or acceptance may be determined for a given subcarrier frequency in relation to the physical angle of scene content ranging between the vertical and horizontal scan axis.

In addition, the effects of combed channel bandwidth and comb-filter response shape may be plotted as indicated for the example of 1.0 MHz with ±500 kHz bandwidth (50% response points) as shown in Fig. 5. The shaded area indicates the spatial angle range of resolution affected from a maximum loss at the 180° point (for 1.0 MHz) to the ±90° points (0.5 MHz and 1.5 MHz) for a ±500 kHz bandwidth system.

The additional information that should be related to the above analysis in order to determine effects upon overall picture quality is the ex-

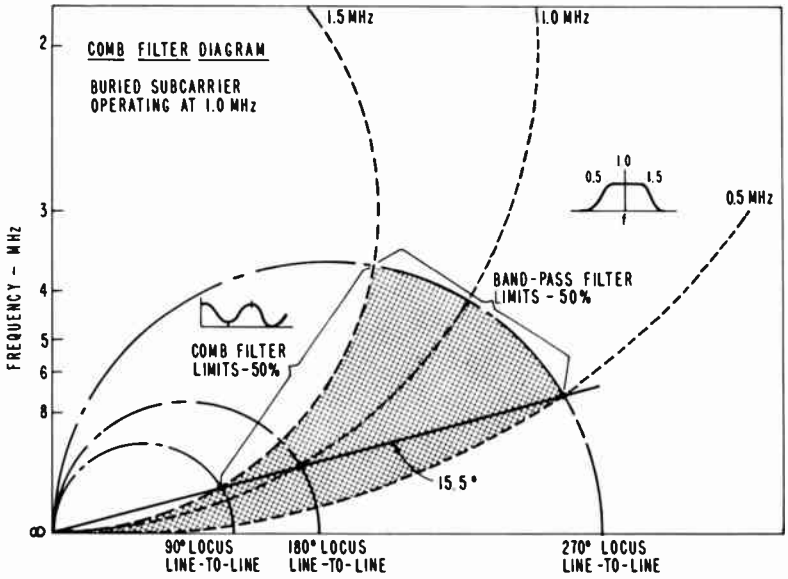


Fig. 5—Buried-subcarrier nomograph #II indicating spatial angles relative to subcarrier frequency and system bandwidths.

tent to which loss of vertical resolution signal components relative to signal bandwidth is subjectively visible. For example, the loss of vertical signal content for a frequency of 3.58 MHz (45° spatial angle) has been judged to be relatively insignificant. As the spatial angle is reduced toward the horizontal (0°), the vertical component increases and a loss, or softening, of vertical detail becomes more and more apparent until the angles reaches zero and the vertical resolution loss reaches a maximum of two-to-one (equivalent bandwidth reduction of from 4.0 MHz to 2.0 MHz).

Extensive tests were conducted to determine the optimum value of color subcarrier frequency relative to acceptable vertical resolution loss as a result of comb filtering. Buried-subcarrier frequency values ranging from 1.8 to 1.0 MHz in one-quarter megahertz intervals were investigated. In each case, a constant color modulation bandwidth of ±500 kHz was maintained and the exact value of the subcarrier frequency was chosen to be interlaced (odd multiple of one-half horizontal line frequency) in order to utilize comb-filter techniques. Subjectively, carrier frequency values substantially below about 1.5 MHz resulted in excessive loss of vertical detail, since the lower sideband extended to at least 1.0 MHz. Therefore, a value of 1.53 MHz was chosen as the optimum compromise and the spectrum range from zero to 1.0 MHz was maintained as exclusively containing luminance information. (It can be seen from

Fig. 4 that a 1-MHz electrical frequency is equivalent to about 3.7-MHz spatial frequency.)

The loss of some signal components comprising vertical detail is one factor in the choice of the use of 1-*H* or 2-*H* combing techniques. The relatively sharp, cusp-like, rejection points in a 1-*H* system result in less loss of vertical components than the sinusoidal filter shape in a 2-*H* system. For this reason, as well as for economic factors, 1-*H* combing is probably preferable at the playback end of the system. The re-insertion of vertical detail to provide sharpness enhancement is symmetrical in nature on the top and bottom of a vertical transition and is subjectively more desirable in the case of 1-*H* combing at both encoder and decoder (equivalent of a 2-*H* combing process).

3.2 Basic Encoder Subsystem

A wide variety of system implementation approaches have been conceived and evaluated. The choice of a specific approach is based upon such factors as picture performance, circuit complexity and stability, and overall system cost/performance trade-offs utilizing state-of-the-art hardware, particularly in the form of 1-*H* delay devices available.[2–4]

In the case of the encoder, the color image information to be recorded may initially appear either in NTSC encoded form or, in some cases, as the initial Red, Green, and Blue simultaneous signals. The basic functions to be performed in each case are as follows:

(1) NTSC Encoded Signal Format

(a) Comb filter that portion of the luminance channel spectrum above about 1.0 MHz to include the 3.58 MHz color subcarrier and its modulation sidebands. This process separates the original chrominance from luminance in the region around 3.58 MHz and also pre-combs the frequency spectrum around 1.53 MHz in preparation for the subsequent insertion of the buried-subcarrier color information.
(b) Comb filter the chrominance signal content for at least ±500 kHz around 3.58 MHz.
(c) Translate the combed 3.58-MHz color subcarrier and its modulation sidebands to 1.53 MHz ± 500 kHz by a heterodyne process. This process employs a 5.11-MHz reference signal synchronized with the basic 3.58-MHz system reference to produce the 1.53-MHz difference frequency product.
(d) Form the buried-subcarrier composite signal in a linear summation process by re-inserting the combed color information at 1.53 MHz,

translated from 3.58 MHz, into the appropriately precombed luminance channel. Fig. 6 is a block diagram of the fundamental functions required for such a transcoder process.

A technique for forming matched high-pass, low-pass, or band-pass filters involves, in this example, an all-pass delay element (matching the delay of the low-pass filter) whose signal output is subtracted from the low-pass filter output to form a complementary high-pass filter. This method insures proper matching of complementary amplitude and phase characteristics in subsequent signal re-assembly into a composite form. Either 1-*H* or 2-*H* comb filtering may be employed and a variety of circuit details may be used to accomplish the functions indicated. The details of the specific approach in use with the RCA VideoDisc system are discussed later in this paper.

(2) *Red, Green, and Blue Simultaneous Signal Format*

(a) A luminance signal (Y) and two color-difference signals (R-Y and B-Y) are formed in a linear resistive matrix from the original Red, Green, and Blue signals.
(b) The luminance signal is comb filtered in a frequency spectrum above about 1.0 MHz to the upper channel band limit (approximately 4.0 MHz).
(c) The color difference signals are fed individually to two balanced modulators operated from a 1.53-MHz reference signal derived from the system reference of 3.58 MHz. A composite chrominance signal at 1.53 MHz is formed by linearly summing the two double-balanced modulator outputs.

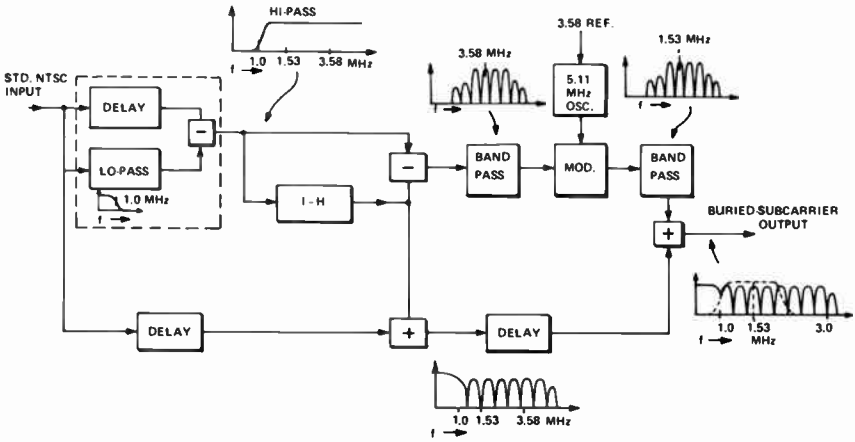


Fig. 6—Basic buried-subcarrier encoder block diagram utilizing a composite NTSC signal format.

(d) The composite chrominance signal at 1.53 MHz in a band of ±500 kHz is comb filtered.
(e) The combed luminance and combed chrominance signals are combined in a linear adder to form the composite buried-subcarrier signal.

Fig. 7 is a basic block diagram of these functions. Obviously, the chrominance signal could be formed at 3.58 MHz and then heterodyned to a frequency spectrum centered at 1.53 MHz. Also, either 1-*H* or 2-*H* combing may be employed.

3.3 Basic Decoder Subsystems

In the playback of a video disc recording of color image information encoded in the buried-subcarrier format, the manner in which the recovered information is processed depends upon the basic nature of the player apparatus. For example, in the case of the player apparatus incorporating a display device, direct decoding of the chrominance signal (after separation by appropriate comb filtering) at the buried-subcarrier frequency would be a good approach. However, where the player apparatus is intended to provide signal of a form suitable to feed any standard NTSC color television receiver, the player must incorporate some form of transcoding function that serves to translate the information recovered in buried-subcarrier format to a standard NTSC encoded form.

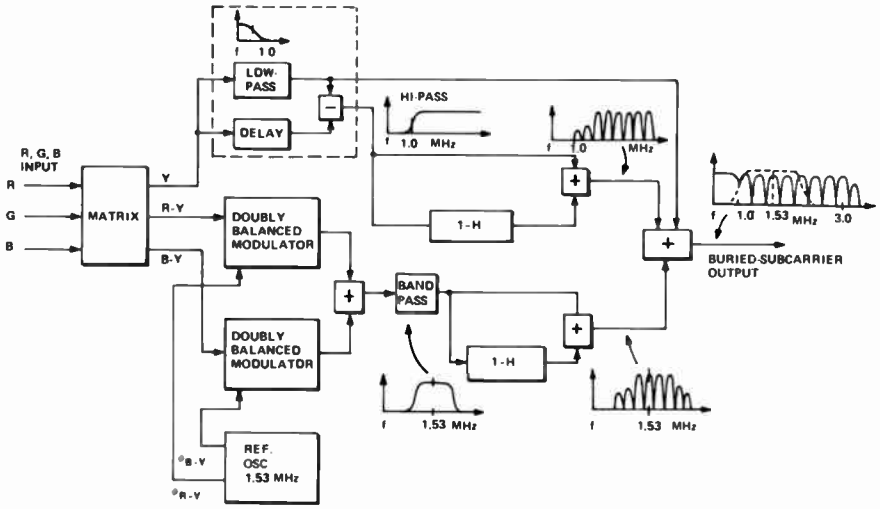


Fig. 7—Basic buried-subcarrier encoder block diagram utilizing the simultaneous Red, Green, and Blue signal format.

A wide variety of specific circuit arrangements to accomplish the above translation function are possible.[3] For discussion purposes, the decoder techniques may be grouped into the following categories:

Category A: Systems in which the luminance channel baseband signal does not pass through the 1-*H* delay elements whereas the chrominance channel must always pass through the delay elements. Such systems may then be further subdivided in relation to the band-pass frequency spectrum of the 1-*H* delay element, i.e., centered around 1.53 MHz, or centered around 3.58 MHz for chroma channel processing.

Category B: Systems in which the composite baseband buried-subcarrier signal is modulated on a reference carrier (typically having a convenient frequency of 5.11 MHz) in either vestigial or double-sideband format and then passed through one or more of the 1-*H* elements and summation processes for the comb-filtering function. These systems may further be subdivided in accordance with the method employed for demodulation or recovery of the baseband signal, i.e., envelope detection or synchronous (heterodyne) detection.

Category C: Systems employing baseband comb filtering of the composite buried-subcarrier signal followed by heterodyne translation of the chrominance information from 1.53 MHz to 3.58 MHz.

Fig. 8 is a block diagram showing the basic functions required for a buried-subcarrier decoder intended to translate the signal from buried-subcarrier form to standard NTSC format.

Representative systems of each of the variations in the above categories have been built and evaluated. All the systems can be made to

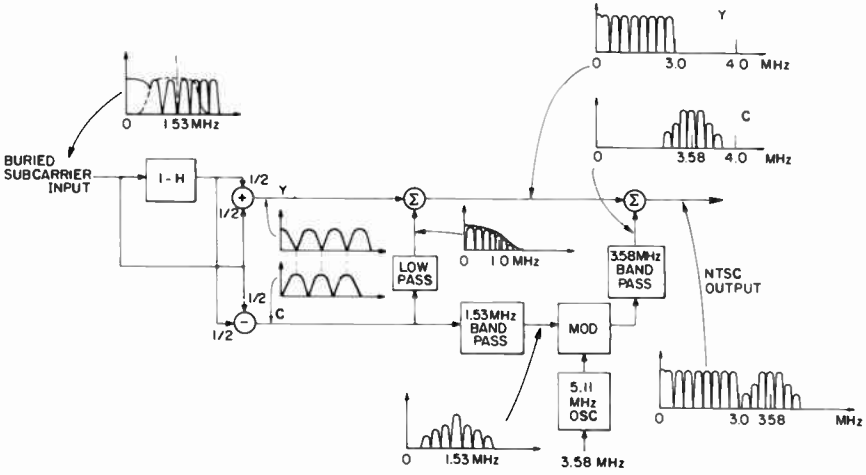


Fig. 8—Block diagram indicating the basic functions required for a buried-subcarrier decoder intended to translate the video signal from buried-subcarrier form to standard NTSC format.

perform satisfactorily, however, some approaches are more easily implemented than others. Therefore, preferred approaches are determined by factors such as simplicity, ease of operation, cost and availability of 1-*H* delay line elements, and overall cost/performance tradeoffs.

The basic functions, therefore, are comb filtering of the buried-subcarrier chrominance information, introduction of a band-pass filter encompassing the 1.53 MHz ± 500 kHz portion of the spectrum, then translating this spectrum to an equivalent band-pass centered around 3.58 MHz and referenced to a continuous, stable, 3.58 MHz local source. This step involves the use of a 5.11 MHz voltage-controlled oscillator (VCO) operated in a phase-locked-loop (PLL) configuration. The incidental timing variations existing in the chrominance signal resulting from velocity errors in the recording and playback process are exactly matched by variations introduced into the 5.11 MHz VCO continuous signal as determined by the incoming signal reference "burst." Thus, the timing *difference* is corrected and the "jitter" is removed by the heterodyne process, with the translated chrominance signal information at 3.58 MHz being stable and locked to the basic system 3.58 MHz reference.

The remaining functions include the comb filtering of the luminance channel in at least the region around the buried subcarrier (above 1.0 MHz) and, finally, the formation of the translated composite signal consisting of the non-combed (or vertical-detail-restored) luminance spectrum from zero to about 1.0 MHz, the combed luminance spectrum from 1.0 MHz to at least 2.0 MHz (usually all the way to the 3.0 MHz limit), and the addition of the combed chrominance around 3.58 MHz after it has been translated from the original 1.53 MHz range. This composite signal is in basic NTSC format and can be used directly to feed appropriate circuitry for transmission to the input of a conventional NTSC color television receiver and display system.

3.4 Typical Implementation Approaches

It is beyond the scope and intent of this paper to describe and compare all of the wide variety of specific decoder circuits that are possible and that have been evaluated. However, a few general forms will be briefly outlined in order to indicate some of the techniques used and factors involved.

A very straightforward approach falls in Category A as outlined above in which a 1-*H* delay element centered around 1.53 MHz is used in the comb filter circuit, and the baseband luminance is bypassed around the comb filter and is combed only in a frequency band about 1 MHz wide centered around 1.53 MHz. One heterodyne circuit is required to translate the chroma signal band to 3.58 MHz. A drawback of this ap-

proach is that inexpensive acoustic delay lines centered at a baseband frequency as low as 1.5 MHz are not readily available, and the use of high-frequency acoustic lines where the baseband signals must be translated up in frequency and then down again becomes relatively expensive. Fig. 9 is a block diagram of such a system.

A second method, still in Category A, is the use of a 1-*H* delay element centered at 3.58 MHz with the baseband luminance bypassed around the comb filter. This method is attractive because it makes use of available, relatively inexpensive acoustic 1-*H* delay elements. However, it requires a 5.11 MHz modulator to translate the buried-subcarrier band to 3.58 MHz and one heterodyne circuit to translate the combed portion of the luminance signal to 1.53 MHz. Fig. 10 is an example of this approach.

Perhaps a more attractive approach that also makes use of available, low-cost, acoustic 1-*H* delay lines falls in Category B. This system modulates the entire buried-subcarrier signal on a 5.11 MHz carrier. The lower sideband of this signal is then selected, in which the 1.53 MHz buried-subcarrier has been translated to 3.58 MHz prior to combing. The signal is then appropriately comb filtered and the chrominance information is directly available at 3.58 MHz. Either a synchronous detector or a straight-forward envelope detector may be used to recover the combed luminance baseband information. Thus, only one modulation process is involved. This method requires good engineering practice in

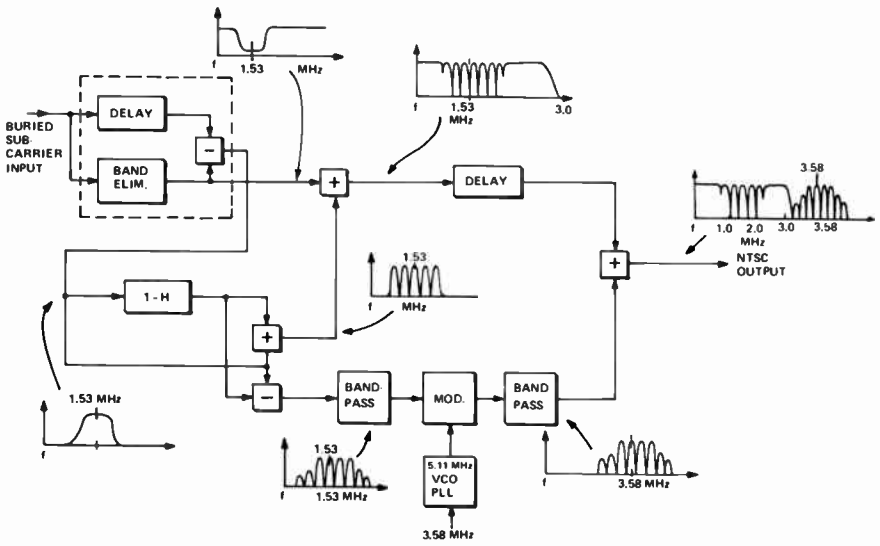


Fig. 9—Buried-subcarrier decoder block diagram, Category A. Utilizes bypassed luminance and combing directly at 1.53 MHz.

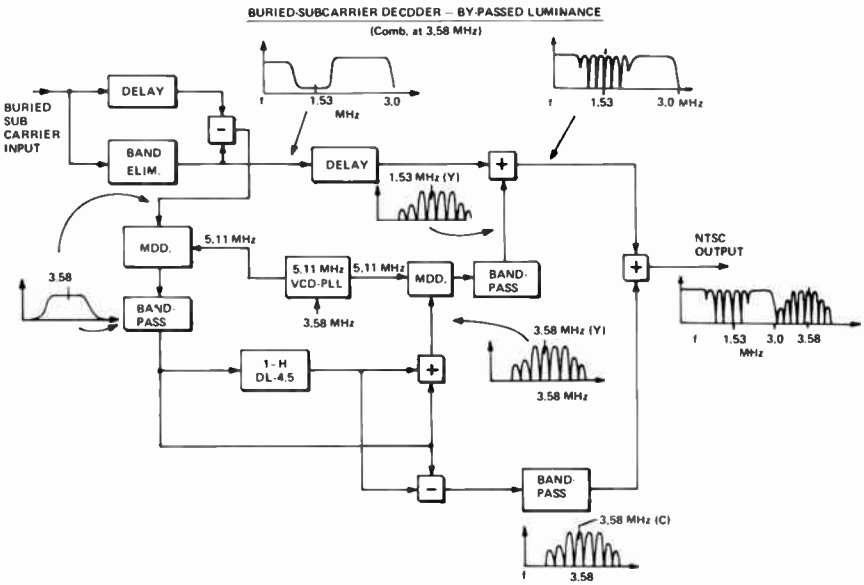


Fig. 10—Buried-subcarrier decoder block diagram, Category A. Using bypassed luminance and combing at 3.58 MHz.

the design and implementation of the pass-band filters as well as in the design of the modulator to minimize problems arising from spurious signals generated by the modulator and the use of a low percentage modulation to minimize transient response distortions. Fig. 11 shows a block diagram of such a system that has been successful and attractive from a cost/performance tradeoff point-of-view.

In each of the above examples, either 1-*H* or 2-*H* combing may be employed. A 1-*H* comb requires a higher order of timing stability than a 2-*H* system for a given depth of comb filter nulls (see Fig. 12), but the 2-*H* approach loses more of the vertical detail components. Experience has indicated that adequate timing stability is obtainable in the PLL circuit design for a 1-H system to maintain satisfactory comb filter null depths (in the order of 30–40 dB). It therefore provides a successful, low-cost approach to a practical VideoDisc player design.

It should be apparent that numerous other circuit variations are possible in the overall player–decoder design. The particular choice for the present RCA VideoDisc player is described later in this paper.

3.5 The Effects of Channel Distortions on the Subcarrier Frequency

All video disc signal-system channels suffer from some degree of non-linearity and phase shifts.[5] The result of these nonlinearities and phase shifts is the production of harmonics of signals and beats between signals.

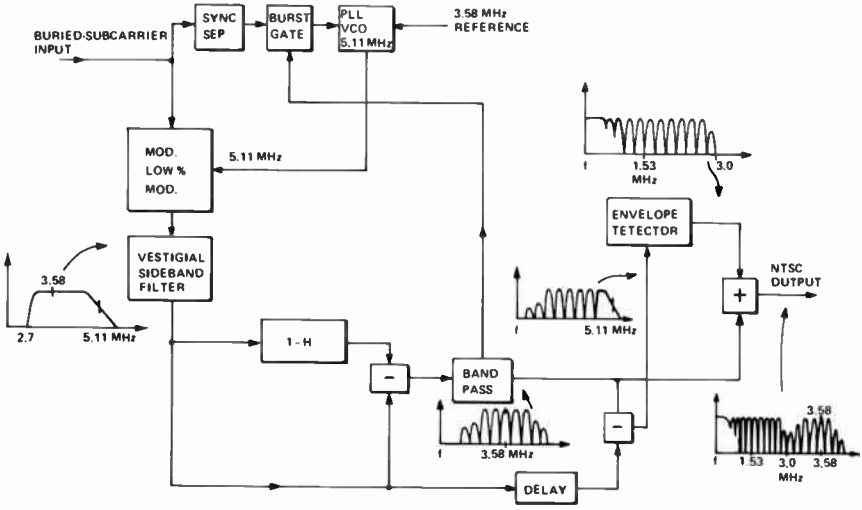


Fig. 11—Buried-subcarrier decoder block diagram, Category B. Employing a 5.11 MHz heterodyne modulator and envelope detection.

Some of the consequences of these nonlinearities on the signal system can be described by considering, for example, an FM carrier of 5 MHz modulated with a deviation of 1 MHz by a 3-MHz sinusoid. This example

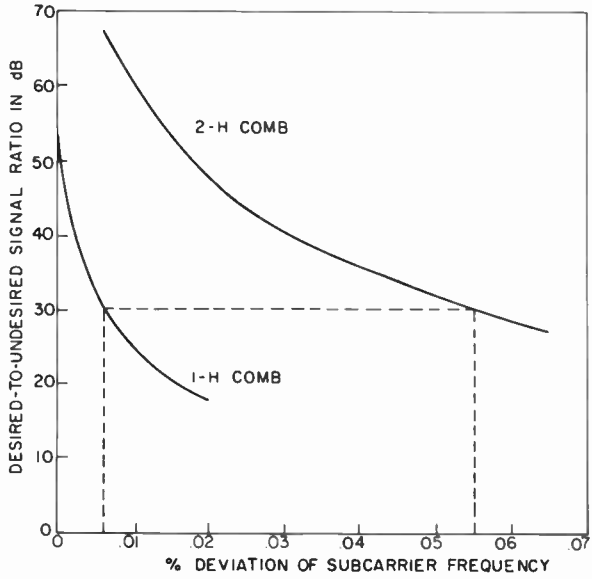


Fig. 12—Comb-filter rejection ratio comparing 1-*H* and 2-*H* configuration versus carrier frequency stability.

is shown in Fig. 13. Let $J_{n,m}$ represent the nth side frequency of the mth harmonic of the main carrier, f_c. Then the desired components at the frequencies

$$f = mf_c \pm nf_m,$$

where m = 1,2,3, n = 0,1,2,3, f_c = 5 MHz, and f_m = 3 MHz, have the amplitudes indicated as solid lines in the figure. The second side frequency of the main carrier, f_c, appears at 1 MHz due to "frequency fold over" at zero frequency.

If the channel distortions are included, a baseband component may appear at 3 MHz. The amplitude of this component will depend upon the nonlinearities and phase distortions, but it is large enough in most cases that it must be considered in the system design. Another set of components due to nonlinearities is the second harmonic of f_c at 10 MHz and its side frequencies. The components due to distortions are shown as dashed lines in Fig. 13.

In the VideoDisc, the video output from the FM modulator is limited to a 3-MHz bandwidth. As a consequence, only components within 3 MHz of the main carrier are of major concern. There are two main interfering components to consider:

$$f_{i1} = f_m = \text{baseband component} \quad [1]$$

$$f_{i2} = 2f_c - f_m, \quad [2]$$

where f_{ik} is an interfering frequency, f_m is the modulating frequency, and f_c is the main carrier frequency. This is illustrated in Fig. 13 for f_c = 5 MHz and f_m = 3 MHz. Note that the interfering frequency after demodulation, f_{id}, satisfies the inequality

$$f_{id} = |f_c - f_{i1,2}| \geq 3 \text{ MHz}, \quad [3]$$

when $f_c \geq 6$ MHz and f_m = 3 MHz. Under this condition, the interference does not appear in the video output.

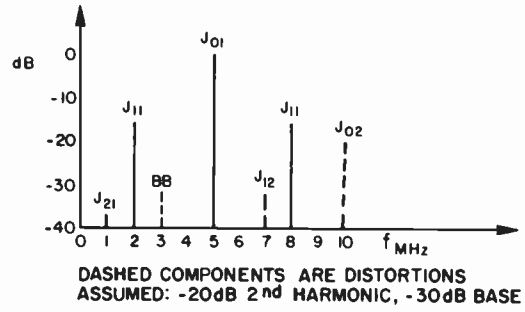


Fig. 13—Spectrum of a 5-MHz carrier modulated by a 3 MHz sinusoid.

From the baseband interference shown in Fig. 13, it would seem that the lowest FM video-carrier frequency would have to be 6 MHz for a 3-MHz luminance bandwidth. The components due to baseband interference at 3 MHz would then be 3 MHz from the carrier. If the video information is only luminance information, this carrier can, in fact, be lowered. If, for example, the picture carrier were lowered to 5 MHz, the interference from the 2- to 3-MHz region of the baseband would cause interfering signals in the 2- to 3-MHz portion of the detected picture. Interference in the frequency band is, however, hardly noticeable, partly because of its small spatial dimensions and partly because it is almost exclusively caused by edge information in the picture. As a consequence, the interference will only appear on the edges. The harmonics of these frequencies are not conveyed by the system since there is a 3-MHz low-pass final filter. In addition, the visual perception of noise components around an edge are masked by the edge. A continuous pattern of edges, such as picket fence, is a special case, but the above description still applies for the most part. The result is that little picture degradation occurs when the video carrier is lowered in frequency, violating the condition of Eq. [3] for the luminance signal. This is not the case, however, for the chroma subcarrier and its sidebands since these signals are not related to the luminance information and the interferences will be very noticeable.

The video carrier frequency thereby is directly related to the subcarrier frequency and it is desirable to keep the subcarrier frequency as low as possible.

3.6 Chrominance Subcarrier SNR Considerations

If the chroma subcarrier is added to the luminance signal, the FM carrier will, of course, be modulated by both signals simultaneously. The amplitude ratio of the two signals must be chosen carefully but there is no reason to assume the subcarrier amplitude should be a function of subcarrier frequency. The subcarrier amplitude will therefore be chosen as large as possible to maximize SNR independent of its frequency.

The power spectral density of noise at an FM discriminator output, however, is a quadratic function of the frequency (f) for an FM system with large carrier-to-noise ratio and with constant spectral density noise at the discriminator input.[9] The chrominance SNR will be determined by the ratio of the subcarrier amplitude to the noise power level at the carrier. The ratio will depend upon the noise power which is quadratic in frequency (f). Therefore, the SNR will be a function of the subcarrier frequency chosen and is another reason to keep the frequency as low as possible.

3.7 Pre-Emphasis and De-Emphasis of Luminance and Chrominance Subcarrier Amplitude

Due to the characteristics of FM modulation, the noise spectral density at the output of the discriminator is quadratic in frequency. This causes high-frequency noise in luminance as already discussed, and adds noise to the chroma subcarrier. The luminance signal is pre-emphasized before FM modulation in order to better match the FM channel noise characteristics. An *RL* pre-emphasis is used in the region between 250 KHz and 1.0 MHz, and a linear phase pre-emphasis of approximately 9 dB is employed between 1.0 MHz and 3.0 MHz.

Since the chroma subcarrier is separated from the luminance in playback, the relative levels during record can be freely chosen. The SNR of chroma is increased with an increase in carrier level. However, if the chroma carrier level is too high, any components of chrominance that are left in the luminance after comb filtering will be noticeable in the luminance. The reverse is also true if the chroma subcarrier is too low in amplitude.

The ratio of the chroma carrier level and luminance level was determined experimentally. The level was not critical and was a balance of smoothly changing parameters. The chroma carrier level was chosen to be 2.7 times the amplitude of the NTSC standard value. Of course, all levels are restored to standard NTSC levels in the player decoding circuitry.

4. Contemporary Encoder

The encoder and player systems that have evolved during the development period have attempted to apply the basic concepts in the most reliable manner. Picture quality was the main objective of the encoders, while the best performance at the lowest possible cost was the primary consideration in the players.

4.1 NTSC Version

Fig. 14 shows a block diagram of an NTSC-to-buried subcarrier encoder. Two wideband 1-*H* delay lines are used to form a baseband 2-*H* comb format of the chrominance signal. The 2-*H* comb insures the removal of high-frequency luminance signals to prevent "cross-color" beat patterns from appearing on edges of objects in the picture. To prevent over-deviation of the FM carrier, the signal is clipped at this point. A 3.58 MHz bandpass filter at the output of the matrix selects only the

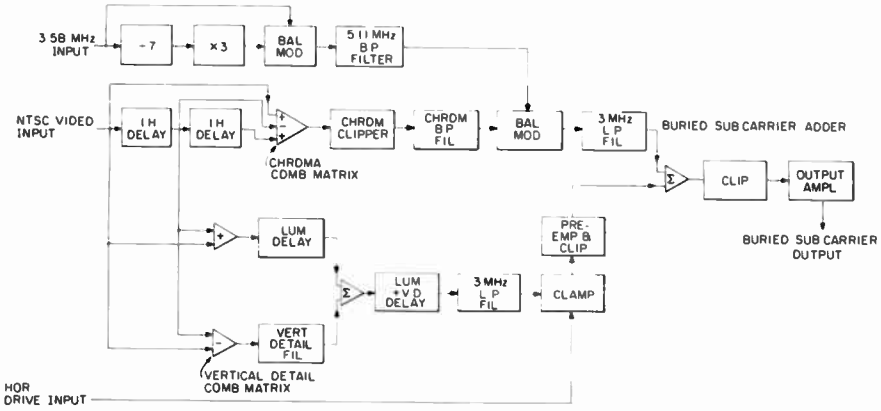


Fig. 14—Block diagram of a contemporary design of an NTSC to buried-subcarrier encoding system.

chrominance signal for translation to the 1.53-MHz buried-subcarrier frequency. The translation is accomplished by mixing the 3.58-MHz combed chrominance signal with a 5.11-MHz carrier and selecting the lower sideband components. A low-pass filter is used for this selection process to insure good rejection of the carrier and upper sideband components.

The 5.11-MHz carrier used in the translation process is locked in frequency with the system reference 3.58-MHz subcarrier. One method to accomplish this is to divide the 3.58-MHz subcarrier by 7 and multiply by 3 to obtain a 1.53 MHz subcarrier. This subcarrier is then mixed with the original 3.58-MHz subcarrier and the upper sideband is selected by means of a bandpass filter.

The luminance channel of the encoder uses a 1-*H* comb to remove the chrominance components and to prepare spaces for the buried subcarrier with a minimum loss of resolution. The output of the comb matrix is inverted and delayed such that after the vertical detail signal is added, the resultant will be combed above 900 kHz and delayed by 1-*H* and uncombed below 900 kHz. The output of the luminance adder is delayed to compensate for the bandpass filter in the chrominance channel and then band limited by the 3-MHz low-pass filter.

The output of the low-pass filter is clamped so that pre-emphasis and clipping can occur. To improve the high frequency signal-to-noise ratio, the luminance signal is pre-emphasized approximately 12 dB using a double break-point network with time constants of 0.64 μsec and 0.16 μsec, respectively. To prevent over-deviation of the video FM modulator, the luminance signal is clipped at −66 IRE units in the direction of sync and +140 IRE units in the white direction. It should be noted that the

luminance signal is clipped before it is added to the chrominance signal. This prevents undesirable edge effects, which could occur in the picture due to line-to-line phase reversals of chroma signals when the luminance signal is near the clipping level.[8]

The pre-emphasized and clipped luminance signal is then added to the 1.53-MHz chrominance signal. The level of the chrominance is increased three times above NTSC values to improve the signal-to-noise ratio and to minimize crosstalk from the pre-emphasized luminance. The resultant is again clipped at −70 and 144 IRE units to prevent overdeviation of the video FM modulator.

4.2 R, G, B Version

Fig. 15 shows a block diagram of a Red, Green, and Blue simultaneous signal to buried-subcarrier encoder. In this system, the luminance signal (Y) and the two color-difference signals (R-Y and B-Y) are formed in a linear resistive matrix from the original Red, Green, and Blue signals. The luminance and vertical detail signals are processed the same as in the NTSC encoder except for the addition of composite sync.

The color difference signals are low-pass filtered at 500 kHz to limit their bandwidth. These signals then modulate separate 1.53-MHz car-

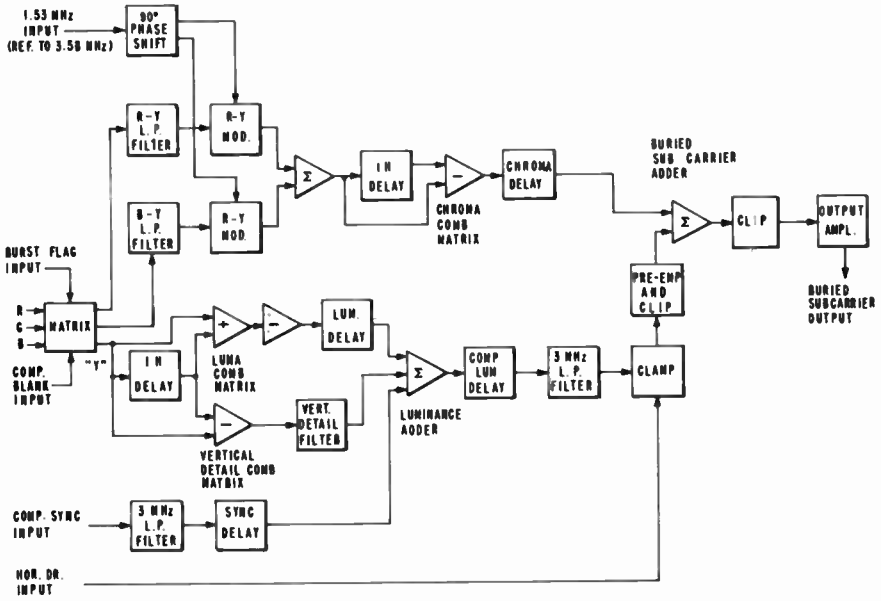


Fig. 15—Block diagram of a contemporary design of an R,G,B signal source to buried-subcarrier encoder.

riers having a phase difference of 90°. The resulting subcarrier components are then summed to form the chrominance signal. It should be noted that in the NTSC system, the lower sideband of a 5.11-MHz carrier was selected as the chrominance signal. For the RGB encoder to be compatible, the phase of the color difference carriers referenced to burst must be the reverse of the NTSC standard.

Since no luminance components exist in the color difference signals, there is no need to use a 2-*H* comb and, therefore, a 1-*H* comb is used. The remainder of the RGB encoder is essentially similar to the NTSC encoder previously described.

5. Developmental Player

The developmental VideoDisc player is described in Ref. [6]. Fig. 16 shows a block diagram of the video processing circuit portions only. It can be seen that the player is a combination of Category A and Category B systems, since part of the luminance signal does not pass through the 1-*H* line while another portion of the luminance signal is translated to a reference carrier frequency for processing.[7] The output of the AM modulator consists of a 5.11-MHz carrier with both modulation sidebands extending to 3 MHz on either side of the carrier. The low-cost glass 1-*H* delay line effectively limits the upper sideband and passes the lower sideband only. It should be noted that the chrominance subcarrier at 1.53 MHz below the 5.11-MHz carrier coincides with the NTSC subcarrier value of 3.58 MHz. Subtracting the input and output of the 1-*H*

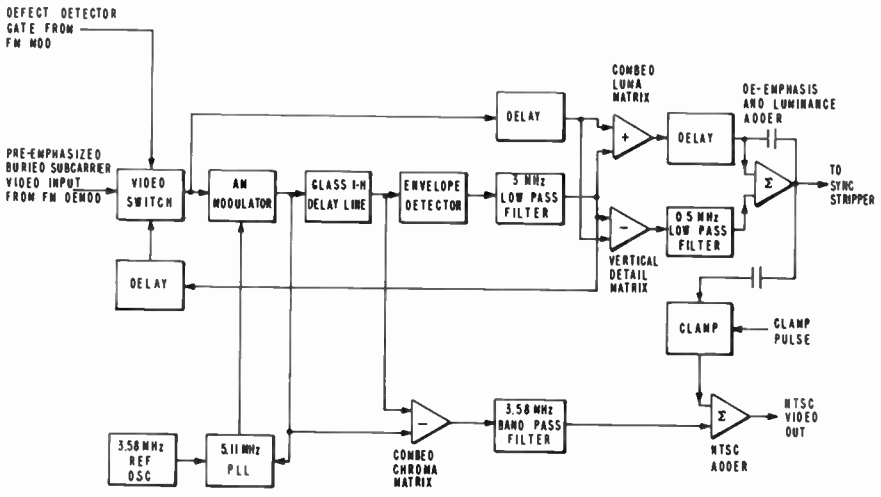


Fig. 16—Block diagram of the developmental design of the video processing portion of a buried-subcarrier player.

line provides the comb-filtering process and also allows the NTSC chrominance to be recovered directly with a bandpass filter.

The delayed buried-subcarrier signal is recovered from the 1-*H* delay line by means of an envelope detector and low-pass filter. The 1-*H* delayed signal and the input signal, after a small delay to compensate for the low-pass filter, are linearily matrixed together to form the combed luminance and vertical detail signals similar to the method used in the encoder. It should be noted that the 1-*H* delayed buried-subcarrier signal can also be used for defect substitution at the input video switch. The delay line in this path is used to shift the phase of the 1.53 MHz chroma signal by 180° such that its phase is the same as the undelayed signal from the video FM demodulator.

The vertical detail and luminance adder is also used to de-emphasize the resultant signal. The output of the adder is fed to a sync stripper which develops a clamp pulse to restore the dc reference of the luminance signal. The clamped luminance and the 3.58 MHz combed chrominance are then summed to form the composite NTSC video output which is used to drive an rf modulator for transmission to a standard TV receiver.

6. Contemporary Player

Fig. 17 shows a simplified block diagram of the buried-subcarrier video processing portion of a contemporary player. It should be noted that this player is a Category C system. The overall VideoDisc player details are described in Ref. [10]. The defect-corrected pre-emphasized BSC video from the FM demodulator is processed by a 1-*H* comb filter employing CCD technology.[11,12] The combed chrominance signal is fed through a low-pass filter to form the vertical detail signal which is reinserted into the combed luminance signal.

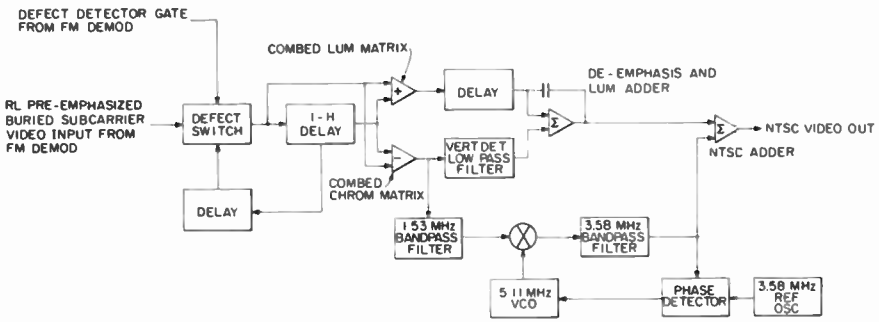


Fig. 17—Block diagram of a contemporary design of the video processing portion of a buried-subcarrier player—Category C.

The combed chrominance signal is also fed to a 1.53-MHz bandpass filter. The output of this filter drives a PLL multiplier which performs the function of translating the 1.53 MHz to 3.58 MHz. As mentioned previously, this step involves the use of a 5.11-MHz voltage-controlled-oscillator (VCO) operated in a phase-locked-loop (PLL) configuration. The translated chrominance signal is then de-emphasized and summed with the de-emphasized luminance signal to form the NTSC composite signal output of the player.

7. Summary

An overall systems approach was utilized in the development of the buried-subcarrier color-television-signal encoding and decoding system for the RCA VideoDisc. Appropriate utilization of the buried-subcarrier principles, coordinated with the FM recording and playback technology in the RCA VideoDisc system, meets the objective of providing high-quality, consumer-acceptable, color TV picture reproduction.

The systems approach involves an innovative application of comb-filter techniques that provides advantages in the areas of bandwidth conservation, longer playing time, elimination or minimizing beat and interference effects, and successful performance within the specific constraints of the system in areas such as nonlinearities and signal-to-noise improvement.

These unique advantages of the buried-subcarrier approach, as realized in the VideoDisc system, are applicable to other video systems particularly where an FM communications process is involved and where both the encoding and decoding functions may be controlled by an overall systems design. For example, magnetic tape recording and playback systems and certain satellite TV communications links involving FM video processing are especially adaptable.

In the VideoDisc system, the buried-subcarrier encoding process utilizes the capacitive pickup coupled with FM video processing to its best advantage in providing high-quality color-television recording and reproduction.

Acknowledgments

The development of the buried-subcarrier encoding system and its application to the RCA VideoDisc involved the dedication and contributions of a great many engineers and scientists at both Princeton Laboratories and at Indianapolis, without whose efforts success could not have been achieved. The authors would especially like to acknowledge the contributions of J. Amery, J. P. Bingham, J. S. Fuhrer, and A. C. Schroeder.

References:

[1] P. Mertz and F. Gray, "A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal to Telephotography and Television," *Bell System Tech. J.*, **13**, p. 464, July 1934.

[2] D. H. Pritchard, "Color Information Translating Systems," US Patent No. 3,872,498, March 18, 1975.

[3] D. H. Pritchard, "Comb Filter for Video Processing," US Patent No. 3,996,606, Dec. 7, 1976.

[4] J. K. Clemens, "Information Records and Recording/Playback Systems Therefore," US Patent No. 3,842,194, Oct. 15, 1974.

[5] J. K. Clemens, "Capacitive Pick-up and the Buried-Subcarrier Encoding System for the VideoDisc," *RCA Review*, **39**, No. 1, p. 33, March 1978.

[6] R. N. Rhodes, "VideoDisc Player," *RCA Review*, **39**, No. 1, p. 60, March 1978.

[7] J. Amery, "Color Image Signal Processing Circuits," US Patent No. 3,969,757, July 13, 1976.

[8] M. D. Ross, "Color Video Signal Processing Circuits," US Patent No. 4,096,513, June 20, 1978.

[9] M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill Book Co., Inc., 1959, pp. 301, 302.

[10] W. Workman, "The VideoDisc Player," Conf. Record, Electro 1981, April 1981.

[11] D. H. Pritchard, "A CCD Comb Filter for Color TV Receiver Picture Enhancement," *RCA Review*, **41**, p. 3, March 1980.

[12] D. J. Sauer, "Design and Performance of a CCD Comb Filter IC," *RCA Review*, **41**, p. 29, March 1980.

The Influence of Carrier-to-Noise Ratio and Stylus Life on the RCA VideoDisc System Parameters

M. D. Ross, J. K. Clemens, and R. C. Palmer

RCA Laboratories, Princeton, NJ 08540

Abstract—In the RCA VideoDisc System, the video and audio information is stored in depressions recorded in grooves that are pressed into plastic discs. During playback of the disc, a capacitive pickup system converts these depressions into electrical signals. The carrier-to-noise ratio obtainable from these signals during the useful life of the pickup stylus influences the choice of many of the system parameters, such as minimum recorded wavelength, groove pitch, playing time, inner recorded radius, rotational velocity, FM deviation, maximum carrier frequency, and video signal-to-noise ratio.

It is shown that the carrier-to-noise ratio data can be used to make trade-offs between the minimum recorded wavelength, maximum carrier frequency, and rotational velocity to obtain a desired value of the weighted video signal-to-noise ratio. For maximum stylus life, the rotational velocity should be chosen as low as possible consistent with desired signal-to-noise ratio and manufacturing margins.

1. Introduction

In the RCA VideoDisc System, the video and audio information is stored in depressions recorded in grooves that are pressed into plastic discs. During playback of the disc a capacitive pickup system converts these depressions into electrical signals as described in Ref. [1]. The carrier-to-noise ratio obtainable from these signals during the useful life of the pickup stylus influences the choice of many of the system parameters, such as minimum recorded wavelength, groove pitch, playing time, inner recorded radius, rotational velocity, FM deviation, maximum carrier frequency and video signal-to-noise ratio.

The RCA Capacitance Electronic Disc (CED) System in its present form evolved over many years[2] with substantial changes occurring in such basic techniques as pickups, recording instruments, mastering, and disc materials. Each of these changes affected the carrier-to-noise ratio of the signal channel whereby information is stored and reproduced in the VideoDisc system. As these changes occurred, it was necessary to make appropriate tradeoffs of the system parameters in order to obtain our basic goal of good picture quality, long playing time, and low cost of both player and disc.

2. Relationship Between Parameters

The mean signal-to-noise power ratio, S/N, for an FM system with de-emphasis is given by Schwartz[3] as

$$\frac{S}{N} = \frac{3\beta^2}{D}\frac{S_c}{N_c}, \qquad [1]$$

where β is the modulation index, D is the de-emphasis factor, and S_c/N_c is the carrier-to-noise ratio of an AM system with the same maximum modulating frequency, carrier power, and noise power spectral density. The modulation index is given by

$$\beta = \frac{\Delta\omega}{\omega_m} = \frac{\Delta f}{f_m}, \qquad [2]$$

where $\Delta f = \Delta\omega/2\pi$ is the peak frequency deviation and f_m is the maximum bandwidth of the modulation signal. Therefore, for a fixed maximum modulation frequency and de-emphasis factor, the signal-to-noise ratio obtainable from an FM system is directly proportional to both the carrier-to-noise ratio at the input to the system and the square of the peak frequency deviation.

In the VideoDisc System, if the lowest carrier frequency is fixed, the peak frequency deviation is proportional to the maximum carrier frequency, f_{max}, and the carrier-to-noise ratio obtainable at that frequency influences the choice of other system parameters. For a constant number of disc rotations per minute, RPM, the minimum groove velocity, v_{min}, will occur at the inner recorded radius, R_i, and is given by

$$v_{min} = 2\pi R_i \cdot \text{RPM}. \qquad [3]$$

Units conversion factors are not shown in these equations but are included in all calculations. The minimum recorded wavelength, λ_{min}, is related to the minimum groove velocity by

$$\lambda_{min} = \frac{v_{min}}{f_{max}} = \frac{2\pi R_i \cdot \text{RPM}}{f_{max}}. \qquad [4]$$

In the VideoDisc System, recording, replication, and pickup apertures cause a reduction in the carrier-to-noise ratio as the wavelengths become smaller. From Eq. [4] it appears desirable to increase the RPM and reduce f_{max} to increase the minimum wavelength and, thus, the carrier-to-noise ratio. However, from Eq. [1] the signal-to-noise ratio increases linearly with carrier-to-noise ratio but increases as the square of the peak frequency deviation, which is proportional to f_{max}. In general, a trade-off must be made between RPM and f_{max} to keep the carrier-to-noisie ratio at the minimum wavelength from falling below the threshold of the FM system,[3] which would cause undesirable defects in the picture, and to obtain a signal-to-noise ratio that is high enough to be consistent with good picture quality. Of course, sufficient margin must be maintained above threshold to allow for manufacturing tolerances. Longer wavelengths are easier to record and replicate and cause less problems in the player from signal loss due to stylus lifting from either debris or severe disc warp.[1]

The stylus tip geometry used in the RCA VideoDisc System is referred to as keel-shaped. The sides of this stylus are relatively straight and the width remains essentially constant with wear. If a stylus is designed to fit a given groove width, its life will be determined by the volumetric wear rate and the length and height of the keel. In the RCA VideoDisc System, the width of the groove is essentially equal to the distance between the grooves or the groove pitch, P_G, which for a constant RPM disc is given by

$$P_G = \frac{R_0 - R_i}{T_p \cdot \text{RPM}}, \qquad [5]$$

where R_0 is the outer recorded radius, R_i is the inner recorded radius, and T_p is the playing time per side. For a fixed recorded area and playing time, the groove pitch will increase as the RPM is lowered. The larger groove width will result in increased stylus life. Also with lower RPM the stylus travels less distance per hour of play which results in less wear.

In general, it is desirable to keep the RPM as low as possible to increase stylus life. However, Eq. [4] shows that for a fixed f_{max}, reducing the RPM causes a reduction in λ_{min}, which results in a reduction of the carrier-to-noise ratio due to aperture effects and a reduction in manufacturing tolerances. Of course, f_{max} could be reduced to increase λ_{min}, but Eq. [1] shows that this would reduce the signal-to-noise ratio obtainable from the disc, since β is proportional to f_{max}. Thus, a trade-off between RPM, λ_{min}, and f_{max} is required to obtain maximum stylus life at a desired level of signal-to-noise ratio and manufacturing margins.

3. Impetus for Development of Buried Subcarrier System

The desire both to keep the RPM as low as possible to increase stylus life and to make λ_{min} as large as possible to provide manufacturing margins led to the development of the buried-subcarrier signal encoding and decoding system for the RCA VideoDisc.[1,4] While λ_{min} occurs at f_{max} (Equation [4]), the lowest carrier frequency that can be used is also an important consideration in FM systems where the carrier is located close to the maximum modulating frequency. These systems typically use a low modulation index such that only the first-order sidebands are significant.[3] They also typically use an FM demodulator that employs frequency doubling to reject carrier frequencies, f_c, close to the maximum modulating frequency, f_m. The frequency doubler output will contain both the desired frequency f_m plus the doubled carrier and sideband frequencies $2f_c \pm nf_m$. The amplitude at the frequency $2f_c - 2f_m$ is generally large enough to cause a disturbance in the picture and must be removed by the video filter. If this frequency is set equal to or higher than the maximum modulating frequency, then

$$2f_c - 2f_m \geq f_m \quad \text{or} \quad f_c \geq (3/2)\, f_m.$$

This is for the ideal case. If second harmonic or baseband components are present as in the VideoDisc channel, then distortion components will appear at $f_c - f_m$ in the video domain.[1,4] If $f_c \geq 2f_m$ the distortion components will fall outside the video filter. Thus the lowest carrier frequency should be higher than $(3/2)f_m$ for the ideal case and higher than $2f_m$ when distortions are present.

In the RCA VideoDisc System, the luminance bandwidth was chosen to be 3 MHz to be consistent with good resolution in a home TV receiver. If only the luminance information were recorded, then the lowest carrier frequency could be 4.5 MHz in the ideal case and 6 MHz if distortions are considered. If color information were recorded using the standard NTSC format with the color subcarrier at 3.58 MHz and an overall bandwidth of 4.2 MHz, then the lowest carrier frequency would be 6.1 MHz in the ideal case and 8.4 MHz if distortions are considered. In addition to the reduced wavelengths required to record the NTSC color subcarrier, a second disadvantage is that in an FM system the signal-to-noise ratio available at the color subcarrier frequency decreases as the subcarrier frequency moves further away from the main luminance carrier. These disadvantages led to the investigation of many methods of recording color information in the RCA VideoDisc System.

The buried-subcarrier composite color signal encoding and decoding system developed for the RCA VideoDisc uses comb-filter techniques to place the color subcarrier at a frequency of 1.53 MHz with modulation

sidebands extending from 1 to 2 MHz. The advantages of this approach are that it can be recorded in a 3 MHz bandwidth, which results in a low minimum FM carrier frequency, and the reduced subcarrier frequency provides improved signal-to-noise ratio of the chrominance over the standard NTSC composite system.

Extensive viewing tests using the buried subcarrier system in the FM domain have shown that with the baseband distortions present in the VideoDisc, the lowest black-level frequency, f_{black}, can be placed such that the highest chrominance modulating frequency, f_{mc}, of the lower sideband color subcarrier, f_{sc}, is greater than the maximum video baseband component, f_m, or

$$\begin{aligned} f_{black} - f_{sc} - f_{mc} &\geq f_m \\ f_{black} &\geq f_m + f_{sc} + f_{mc} \end{aligned} \qquad [6]$$

For f_m = 3 MHz, f_{sc} = 1.53 MHz, and f_{mc} = 0.5 MHz, this results in a black-level frequency of 5 MHz. Placing the black-level frequency at 5 MHz instead of at 6 MHz as required by the baseband distortions causes an overlap between the baseband luminance interference from 2 to 3 MHz and the lower sideband luminance information that is 2 to 3 MHz from the black-level carrier. Interference in this frequency range is hardly noticeable in the picture partly because the signal energy is low and partly because it is almost exclusively caused by the edge information in the picture. Note, however, that this would not be the case if the color subcarrier appeared in this range. Since the subcarrier and its sidebands are not necessarily related to the luminance information, these signals would cause unacceptable beats to appear in flat areas of the picture. Applying Eq. [6] to the standard NTSC system with f_m = 4.2 MHz, f_{sc} = 3.6 MHz and f_{mc} = 0.5 MHz would result in a black-level frequency of 8.3 MHz. The reduced black-level frequency obtained by using the buried subcarrier composite color system results in a proportional reduction in the maximum modulating frequency required, thus making λ_{min} as large as possible and allowing the use of a low RPM to increase stylus life.

4. Signal-to-Noise Ratio Definitions

Eqs. 1 and 2 show that for a fixed maximum modulation frequency, f_m, the signal-to-noise power ratio obtainable from an FM system is directly proportional to both the carrier-to-noise ratio and the square of the peak frequency deviation and inversely proportional to the de-emphasis factor. To make the proper trade-offs between these variables, it is desirable to know the value of signal-to-noise ratio required. In an FM system, the noise power at the output of the detector increases as the

square of the frequency and is modified by the de-emphasis factor and other system components, such as filter responses. In comparing a system that has a non-flat noise response with test results that in general are for systems with a flat noise response, it is common pratice to use a noise weighting function. The weighting function attempts to make equal measured values denote equal visual interfering effects regardless of the shape of the noise spectrum being measured. The noise weighting function used for the RCA VideoDisc is that adopted by the CCIR in 1974.[5]

The definition of signal-to-noise ratio used for the RCA VideoDisc is the peak-to-peak signal amplitude measured from blanking to white level versus the rms noise weighted with the CCIR function in a 4.2-MHz bandwidth. This definition is different from the CCIR definition in that a 4.2-MHz bandwidth is used instead of the specified 5-MHz bandwidth. 4.2 MHz is chosen because most of the noise measuring equipment for NTSC systems uses this bandwidth. The CCIR noise weighting factor for flat noise is 7.4 dB in a 5-MHz bandwidth and 6.8 dB in a 4.2-MHz bandwidth.

Extensive subjective tests have been performed using experienced viewers to determine the effect of a flat noise spectrum on the impairment to television pictures.[6] These tests showed that for a viewing distance equal to eight times the picture height, a signal-to-noise ratio equal to or greater than 39 dB resulted in a picture in which 50% of the viewers said the noise was just perceptible and more than 90% of the viewers said the noise was not objectionable. In these tests, the peak-to-peak signal was measured from sync tip to white level and the rms noise was unweighted in a 4.2-MHz bandwidth. Subtracting 3 dB to convert the signal from sync tip to blanking level and adding 6.8 dB for noise weighting converts this data to a weighted signal-to-noise ratio of 42.8 dB. Similar tests conducted at a viewing distance equal to four times the picture height resulted in a weighted signal-to-noise ratio of 45 dB to give the same 50% and 90% viewer responses.[7] Later tests at a viewing distance of four times the picture height resulted in a weighted signal-to-noise ratio of 50 dB to obtain the same viewer responses.[8]

While noise weighting attempts to denote equal impairments due to the presence of noise, the effects of the signal-to-noise measured at the output of the RCA VideoDisc player cannot be directly compared with the previously determined signal-to-noise ratio standards. Since the buried subcarrier system employs comb-filtering techniques for video processing, two additional factors must be considered. First, the chrominance signal is converted from 1.53 MHz to 3.58 MHz in the player. The noise in the chrominance signal is the sum of the chrominance noise from midband and any luminance noise at 3.58 MHz. Since

the previous noise shapes do not account for this type of processing, viewing test and measurements were conducted to determine a comparison with flat noise data. Secondly, comb filtering a video signal with a 1-H delay line improves the measured signal-to-noise ratio by 3 dB if consecutive scanning lines carry identical picture information. This is because the delayed and undelayed random noise add only on a power basis. While combing improves the measured signal-to-noise ratio by 3 dB, the visibility of combed noise is not reduced by the same amount. Viewing test and measurements were conducted to show that the visible signal-to-noise ratio improvement due to combing is approximately 1.25 dB. Also, the luminance signal is not completely combed due to the addition of the vertical detail signal.[1,4]

The net effect of the previous considerations is that the equivalent visible signal-to-noise ratio of the RCA VideoDisc Player is approximately 2.5 dB less than the measured signal-to-noise ratio at the output of the player. In other words, it would be necessary to measure a signal-to-noise ratio of approximately 45.5 dB at the output of the player in order to obtain performance equivalent to a weighted signal-to-noise ratio of 42.8 dB. As noted previously, this latter value is such that 50% of the viewers at a distance of eight times the picture height would say the noise was just perceptible and more than 90% would say the noise was not objectionable. The worst-case weighted signal-to-noise ratio acceptable at the innermost radius of the disc has been specified to be 46 dB.

5. Relationship between Carrier-to-Noise Ratio and Signal-to-Noise Ratio

As stated in Sec. 2, a trade-off between RPM, λ_{min}, and f_{max} is required to obtain maximum stylus life at a desired level of signal-to-noise ratio. Combining Eqs. [1] and [2] gives

$$S/N = \frac{3}{D}\left(\frac{\Delta f}{f_m}\right)^2 \frac{S_c}{N_c}. \qquad [7]$$

The worst case weighted signal-to-noise ratio acceptable at the innermost radius of the disc has been established at 46 dB, the maximum luminance bandwidth (f_m) has been established at 3 MHz and from Eq. [6] the black-level frequency is established at 5 MHz. The de-emphasis factor, D in Eq. [7] involves a trade-off between picture quality and improved signal-to-noise ratio that will be discussed later. For a fixed black-level frequency, the peak frequency deviation, Δf in [7] is proportional to the frequency chosen for the white level, f_{max}. From Eq. [4] for a given f_{max}, the minimum wavelength, λ_{min}, is proportional to the RPM for a given

inner recording radius, R_i. Due to recording, replicating, and pickup apertures a reduction occurs in the carrier-to-noise ratio, S_C/N_C, in Eq. [7] as the wavelength, λ_{min}, is reduced. Thus the trade-offs involved in the solution of Eq. [7] are the selection of D and f_{max}, and also an RPM that is high enough to meet the signal-to-noise ratio requirement but is as low as possible to obtain maximum stylus life.

The choice of the de-emphasis factor, D, affects not only the signal-to-noise ratio but also distortions of the luminance signal caused by pre-emphasis and clipping in the encoder.[4] In general, more spacing between the lower and upper breakpoint frequencies in the pre-emphasis network results in a large improvement in the signal-to-noise ratio. For given values of the pre-emphasis factor, D, and clipping levels, the weighted signal-to-noise ratio can be traded for low-frequency distortions by shifting the breakpoint frequencies up or down. Lowering the breakpoint frequencies improves the weighted signal-to-noise ratio by reducing low-frequency noise, but increases smear in the picture. Increased smear occurs because the lower-frequency components with relatively high signal amplitudes become more pre-emphasized and clipped causing a loss of rise time. Increasing the breakpoint frequencies reduces the smear but also reduces the weighted signal-to-noise ratio. Similarly, for a fixed low-frequency breakpoint and clipping levels, the weighted signal-to-noise ratio can be traded for high-frequency distortions by shifting the upper breakpoint frequency up or down. Raising the upper breakpoint frequency improves the weighted signal-to-noise ratio but reduces the contrast of small details in the picture, since the higher frequency components become more pre-emphasized and clipped causing a deterioration in rise time.

The choice of the de-emphasis factor, D, in Eq. [7] involves a compromise between weighted signal-to-noise ratio and both low-frequency and high-frequency distortions in the picture. In the RCA VideoDisc system, the luminance signal is pre-emphasized approximately 12 dB using a double breakpoint network with R-L time constants of 0.64 microseconds and 0.16 microseconds, respectively. In addition, a linear pre-emphasis of approximately 9 dB is used between 1.0 and 3.0 MHz. The luminance clipping levels are set at −66 IRE units in the direction of sync and +140 IRE units in the white direction. The calculated de-emphasis improvement factor, D, for these choices is 14.8 dB.

Fig. 1 shows the reduction in carrier-to-noise ratio, S_C/N_C, in Eq. [7] as the wavelength, λ_{min}, is reduced. The carrier-to-noise ratio was measured using a spectrum analyzer with a 30-kHz bandwidth. The data were taken from noncoated discs[9] recorded with an electromechanical recorder.[10] Note that the carrier-to-noise ratio is essentially flat from 1.5 to 1 μm and decreases about 6 dB per octave from 1 to 0.5 μm. The

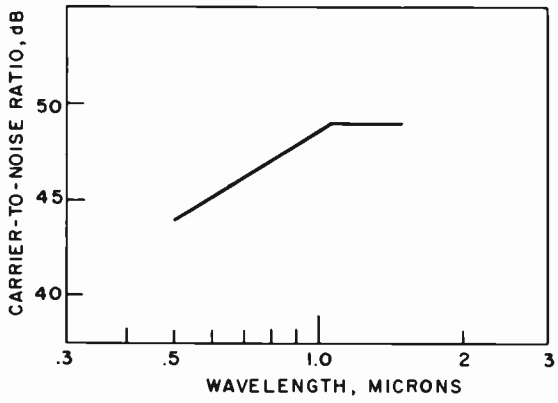


Fig. 1—Carrier-to-noise ratio versus wavelength at 2.9-inch radius for noncoated discs measured with a spectrum analyzer in a 30-kHz bandwidth.

data have been corrected to show the carrier-to-noise ratio that would be measured at the inside radius of a disc. The carrier-to-noise ratio measured at the outside radius of the disc would be approximately 3 dB higher due to the doubling of the stylus-to-disc velocity caused by the doubling of the radius, as can be seen from Eq. [3].

At a given radius, for both fixed playing time and recorded area, the carrier-to-noise ratio measured in a given bandwidth is independent of the RPM assuming the groove pitch is adjusted according to Eq. [5]. For example, assume that under these conditions the RPM of a disc is doubled. From Eq. [5] the groove pitch would be reduced by one-half. In order to play the narrower groove, the stylus width would have to be reduced by one-half. The decrease in the detected signal amplitude due to the narrower groove and stylus would be 6 dB. Since the noise is incoherent across the groove, the decrease in the noise under these same conditions is 3 dB. However, since RPM and thus the velocity have been doubled, disc noise per unit bandwidth is further reduced 3 dB resulting in no change in the carrier-to-noise ratio apart from aperture effects. While the carrier-to-noise has not changed, it can be seen from Eq. [4] that doubling the RPM will double the wavelength for a given frequency at a given radius. As shown in Fig. 1, if the wavelength change is in the range from 0.5 to 1.0 μm the carrier-to-noise ratio will increase due to the increased wavelength. One further note on this example is that doubling the RPM causes the stylus width to be reduced by one-half, which will reduce the life of the stylus as determined by the volumetric wear rate and make the stylus more susceptible to damage by disc imperfections.

From the previous example, it can be seen that the carrier-to-noise

ratio data shown in Fig. 1 can be used to make the trade-offs involved in the solution of Eq. [7] at various values of RPM, and that the desire to maximize stylus life will affect the choice of other system parameters. In general, the calculations involved in the solution of Eq. [7] are lengthy. However, if the simplifying assumption is made that the noise into the FM demodulator is white, then it is possible to make some relatively simple calculations. In general, these calculations have shown reasonably good agreement with measurements from noncoated discs.

The noise in the video signal from a disc is normally measured in the vertical blanking interval at black level. Using a black-level frequency of 5 MHz, an inside radius of 2.9 inches, a RPM of 450, and the proper conversion constants, Eq. (4) gives a wavelength of approximately 0.69 μm. From Fig. 1 the carrier-to-noise ratio measured in a 30-kHz bandwidth using a spectrum analyzer is approximately 46.2 dB. Table 1 summarizes the calculations leading to the unweighted luminance only peak-to-peak signal-to-rms noise ratio. Line 2 represents a spectrum analyzer correction factor to account for the fact that the noise bandwidth of the analyzer is wider than the 3-dB bandwidth and for an inherent error in log amplifiers when measuring noise. Line 5 is the FM improvement factor where $\Delta f = (f_{white} - f_{sync})/2 = (6.3 - 4.3)/2 = 1$ MHz. Line 7 represents the solution to Eq. [7]. Lines 8 and 9 convert line 7 to the ratio of peak-to-peak signal to rms noise as measured from blanking to white definition of signal-to-noise ratio. Line 10 represents the improvement in the measured signal-to-noise ratio due to combing, which is less than 3 dB due to the addition of the vertical detail signal to the combed luminance signal as previously discussed. Line 11 may appear to be a low value for the unweighted luminance signal-to-noise ratio, but it should be noted that the noise distribution of the RCA VideoDisc System is such that the CCIR weighting factor is higher than that for a system with this same value that has a flat noise spectrum.

Table 1—Unweighted Luminance Calculations

Line No.	*C/N* and *S/N* Calculations	Description
1	46.2 dB	*C/N* ratio in 30 kHz BW
2	−1.7 dB	Spectrum analyzer error
3	−23.0 dB	Conversion to 6 MHz BW
4	21.5 dB	10 log S_c/N_c Eq. [7]
5	−4.8 dB	10 log 3 $(\Delta f/f_m)^2$ Eq. [7]
6	14.8 dB	10 log $1/D$ Eq. [7]
7	31.5 dB	rms *S/N* ratio, 10 log *S/N* Eq. [7]
8	+9.0 dB	Conversion from RMS to p-p
9	−3.0 dB	Conversion to blank to white
10	+2.2 dB	Combing improvement
11	39.7 dB	Unweighted luminance only p-p signal to rms noise

However, before a value for the weighted signal-to-noise ratio at the output of the player can be obtained, it is necessary to consider the noise contribution from the chrominance signal.

It can be shown that the signal-to-noise power ratio, $(S/N)_{sc}$, for a subcarrier measured in a narrow bandwidth (i.e., 30 kHz) is given by[3]

$$(S/N)_{SC} = \frac{1}{2}\left(\frac{\Delta f}{f_{sc}}\right)^2 \frac{S_c}{N_c}, \qquad [8]$$

where S_c/N_c is the video carrier-to-noise power in a 30-kHz bandwidth, Δf is the peak chrominance deviation, and f_{sc} is the chrominance subcarrier frequency of 1.53 MHz. Table 2 shows the calculations for the unweighted chrominance signal-to-noise ratio. The first two lines are similar to the luminance calculations in that the carrier-to-noise ratio is obtained from Fig. 1 for a 0.69 μm wavelength and corrected for noise measurement errors in the spectrum analyzer. Line three is the video carrier-to-noise power, S_c/N_c, shown in Eq. [8]. Line 4 is the FM improvement factor. In the buried-subcarrier system the chrominance subcarrier is increased 2.7 times the NTSC value to improve the final signal-to-noise ratio. In the encoder, this enhanced chrominance signal is added to the pre-emphasized luminance signal to form the composite buried-subcarrier signal. The composite signal is then clipped at +144 IRE units and −70 IRE units to prevent over deviation of the FM modulator.[4] In order to calculate the FM improvement factor, a chrominance signal within this linear range will be considered (e.g., 50 IRE units peak-to-peak). Increasing this signal 2.7 times results in an enhanced chrominance of 135 IRE units, which corresponds to a peak-to-peak FM deviation of approximately 1.92 MHz. Substituting a peak deviation, Δf, of 0.96 MHz and a subcarrier frequency, f_{sc}, of 1.53 MHz into Eq. [8] results in an FM improvement factor of −7 dB for the 50 IRE peak-to-peak chrominance signal. However, the definition of signal-to-noise ratio is based on a measurement from blanking to white level,

Table 2—Unweighted Chrominance Calculations

Line No.	*C/N* and *S/N* Calculations	Description
1	46.2 dB	*C/N* ratio in 30 kHz BW
2	−1.7 dB	Spectrum analyzer error
3	44.5 dB	Actual video *C/N* in 30 kHz BW
4	−1.0 dB	10 log ½ $(\Delta f/f_{sc})^2$ Eq. [8]
5	43.5 dB	10 log $(S/N)_{sc}$ Eq. [8]
6	−15.2 dB	Conversion to 1 MHz BW
7	28.3 dB	rms chroma *S/N* ratio
8	+9.0 dB	Conversion from rms to p-p
9	+3.0 dB	Combing improvement
10	40.3 dB	Unweighted chroma only p-p signal to rms noise

which corresponds to 100 IRE units. Thus, a correction factor of 6 dB is added to this calculated value, which results in the −1 dB value shown in line 4 of Table 2. Line 5 represents the solution to Eq. [8]. The remaining conversion factors are similar to those explained for the luminance calculation and result in the unweighted chrominance-only signal-to-noise ratio shown in line 10.

The unweighted composite signal-to-noise ratio is obtained by adding the unweighted noise power of the luminance in Table 1, −39.7 dB, to the unweighted noise power of the chrominance in Table 2, −40.3 dB, resulting in a value of −36.9 dB. Applying the CCIR weighting function to the noise spectrum of the RCA VideoDisc results in a 9.1 dB improvement. Thus, the worse-case weighted signal-to-noise ratio measured at the inside of the disc would be 46 dB. Note that this is the value established by the specification and is equivalent to being 0.5 dB above the value where over 90% of the experienced viewers said the noise in the picture was not objectionable.

If the calculations shown in Tables 1 and 2 are repeated for various f_{max} and RPM while maintaining the black level fixed at 5 MHz, then the results shown in Fig. 2 are obtained. The values of RPM shown result in an even number of TV fields being recorded on the disc. It appears from Fig. 2 that both increased stylus life and the desired weighted sig-

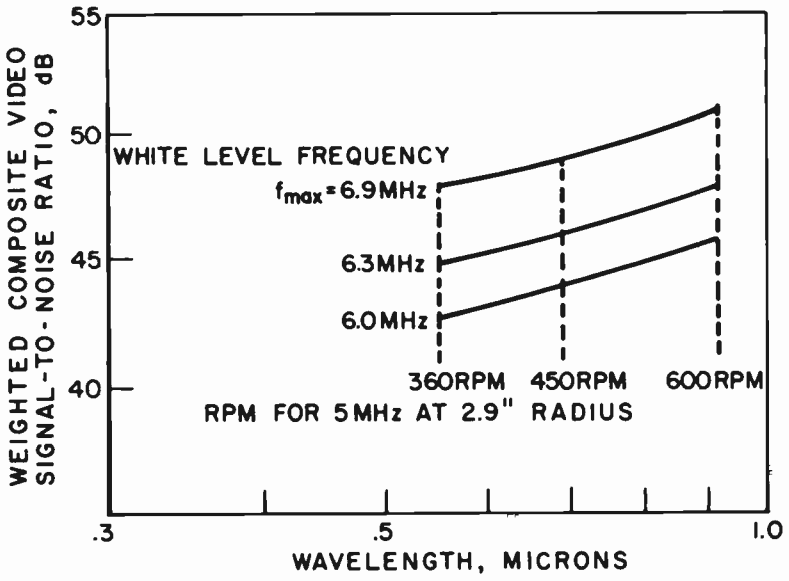


Fig. 2—Calculated weighted composite video signal-to-noise ratio at black level frequency versus wavelength at 2.9-inch radius for various white-level frequencies and RPM.

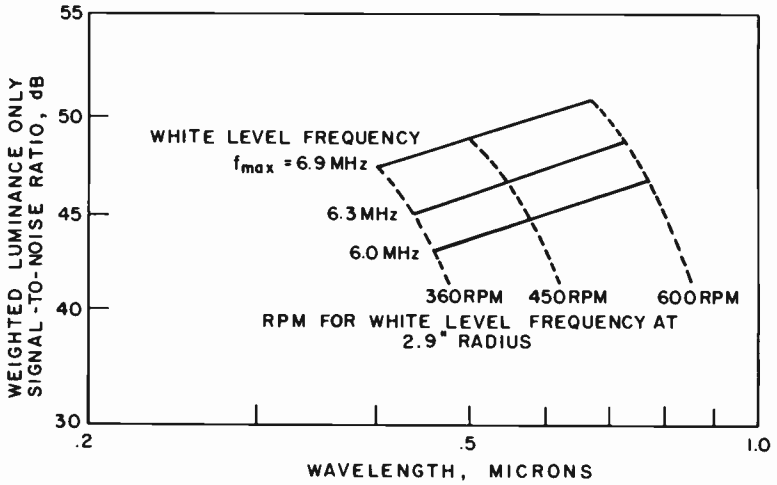


Fig. 3—Calculated weighted luminance-only signal-to-noise ratio at white-level frequencies versus wavelength at 2.9-inch radius for various RPM and white-level frequencies.

nal-to-noise ratio of 46 dB could be obtained by choosing the parameters of 360 RPM and a white level of approximately 6.9 MHz. This is somewhat misleading, however, since the wavelengths shown were calculated for the black-level frequency. Fig. 3 shows the effect of calculations made for the luminance-only weighted signal-to-noise ratio for wavelengths at the white-level frequencies. No specification exists for this ratio, but the calculations are useful in showing some of the tradeoffs involved in the choice of the system parameters. A luminance-only-weighted signal-to-noise ratio of 47 dB approximately corresponds to the previously specified composite value of 46 dB. The figure shows the reduction in minimum wavelength, λ_{min}, that would occur at 360 RPM in order to achieve a weighted luminance signal-to-noise ratio of 47 dB. Tests have shown that signal loss in a player increases and that disc manufacturing margins are reduced as the wavelength is reduced below approximately 0.5 μm, which makes the choice of 360 RPM undesirable.

Fig. 3 shows that a luminance-only weighted signal-to-noise ratio of 47 dB could also be obtained with a white-level frequency of 6.0 MHz and 600 RPM. While this would result in larger wavelengths, the increased number of grooves to obtain the same playing time as the lower RPM systems would result in reduced stylus width and a costly reduction in stylus life.

Fig. 3 also shows that with the present parameters of 450 RPM and a white level of 6.3 MHz, the weighted luminance-only signal-to-noise ratio is approximately equal to the 47 dB level. It may appear desirable

to increase the white level to 6.9 MHz in order to provide more margin; however, there are several considerations involved in such a decision. From Fig. 3 it can be seen that for 450 RPM, raising the white level to 6.9 MHz will decrease the wavelengths at the inner radius to approximately 0.5 μm which will increase the risk of signal loss in the player and reduce the disc manufacturing margins. In addition, wavelengths are further reduced due to pre-emphasis which extends approximately 0.6 MHz above white level. As stated previously, it is expected that the carrier-to-noise ratio shown in Fig. 1 will improve with production, so that the present parameters provide sufficient margin.

6. Conclusions

The carrier-to-noise ratio data shown in Fig. 1 can be used to make tradeoffs between f_{max}, λ_{min}, and RPM to obtain a desired value of the weighted video signal-to-noise ratio as shown in Figs. 2 and 3. For maximum stylus life, the RPM should be chosen as low as possible consistent with manufacturing margins and loss of signal performance in the player caused by wavelengths smaller than approximately 0.5 μm. It appears that the existing system parameters are compatible with the present state of the art in noncoated disc materials, recording, mastering, and manufacturing.

Acknowledgment

The authors would like to acknowledge the contributions of Hans Schwarz in the areas of video pre-emphasis and in establishing the equivalent visible signal-to-noise ratio of the RCA VideoDisc player.

References:

[1] J. K. Clemens, "Capacitive Pickup and the Buried Subcarrier Encoding System for the RCA VideoDisc," *RCA Review*, **39**, p. 33, March, 1978.

[2] E. O. Keizer and D. S. McCoy, "The Evolution of the RCA "SelectaVision" VideoDisc System—A Historical Perspective," *RCA Review*, **39**, p. 14, March, 1978.

[3] M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw Hill Book Co., New York (1959).

[4] D. H. Pritchard, J. K. Clemens, and M. D. Ross, "The Principles and Quality of the Buried Subcarrier Encoding and Decoding System," *RCA Review*, **42**, p. 367, Sept. 1981 (this issue).

[5] Report 410-2, "Single Value of the Signal-to-Noise Ratio for all Television Systems," CCIR XIII, Vol. XII, Plenary Assembly, 1974.

[6] D. N. Carson, "CATV Amplifiers; Figure of Merit and the Coefficient System," *IEEE International Convention Record*, Part 1, pp. 87–97, 1966.

[7] J. R. Cavanaugh, "A Single Weighting Characteristic for Random Noise in Monochrome and NTSC Color Television," *J. SMPTE*, **79**, p. 105, 1970.

[8] J. R. Cavanaugh and A. M. Lessman, "The Subjective Effect of Random Noise Spectra on 525 Line NTSC Color Television," *J. SMPTE*, **83**, p. 829, 1974.

[9] L. P. Fox, "The Conductive VideoDisc," *RCA Review*, **39**, p. 116, March 1978.

[10] E. O. Keizer, "VideoDisc Mastering," *RCA Review*, **39**, p. 60, March 1978.

Technical Standards for Direct Broadcast Satellite Systems

M. R. Freeling

RCA Americom, Princeton, NJ 08540

L. Schiff

RCA Laboratories, Princeton, NJ 08540

Abstract—This paper presents the technical considerations involved in the derivation of standards for direct broadcast satellite systems. In particular, it considers the crucial interaction between present and future hardware availability and system parameter determination. The objective of direct broadcast satellite design and the accompanying technical standards is the delivery to the home of the highest overall quality television picture — taking into account both clear weather reception and rain degraded reception. In the Eastern United States, system performance degradation due to rain is of particular importance. Anticipated improvements in receiver front-end noise temperature or satellite EIRP mandate wider channel bandwidth to maximize overall quality of reception, while currently available equipment restricts the design to a narrower bandwidth. As will be shown in a specific example, a 500°K LNA restricts the channel bandwidth to 16 MHz, while a design utilizing a noise temperature of 300°K results in an optimum bandwidth of 24 MHz. Increase in bandwidth to improve quality places increased importance on frequency reuse to conserve the spectrum. This paper will show how frequency reuse is more easily implemented by means of linear polarization than circular polarization, particularly in regions of heavy rainfall. System performance as a function of satellite orbital slot is also discussed.

1. Introduction

In recent years there has been increasing international interest in the transmission of television directly from a communications satellite to

the viewer's home. Direct broadcast satellite (DBS) experiments have been conducted or systems have been planned in Australia, Canada, China, France, Germany, Japan and the United States.

This paper describes technical standards that provide reception by the home earth station of television signals of high quality and availability. It also discusses the impact of state-of-the-art limitations on the freedom of the system designer to determine system parameters. It will be shown that a design utilizing a 300°K low-noise amplifier (LNA) results in an optimum bandwidth of 24 MHz. Utilization of the spectrum available for the direct broadcast service can be made more efficient by means of frequency reuse. This paper shows that frequency reuse is more easily implemented by means of linear polarization rather than circular polarization, particularly in regions of heavy rainfall. System performance as a function of orbital slot is discussed. The conclusion is reached that a satellite serving the eastern zone of the United States should be located between 100° West Longitude and 120° West Longitude. There should be similar angular separations between each of the other three zones and the satellite serving each zone.

2. Quality Standards

The overall objective of the direct broadcast satellite (DBS) system described in this paper is to provide excellent video and audio quality for a very high percentage of the year. Specifically, the system has been designed to provide a minimum video signal-to-noise ratio (peak-to-peak luminance video signal to rms weighted noise) of 42 dB everywhere in the United States and for all weather conditions, except for 8.8 hours a year (system availability = 99.90%). Subjective tests conducted by the Television Allocations Study Organization (TASO) have established that 98% of viewers will perceive a picture with $(S/N)_{\text{video}}$ of 42 dB as "fine" or better, and 50% of all viewers will rate this picture quality as "excellent." The DBS system will also provide a high quality program audio signal with a top baseband frequency of 15 kHz. The ratio of the audio signal at average program level (+8 dBm) to the unweighted rms noise, $(S/N)_{\text{audio}}$, will be a minimum of 53 dB. Subjective listening tests[1] have determined that for a system with $(S/N)_{\text{audio}}$ of 52 dB or greater, the noise is undetectable.

3. Physical Constraints

The freedom of the DBS system designer is limited by a number of physical constraints. Outstanding among these constraints are the

achievable noise temperature of the Ku-band LNA, the diameter of the antenna of the home earth station, and the power of the traveling-wave-tube-amplifier (TWTA) in the satellite.

An investigation of the present state-of-the-art of LNAs at 12.0 GHz indicates that the lowest practical noise temperature achievable either at present or in the near future is 300°K.[2] Economic and esthetic considerations dictate that the antenna dish mounted on the roof of the customer's residence be a maximum of 0.6 to 1.0 meter in diameter. The DBS system design described in this paper makes use of two antenna diameters: 0.6 meter and 1.0 meter. For direct broadcast reception everywhere throughout the geographic area of the United States, except in the worst-case rain degradation region, the home earth station would employ an antenna 0.6 meter in diameter and a 300°K LNA, resulting in a G/T of 9.5 dB/°K.

The highest saturated output power of Ku-band TWTAs that are either presently space qualified or under development for satellite applications is in the 250 to 450 watt range. Use of amplifiers in this power range on a DBS irradiating the entire continental United States would result in an effective isotropic radiated power (EIRP) of 51 dBW at the edge of the beam. Conservative system design and provision of adequate margin for fade due to rain both require that the downlink carrier-to-noise ratio, $(C/N)_D$, be at least 13 dB, or 3 dB above the threshold of the demodulator. For a satellite EIRP of 51 dBW, a home earth station G/T of 9.5 dB/°K, and a minimum $(C/N)_D$ value of 13 dB, the television channel bandwidth may be no greater than 10 MHz. This narrow

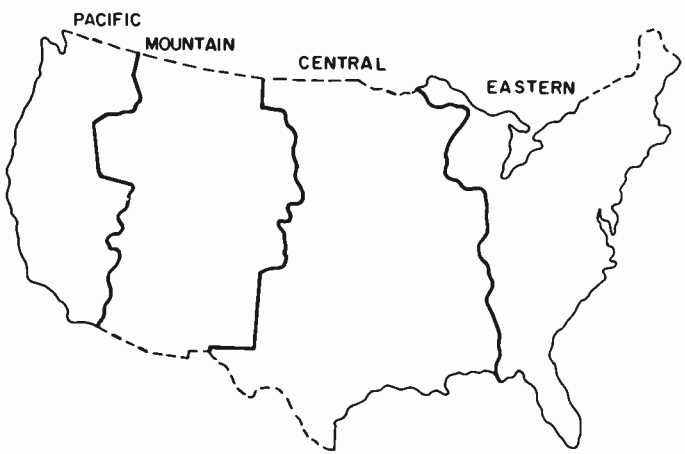


Fig. 1—Time zones in the United States.

bandwidth is completely inappropriate for the transmission of television. It can be concluded that with the existing technology or even with that projected for the near future, it is not possible for a single DBS to provide coverage of the entire land mass of the continental United States.

An appropriate coverage area that would be acceptable for commercial programming for direct broadcast service would be a single time zone. The four time zones into which the United States is divided are shown on the map in Fig. 1. A single DBS equipped with TWTAs with saturated output power in the 250 to 450 watt range and appropriate antenna sizing would provide a minimum EIRP of 57 dBW at the edge of a time zone. As will be shown below, an EIRP of 57 dBW will be adequate to achieve television reception satisfying the established standards of quality and availability.

4. Bandwidth

In the United States, particularly in the eastern and southeastern regions, attenuation due to rainfall has a great impact on satellite system performance in Ku band. Fig. 2 shows the regions of different rainfall rates into which the United States may be divided. The curves in Fig. 3 show the relative rain rates for each of these regions. For example, in region B, a rain rate of 5.6 mm/hr is exceeded 0.1% of the time, or 8.8 hours in a year. On the other hand, in the worst-case rainy southeastern area, region E, a rain rate of 35.4 mm/hr is exceeded for the same time interval. Fig. 4 shows the attenuation at 12.0 GHz as a function of rain rate for elevation angles at the receive earth station of 20°, 40°, and 60°.

In addition to increasing the uplink and downlink losses, rainfall raises

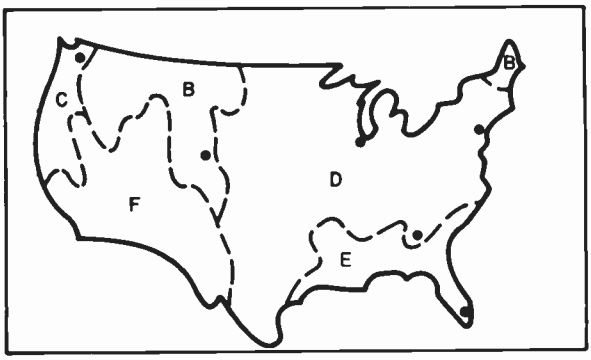


Fig. 2—Rain rate regions.

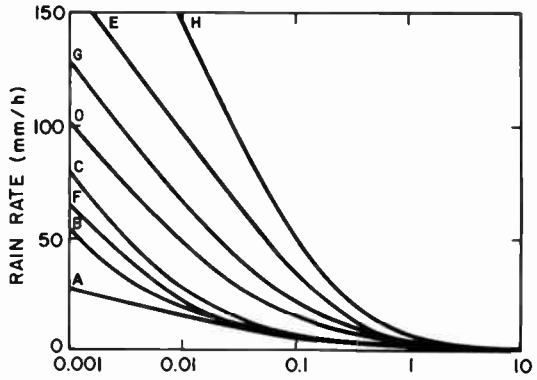


Fig. 3—Percent of time rain rate exceeded.

the noise temperature of the system, as the rain is at 290°K, substantially higher than the background temperature of space. The decrease in G/T of the receive earth station corresponding to the increase in noise temperature caused by rain is given by:

$$\Delta(G/T) = 10 \log \left\{ \frac{T_s + T_o([L_{rain} - 1]/L_{rain})}{T_s} \right\}$$

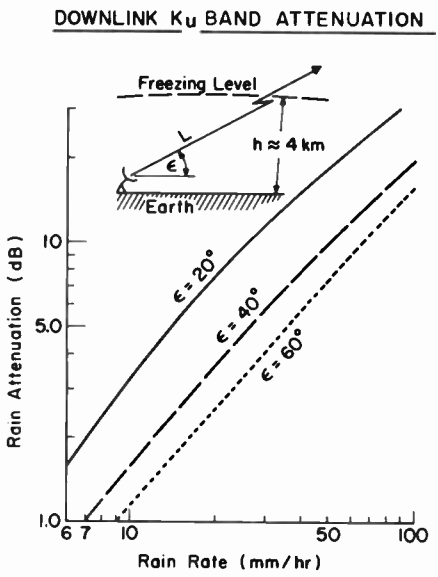


Fig. 4—Attenuation as a function of rain rate.

Table 1—Rain Degradation at 12.0 GHZ for Receiver Noise Temperatures of 300°K and 500°K and System Availability of 99.90% (Rain Rate Exceeded 0.1% of the Time)

City	Elevation Angle (°)	Rain Rate (mm/hr)	Attenuation (L_{rain}) (dB)	Δ (G/T) (dB) 300°K	500°K
Satellite at 135° W					
Atlanta	25.4	21.0	5.5	2.0	1.4
Chicago	23.0	15.0	4.5	1.8	1.2
Houston	34.5	35.0	7.5	2.2	1.5
Miami	24.0	35.4	10.0	2.3	1.7
New York	14.0	15.0	7.0	2.1	1.5
Seattle	33.5	8.0	1.5	0.9	0.6
Satellite at 95° W					
Atlanta	48.4	21.0	3.0	1.5	1.0
Chicago	41.3	15.0	2.3	1.2	0.8
Houston	52.4	35.0	5.5	2.0	1.4
Miami	55.5	35.4	5.3	1.9	1.3
New York	38.7	15.0	2.3	1.2	0.8
Seattle	30.0	8.0	1.6	1.0	0.6

where L_{rain} is the attenuation due to rain, T_s the system noise temperature, and T_o the ambient (rain) temperature = 290°K.

Table 1 shows, for selected cities in the United States, the rain rate for a system availability of 99.90%. The downlink attenuation and decrease in G/T corresponding to these rain rates are shown in the table for a receiver front-end noise temperature of 300°K.

Table 1 also shows the same parameters computed for an earth station equipped with a receiver with a noise temperature of 500°K. Rain degradations are calculated in the table for two satellite orbital slots, 95° West Longitude and 135° West Longitude.

For direct broadcast reception in the area experiencing the worst rain degradation, the southeast, the home earth stations would employ antennas 1.0 meter in diameter in conjunction with 300°K LNAs (G/T = 14.0 dB/°K). Fig. 5 shows bandwidth versus system carrier-noise ratio, $(C/N)_{system}$, at Miami, Florida. These curves were computed for a satellite located at 95° West Longitude and for saturated EIRP in Miami of 57.0 dBW. The curve for faded operation was computed for a rain rate that is exceeded 0.10% of the time. As may be seen from Fig. 5 for a bandwidth of 30 MHz, $(C/N)_{system}$ is below threshold for 8.8 hours a year, i.e., 30 MHz could be established as the channel bandwidth if operation at Miami were the limiting case. In the area of second-worst rain degradation, region D in Fig. 2, home earth stations would employ 0.6 meter antennas and 300°K LNAs (G/T = 9.5 dB/°K). Fig. 6 shows clear weather and faded performance for 99.90% availability in New York City for a satellite located at 95°W. This figure shows the upper channel bandwidth limit to be 26 MHz. Fig. 7 shows faded $(S/N)_{video}$ at New York

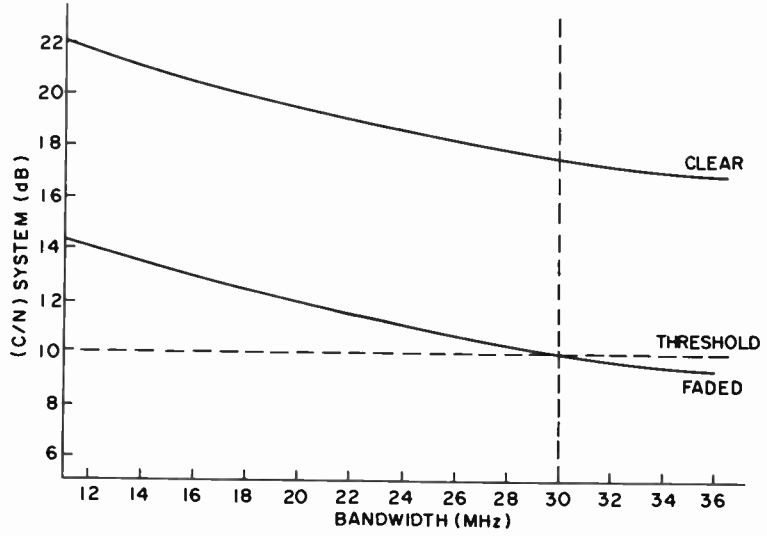


Fig. 5—Performance in Miami of earth station with 1.0 meter antenna.

City as a function of bandwidth. It can be seen that a minimum bandwidth of 22 MHz is required for the achievement of a faded $(S/N)_{\text{video}}$ of 42.0 dB. The above considerations have led to the conclusion that for a DBS system with a beam edge EIRP of 57.0 dBW, receiver noise

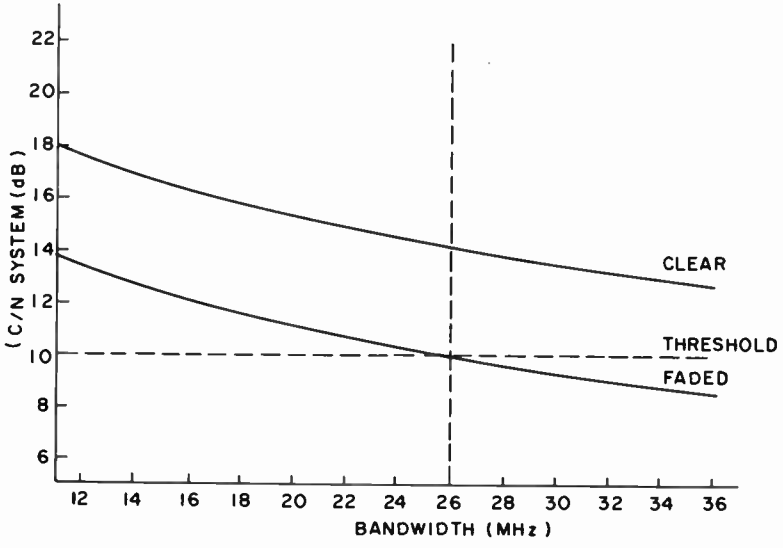


Fig. 6—Performance in New York City of earth station with 0.6 meter antenna.

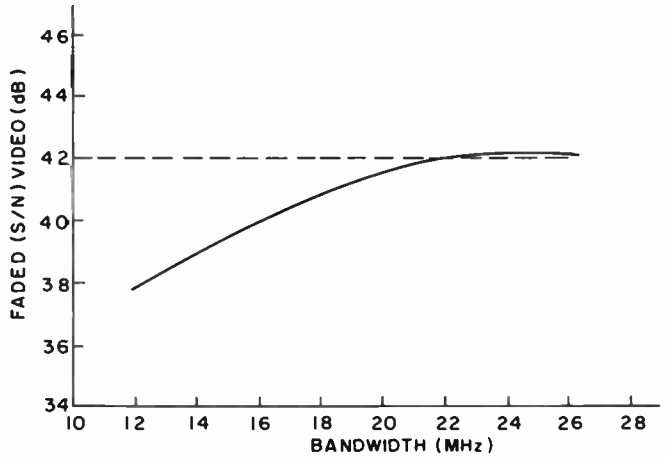


Fig. 7—Faded video signal-to-noise ratio in New York City.

temperature of 300°K, and receive antenna diameters of 0.6 to 1.0 meter, the optimum bandwidth is 24 MHz.

As indicated above, a study of the present state-of-the-art of LNAs at both 4.0 GHz and 12.0 GHz and a reasonable projection of development trends shows that a 300°K LNA will be achievable in the appropriate time frame for use in a direct broadcast system. Use of a higher noise temperature restricts the system to a narrower channel bandwidth. For example, Fig. 8 shows the performance in Miami of a DBS system

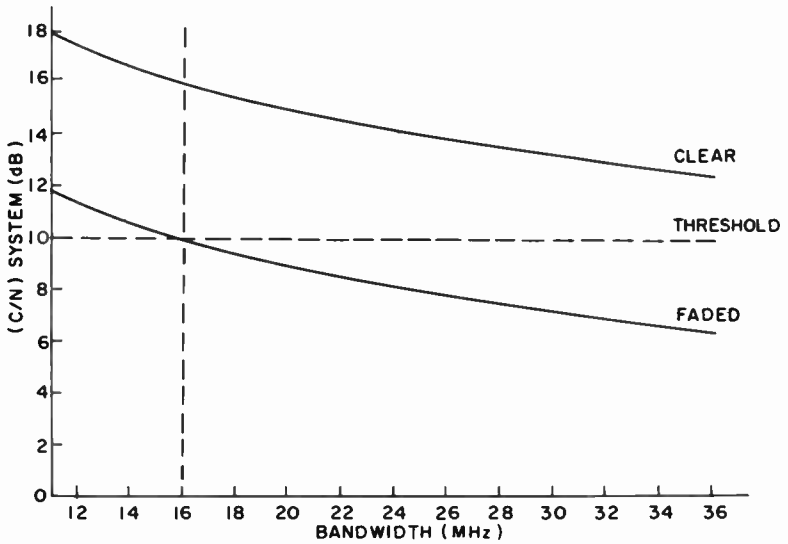


Fig. 8—Performance in Miami of an earth station with 0.75 meter antenna.

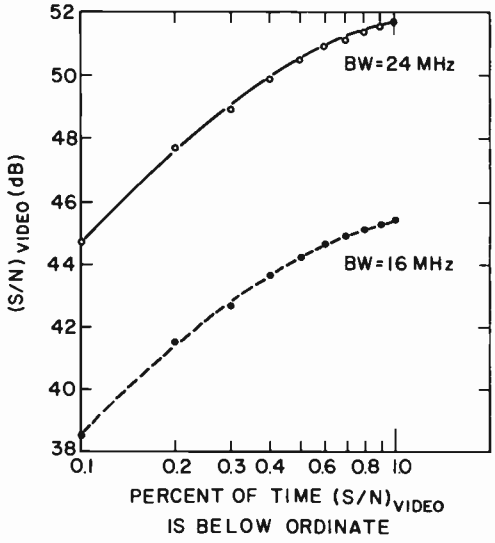


Fig. 9—Television reception at Miami (rain zone E).

(satellite location 95° West Longitude) with a receive antenna diameter of 0.75 meter and a 500°K LNA (G/T = 8.8 dB/°K). The curve for faded performance was calculated for a rain rate that results in a system availability of 99.83%; i.e., for a $(C/N)_{system}$ that is below threshold for 14.9 hours a year. Fig. 8 shows that choice of a noise temperature of 500°K restricts the channel bandwidth to a maximum of 16 MHz. For a bandwidth of 16 MHz, $(S/N)_{video}$ at threshold will be only 37 dB, a value that only 20% of TASO viewers rated as "excellent." The superiority of a 24-MHz bandwidth system over a 16-MHz bandwidth system can also be demonstrated for the case in which both systems employ the same receive earth station G/T. Fig. 9 shows the performance in Miami of systems with bandwidths of 16 MHz and 24 MHz. In both instances the home earth stations employ a 1.0-meter-diameter antenna and 300°K LNA (G/T = 14.0 dB/°K). $(S/N)_{video}$ for a given availability is 6 dB better for the 24 MHz bandwidth than for the 16 MHz bandwidth. Viewed differently, the system availability at a given level of $(S/N)_{video}$ is substantially greater for 24 MHz than for 16 MHz.

5. Polarization

The discussion above shows how important rain effects are in determining the quality of the received signal and hence in determining the channel bandwidths. The effect of rain is equally important in the choice

of polarization standards. The choice of a wide rather than narrow bandwidth is facilitated by improved equipment, but is motivated by a desire to provide a higher quality picture. At the same time, it obviously cuts down on the number of channels that can be carried in the band on a given polarization. Hence, frequency reuse via orthogonal polarization becomes very important. Further, one is driven to select the polarization scheme that provides the best isolation. One does not want to choose system standards (such as bandwidth) that provide higher quality and, at the same time, choose a polarization scheme with poorer isolation and, therefore, large interference.

The two possible orthogonal polarization modes are circular (left- and right-hand) and linear (vertical and horizontal). There is little question of the superiority of linear over circular polarization to achieve greater isolation. To demonstrate how important depolarizatin can be, consider a rain rate high enough to produce a 5-dB attenuation—a situation that occurs for more than 10 hours per year in much of the eastern United States. For earth stations having elevation angles in the 20°–40° range, such a rain will produce a 20–25 dB median depolarization with circular polarization.[3] In other words, at just the moment a 5-dB fade is being experienced, the carrier-to-interference (C/I) ratio (caused by an oppositely polarized signal) drops to 20–25 dB.

Fig. 10 demonstrates the superiority of linear polarization as compared to circular. The curve is taken from Chu[3] under assumptions of rain drop canting that understate this advantage. In fact, experiments have shown greater advantages at a specific site.[4] For vertical polarization (polarization angle = 0) the improvement over circular is 14 dB, making the C/I ratio in the previous example 34–39 dB. Polarization is vertical when the electric field vector lies in the plane formed by the local vertical and the line connecting satellite and earth station. Hence, if a transmission is vertical at one part of the served region it will not necessarily be vertical in another. The polarization angle is simply the angle between the electric

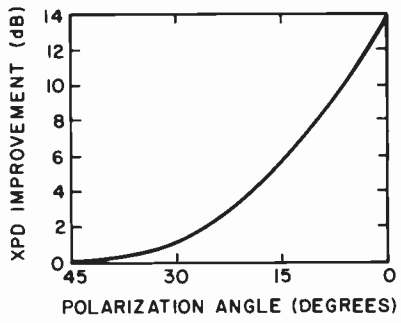


Fig. 10—Advantage of linear over circular polarization.

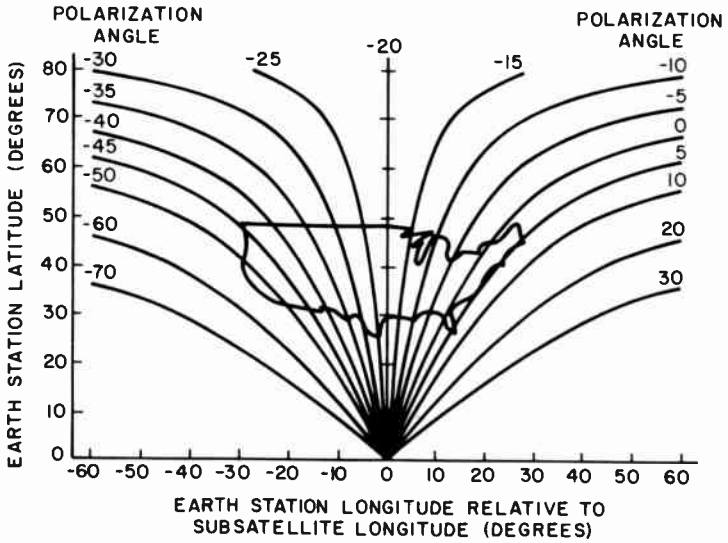


Fig. 11—Polarization angles across the United States for a satellite at 95°W.

field vector and the plane defined above. The situation for horizontal polarization is essentially the same.

Fig. 11 shows the variation in polarization angle through the coverage region when the satellite is slightly west of the coverage region (eastern United States). Note the feed on the satellite antenna is assumed to be

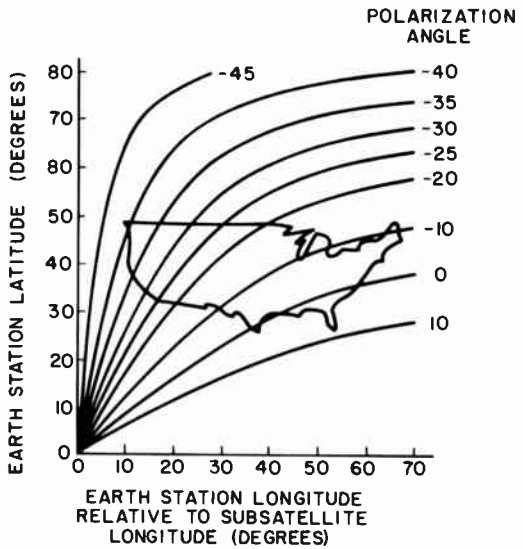


Fig. 12—Polarization angles across the United States for a satellite at 135°W.

rotated such that the polarization angle varies by only ±5° from true vertical over the entire region.

Fig. 12 shows that even if one locates the satellite far to the west of the service area, the variations from true vertical can be made ±10°.

The arguments favoring circular polarization involve earth station cost. Using linear polarization requires an antenna with a rotatable feed, as opposed to circular which does not. Further, that feed must be adjusted to different values in different parts of the coverage region.* In regions where high intensity rain is unlikely, these considerations can override the enhanced quality that linear polarization provides. It is felt that these extra costs will be relatively small, and the extra isolation provided in heavy rain regions is worth these costs.

6. Satellite Longitude

Another consideration is at what longitude the satellite should be located relative to the area served. There are certain factors that indicate that the satellite be located at the longitude of the served area, while others indicate that the satellite be positioned to the west of the area served. Again, the problem is most critical for the eastern United States. The list below gives a brief summary of these factors.

Factors That Indicate That the Satellite Should Be at the Longitude of the Served Area

(1) Elevation Angle
The closer the satellite is to overhead, the higher the elevation angle at the receiver. The higher the elevation angle, the fewer the potential number of obstructions. Experiments carried out in Japan[5] indicate that when the receiving antenna view is half shielded by a building, the attenuation is about 8 dB. When the antenna is obstructed by a dense foliage tree, attenuation is sometimes greater than 10 dB, and attenuation is 7 to 8 dB even for low foliage trees.

(2) Rain Degradation
The higher the elevation angle, the shorter the path through a potential rainy atmosphere and the lower the outage time at any given level of fade.

* The difficulties in doing this should not be overrated. The feed can be adjusted by the local installer prior to installation or at the customer location by nulling on the orthogonally polarized signal. Note also that, on installation, accurate alignment of elevation and azimuth must be done anyway.

(3) Polarization

Figs. 11 and 12 show that for satellites using linear polarization and located far to the west of the served region, the polarization angle variation is larger. Because of the particular shape of the eastern United States, it is actually easier to keep the polarization closer to true vertical (or horizontal) when the satellite is slightly to the west rather than directly overhead.

Factors That Indicate That the Satellite Should Be Well to the West of the Served Region

(4) EIRP

As one moves the satellite farther to the west of the served region, the solid angle subtended by the served region diminishes. If one uses larger antennas on the satellite, one can therefore obtain higher EIRP while still serving the whole region.

(5) Time of Eclipse

The satellite is in eclipse for a small number of days at the time of the equinox. Because of the inability to power the transponders through an eclipse, transmissions cease during this period. The duration of the eclipse as a function of the number of days into the equinox is shown in Fig. 13. Of course, the middle of the eclipse period each day is at astronomical midnight at the longitude of the satellite. Hence, if the satellite is directly overhead, eclipse will occur during the very late part of prime time for viewer watching. If the satellite is far to the west of the region served, the eclipse oc-

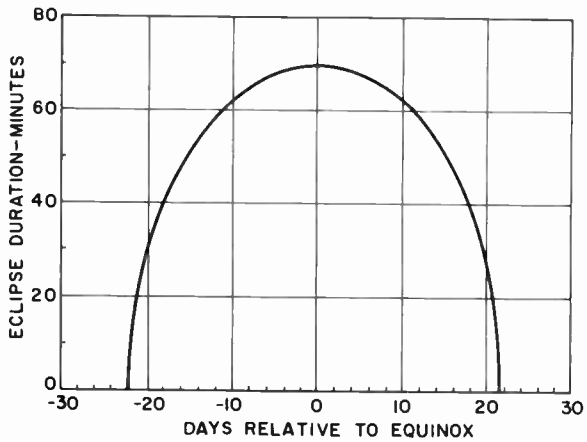


Fig. 13—Duration of eclipse versus days relative to equinox.

curs during the hours well after midnight in the region served when there is less impact to viewers by cessation of transmission.

7. Discussion

Factor (3) is really a very small matter. Even a satellite located at a longitude off the coast of California (Fig. 12) can orient the feed to provide less than 10° variation from true vertical (or horizontal) polarization, and a 10° polarization angle will still provide an 8–10 dB gain over circular polarization.

Factor (1) is extremely important, but it is almost impossible to quantify, since one is concerned with the minimum elevation angle (which presents no physical obstruction) and the region of concern is one of immense topographical variation. Factor (5) is also important, but it is an operational rather than a technical question.

Factors (2) and (4), can be considered together. Since considerations of Factor (1) are neglected, a far westerly satellite location will deliver a higher carrier-to-noise ratio (C/N) for a given raw satellite power and given receiver than will a more easterly location in clear weather. However, the westerly location will have a higher degradation in rain than will the easterly. The question is which location will produce the higher C/N for rain at a given percentage of the time.

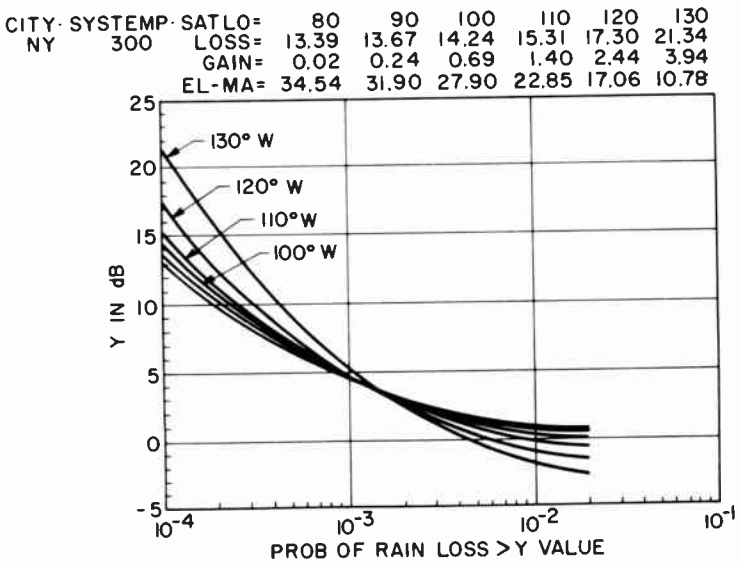


Fig. 14—Degradation of carrier-to-noise ratio at New York relative to clear weather performance.

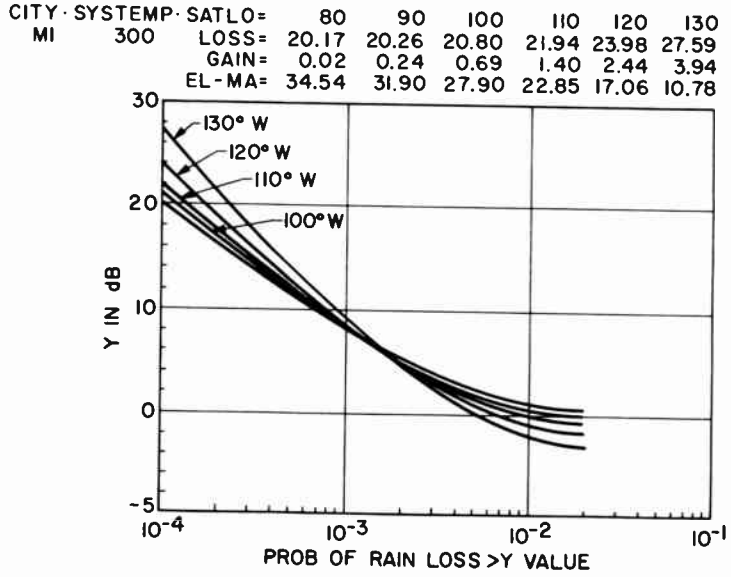


Fig. 15—Degradation of carrier-to-noise ratio at Miami relative to clear weather performance.

Figs. 14 and 15 give the answer to this question. Curves are plotted of *C/N* degradation relative to clear weather conditions for a satellite at 75°W. Fig. 14 is for New York City and Fig. 15 is for Miami, with the different curves corresponding to satellite positions from 80° to 130° in 10° increments. It will be noted that the asymptotic values (lower right, corresponding to clear weather) for more westerly satellites are negative (i.e., negative loss), showing the higher EIRP effects because of the smaller solid angle subtended. Both curves are for a 300° LNA assumed, but the results for 500° differ only slightly. The data up at the tops of the figures show the loss at the 10^{-4} level and the gain in the EIRP for each satellite longitude, allowing one to identify which curves belong to which longitudes. In addition, the elevation angles for the poorest location (northeastern Maine) are also given for each satellite.

At the 10^{-3} (or 0.1%) level, the curves show that at both New York and Miami there is the same degradation for locations between 80° and 120°, while a 130° location gives slightly poorer performance. Of course, for larger percentages of the time (10^{-2} or 10^{-1}) there is significant gain for more westerly locations when viewed from either city.

In attempting to sum up the important technical Factors (1), (2), and (4) (Factor (3) is small in effect and Factor (5) is operational), we believe that satellite longitudes significantly to the west of the served area are to be preferred, up to an orbital slot at 120° West Longitude.

Going west of 120° is undesirable, not only because the performance at the 10^{-3} level gets worse but also for elevation angle reasons. Note that in Maine the elevation goes from 17° to 10.7° as the satellite goes from 120° to 130°. A location of the satellite somewhere between longitude 100° and longitude 120° gives good elevation angles, clear weather C/N performance improvement, and excellent eclipse timing. This band is the preferred set of locations.

8. Conclusion

To provide a DBS capability for the United States, the present capability of power tubes (TWTAs) requires the separation of the United States into four distinct regions. Of these regions, the eastern time zone presents the greatest difficulty. Studying the method of providing service to the eastern time zone led to a number of tentative conclusions, which are restated below.

- The channel bandwidth should be 24 MHz. Higher bandwidth reduces the total number of channels and presents threshold problems. Lower bandwidth results in less SNR than we believe the consumer will accept.
- Polarization should be linear and oriented so that the heaviest rain regions get true vertical and horizontal. We believe the cost penalty of linear over circular is slight and the superior isolation provided by linear is well worth the price because of the heavy rain.
- The longitude of the satellite serving the eastern region should be between 100° and 120° West Longitude. Similar angular separation between the region served and the satellite serving it should hold for the other three regions.

References:

[1] B. Blesser, F. Ives, "A Reexamination of the S/N Question for Systems with Time-Varying Gain or Frequency Response," *J. Audio Engineering Soc.*, **20**, No. 8, p. 638, Oct. 1972

[2] J. G. Chambers, "An Evolutionary Approach to the Introduction of Direct Broadcast Satellite Service," *NTC '80 Conf. Record*, Paper 73.2.2

[3] T. S. Chu, "Microwave Depolarization of Earth-Space Path," *BSTJ*, **59**, No. 6, p. 987.

[4] H. W. Arnold et al., "Measurements and Predictions of the Polarization-Dependent Properties of Rain and Ice Depolarization," *NTC '80 Conf. Record*, Paper 43.3.1

[5] R. Takahashi, "Planning and Experimentation for an Operational Broadcasting Satellite for Japan," *NTC '80 Conf. Record*, Paper 73.4.4

GIMOS—A Nonvolatile MOS Memory Transistor

S. T. Hsu

RCA Laboratories, Princeton, NJ 08540

Abstract—Characteristics of electrically alterable stacking gate nonvolatile MOS memory transistors are presented. The fabrication processes for this device is compatible with most MOS technologies. The memory mechanism is charge storage at the floating gate. The charges stored at the floating gate are injected into the floating gate from the control gate by the Fowler Nordheim tunneling process. This memory device is shown to have a very large memory window, extremely long retention time, and excellent endurance and temperature stability.

Transistors have been fabricated in SOS technology. Threshold voltages for a self-aligned GIMOS p-channel device can be varied from +15 V to −25 V. Similar characteristics were obtained for n-channel transistors. Based on limited data the retention time is at least 1,000 years at room temperature. Transistors were stable after 20,000 write-erase cycles and 3×10^9 read cycles. Other stability data are also presented.

1. Introduction

In recent years a number of electrically alterable nonvolatile memory devices, other than MNOS structures, have been developed.[1–4] These electrically alterable nonvolatile MOS memory devices may require an extra electrode or complicated fabrication process compared to the standard MOS devices. The increased complexity limits the application of such devices as a memory cell for use in large scale integrated circuits.

In 1975 DiMaria and Kerr[5] found that for a given field intensity and polarity the dark current flow through a polycrystalline silicon MOS structure is many orders of magnitude larger than that flowing through

a single-crystal silicon MOS structure. They also showed that when polycrystalline silicon or a single-crystal silicon MOS structure was oxidized for various lengths of time, the dark current showed an abrupt conductivity decrease when the polycrystalline silicon became completely oxidized, leaving only oxide on the single-crystal silicon substrate. The transport mechanism for the dark current in a polycrystalline silicon MOS structure is the Fowler-Nordheim tunneling process. After taking the Schottky barrier lowering effect into consideration, DiMaria and Kerr found the activation energies were equal to 1.4 eV and 1.8 eV when the aluminum electrode was biased with positive and negative voltages, respectively. They believed that 1.4 eV was the high-field electron energy barrier height, i.e., the energy difference between the edge of the conduction bands of polycrystalline silicon and the thermal oxide at high fields. The 1.8 eV activation energy was believed to be the barrier height between the Fermi energy of aluminum and the edge of the conduction band of the SiO_2. They concluded that this small energy barrier was due to the local field enhancement near the injection electrode. The energy barriers of polycrystalline silicon MOS structures determined by the photocurrent method are identical to those of single-crystal silicon MOS structures.

The properties of the Si-SiO_2 interface has been studied very extensively during the past fifteen years. Studies carried out using scattering and channeling measurement,[6] low energy ion-scattering spectrometry,[7] and Auger analysis[8–10] concluded that there was a silicon-rich transit region between Si and SiO_2. Based on this finding Johannessen et al[10] proposed that the energy diagram at the Si-SiO_2 interface was not an abrupt function but is a smooth transit region between the Si and SiO_2. These experimental results provide an alternative explanation to DiMaria and Kerr's data. At a high electric field, the energy barrier for the current carriers is determined by the transition region between silicon, or polycrystalline silicon, and the SiO_2. The transition region of polycrystalline silicon–SiO_2 is presumably larger than that of single-crystal silicon SiO_2. As a result, at high electric field the energy barrier height is smaller for a thermal oxide grown on polycrystalline silicon than for thermal oxide on single crystal silicon. The 1.4 eV is the energy barrier between the edge of the conduction band of the polycrystalline silicon and the edge of the conduction band of SiO_2 when a large positive voltage is applied to the aluminum electrode. When a large negative voltage is applied to the aluminum electrode the energy difference between the top of the valence band of polycrystalline silicon and the tops of the valance band of SiO_2 is equal to 1.8 eV. The energy barrier obtained from the photocurrent (low field) is determined by the bulk property of SiO_2. Consequently when the aluminum electrode of an Al-SiO_2–polycrys-

talline silicon structure is biased with a large positive voltage, electrons are injected from the polycrystalline silicon to the aluminum. When the aluminum electrode is biased with a large negative voltage, holes are injected from the polycrystalline silicon to the aluminum.

2. Device Structure and Memory Mechanism

Based on the results discussed above, a simple stacking gate nonvolatile electrically alterable memory MOS transistor was constructed. The basic structure of this device is sketched in Fig. 1. This structure is similar to that proposed by Lee.[4] However, there are some basic differences. The

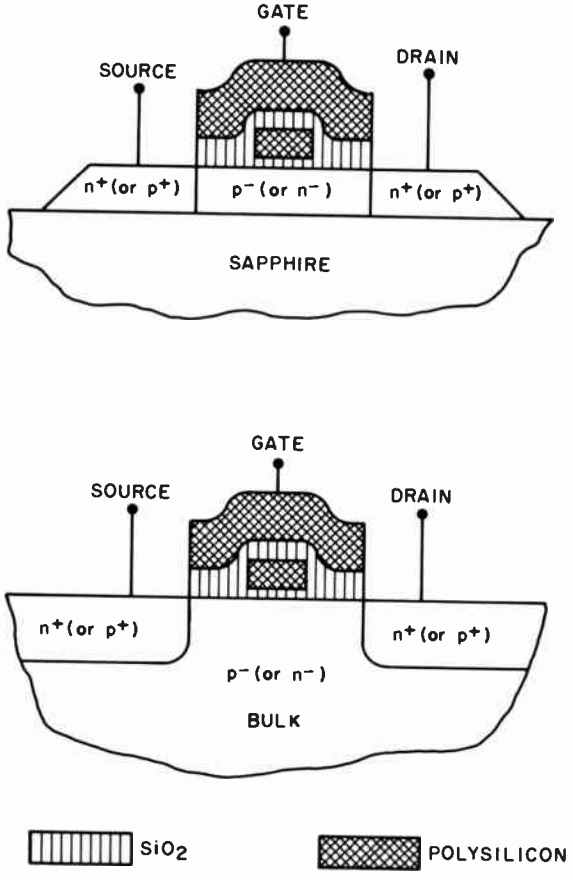


Fig. 1—Structures of GIMOS memory transistors (a) on SOS substrate (b) on bulk substrate.

sketches shown in Figs. 1(a) and 1(b) are the device on SOS, and on bulk Si, respectively. The floating gate covers only the central portion of the channel. The device can be operated at high voltages, since avalanche breakdown at the drain junction cannot disturb the potential at the floating gate. A large memory voltage window can be obtained by using a thin oxide between the substrate and the floating gate and a thick oxide between the floating gate and the control gate. However, the oxide between the substrate and the floating gate must be sufficiently thick that direct tunneling of electrons or holes between the substrate and the floating gate is not possible. Direct tunneling between any electrode and the floating gate will degrade the memory properties of the device. For the sake of process simplicity, the sum of the thickness of the two layers of SiO_2 should be equal to the thickness of the gate oxide of the non-memory MOSFET's. The thickness of the polycrystalline silicon floating gate may be thinner than 1000 Å. The doping concentration of the floating gate polycrystalline silicon is not critical. For simplicity, undoped polycrystalline silicon was used for the floating gate. The control gate can be p+ or n+ polycrystalline silicon or any convenient metal. This device will be referred to as a GIMOS (Gate Injection MOS) transistor.

The energy diagram of a GIMOS transistor is sketched in Fig. 2. The solid lines sketch the ideal energy diagram and the dotted lines sketch the actual energy diagram. Fig. 2(a) is the energy diagram when the system is at the flat band condition, while Fig. 2(b) is the situation when a large positive voltage is applied to the control gate. In the latter case, electrons may be injected into the floating gate from the substrate and holes from the control gate to the floating gate. The electrons at the floating gate see a barrier height of 1.4 eV while the holes see a barrier height of 3.6 eV. Therefore, a large fraction of electrons flow to the control gate and most of the holes injected to the floating gate are trapped at the floating gate. As a result the floating gate is positively charged.

When the control gate voltage is reversed, holes injected into the floating gate see a 1.8 eV barrier, while electrons see a barrier height of 3.2 eV. The floating gate is, therefore, negatively charged. The gate injection mechanism thus allows the sign of the charge stored on the floating gate to be electrically altered in a non-distructive manner. Since the barrier height for holes is larger than that for electrons, we expect the floating gate can store more holes than electrons. Thus the memory voltage window of a p-channel GIMOS is expected to be larger than that of an n-channel GIMOS.

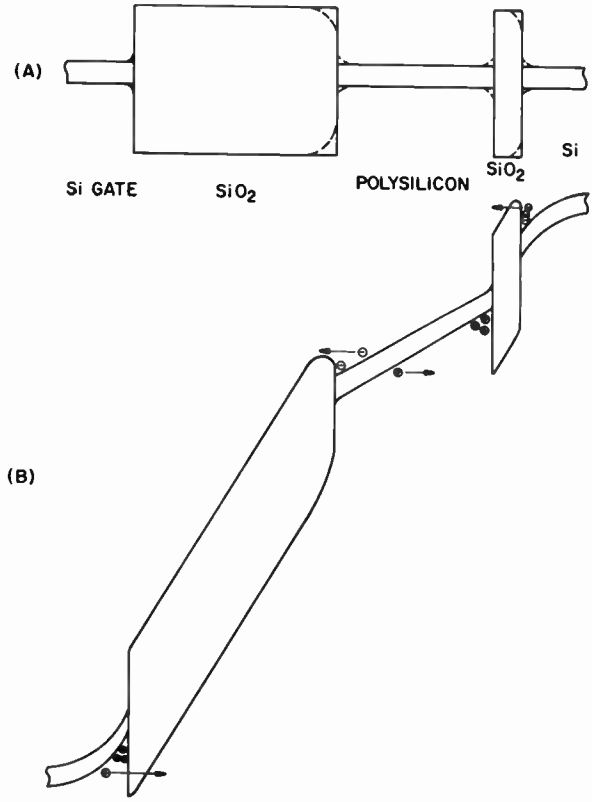


Fig. 2—Sketches of energy diagrams of GIMOS memory transistors: (a) at flat band condition and (b) the control gate is buried with a large positive voltage.

3. Device Analysis

For the purpose of analysis of the structure let the thickness of oxide 1 (between the substrate and the floating gate) and oxide 2 (between the floating gate and the control gate) be t_{ox1} and t_{ox2} respectively, let the charge density at the floating gate be Q_f and let Q_n and Q_B be the mobile and the fixed substrate charges respectively. It can then be shown that

$$V_g = V_{FE} + 2\phi_F + \frac{Q_n + Q_B}{C_g} + \frac{Q_f}{C_2}, \qquad [1]$$

where $C_g = \epsilon/(t_{ox1} + t_{ox2})$, C_2 is the capacitance of oxide 2 between the floating gate and the control gate, and V_{FB} and ϕ_F are the flat band voltage and the Fermi energy of the substrate material, respectively. The

turn-on voltage, V_T, is defined as the gate voltage at the onset of the conducting channel formation:

$$V_T \equiv V_G \ (Q_n = 0)$$

$$V_T = V_{FB} + 2\phi_F + \frac{Q_B}{C_g} + \frac{Q_f}{C_2} \qquad [2]$$

or

$$Q_f = C_2 \ (V_T - V_{To}) \qquad [3]$$

where V_{To} is the virgin threshold voltage. The density of charge stored at the floating gate can be obtained from Eq. [3].

4. Experimental Results

GIMOS nonvolatile memory transistors of the structure shown in Fig. 1 were fabricated on SOS substrates with a complementary polycrystalline silicon gate MOS process. Both p^+- and n^+-polycrystalline silicon control gate devices have been made. The type of doping and the doping concentration of the control gate polycrystalline silicon was found to have no significant effect on the memory characteristic of the device.

The memory characteristics of an n-channel GIMOS is shown in Fig. 3. The source-drain spacing of the device is 20 μm. The width of the

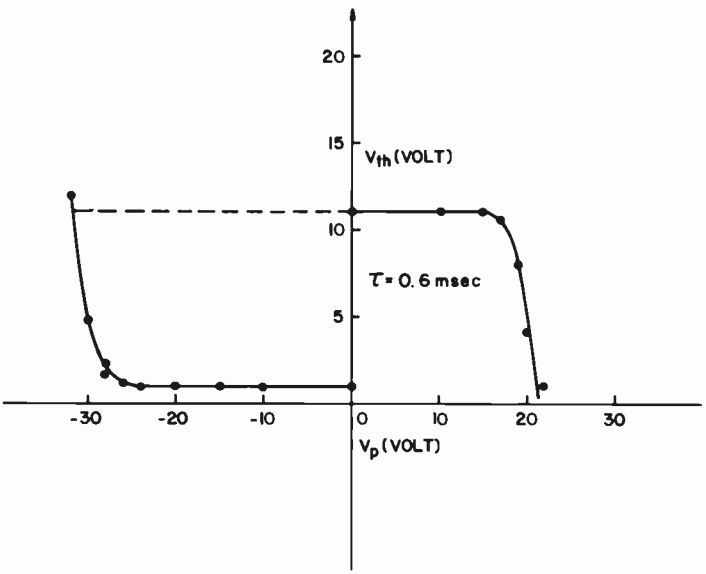


Fig. 3—Memory characteristics of a typical n-channel GIMOS memory transistor.

floating gate is equal to 10 μm and the width of the device is 300 μm. The thickness of oxide 1 and oxide 2 are equal to 145 Å and 740 Å, respectively. In Fig. 3 the threshold voltage of an n-channel GIMOS transistor is plotted as a function of pulse amplitude applied to the gate of device as to write or erase data. The pulse width used was 0.6 msec. After each pulse was applied the threshold voltage was readjusted to 1 V and 11 V for a negative and a positive pulse, respectively. Therefore the threshold voltage measured after each pulse corresponds to the amount of charge injected into the floating gate by the given pulse. For the sake of measurement simplicity the threshold voltage was defined as the gate voltage when the drain current at $V_D = 2$ V was equal to 5 μA. The maximum stable threshold voltage was found to be approximately equal to 12 V. Therefore, the maximum number of electrons that can be stored at the floating gate is approximately equal to 3×10^{12} cm^{-2}.

The memory characteristics of a p-channel GIMOS transistor located on the same chip of the device used in Fig. 3 is plotted in Fig. 4. The structure of the p-channel GIMOS is the same as that of n-channel GIMOS except the width of the device is equal to 425 μm. The pulse width used in this measurement was 1 msec. After each pulse the threshold voltage was readjusted to -1.5 V and -20 V for positive and negative pulses, respectively. The maximum stable threshold voltage is approximately equal to 23 V. The maximum number of holes that can be stored at the floating gate of this device is thus approximately equal to 7×10^{12} cm^{-2}.

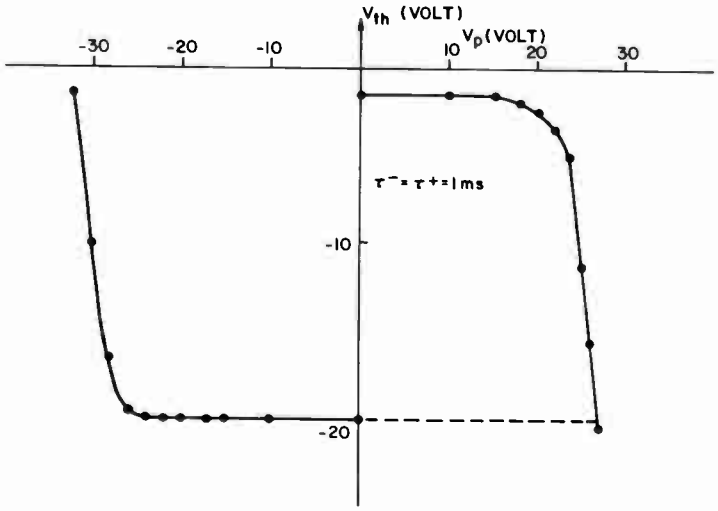


Fig. 4—Memory characteristics of a typical p-channel GIMOS memory transistor.

Self aligned GIMOS transistors were also made on SOS substrates. In this structure the source, the drain, and the floating gate are self-aligned to the control gate electrode. The memory characteristic of a self-aligned p-channel GIMOS is shown in Fig. 5. The pulse width in this measurement was 10 msec. It was shown in this figure that the threshold voltage can be changed from +15 to −25 V. A similar characteristic was also obtained from a self-aligned n-channel GIMOS. As predicted it is consistantly observed that the maximum negative threshold voltage is larger than the maximum positive threshold voltage.

A GIMOS nonvolatile memory transistor can be operated at temperatures below −55°C to above 200°C. We wrote p- and n-channel GIMOS transistors to threshold voltages of −11 V and +11 V, respectively, and stored these devices in a 150°C ambient with all electrodes shorted together for more than four years. No threshold voltage shift could be measured. The threshold voltage of a p-channel GIMOS transistor was written to −10 V and stored in 150°C ambient with −10 V applied to the gate with respect to the source and drain for four years. Again there was no threshold voltage shift. A p-channel GIMOS transistor was written and erased 20,000 times, and the memory window measured was never less than 20 V. The virgin threshold voltage of this device was equal to −1.4 V. The threshold voltage of this device was then

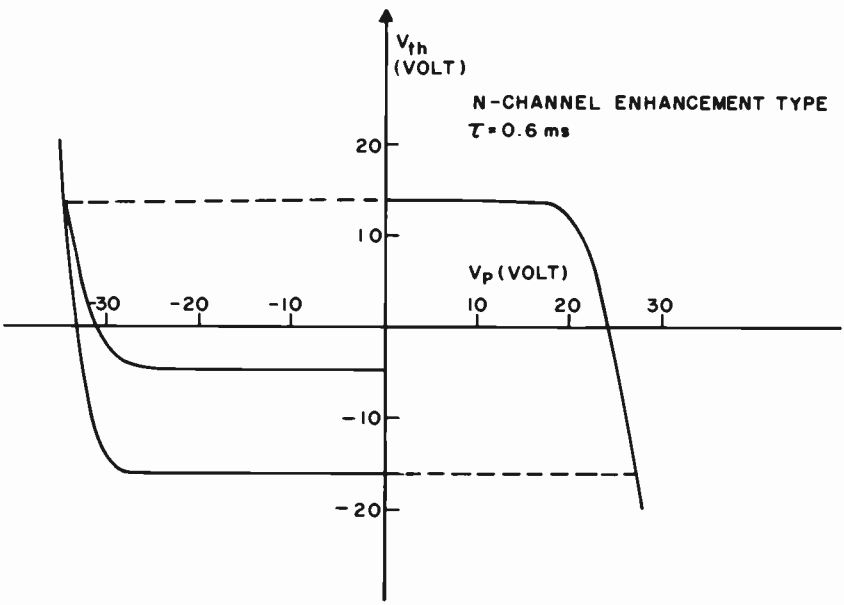


Fig. 5—Memory characteristic of a typical self-aligner GIMOS memory transistor.

adjusted to −19.6 V and stored at room temperature for 40 months. The threshold voltage was found to be equal to −18 V. If we plot the threshold voltage as a function of the logarithm of time and extrapolate the result, the extrapolated line reaches −1.4 V at more than 2000 years. In another experiment, the threshold voltage of an n-channel GIMOS transistor was written to 10 V. Positive pulses of 6 V amplitude and 0.6 msec width simulating a read condition were applied to the gate with respect to the source and the drain. No threshold voltage change could be measured after 3×10^9 pulses were applied.

5. Conclusion

Electrically alterable stacking gate nonvolatile MOS memory transistors have been made and tested. The charge stored at the floating gate are injected from the control gate by Fowler Norheim tunneling process. Therefore the memory characteristics of this type of devices are not sensitive to operating temperature. Memory voltage windows of −20 V and +10 V for p-channel and n-channel transistors, respectively, can be easily obtained. The device has extremely long retention time and excellent endurance.

Acknowledgment

The author wishes to express his appreciation to M. Jones and J. M. Cartwright for their help in device fabrications and some electrical measurement. He is also indebted to N. Goldsmith and J. H. Scott, Jr. for helpful discussions.

References:

[1] Y. Tarui, Y. Hayashi, and K. Nagai, "Electrically Reprogrammable Non-Volatile Semiconductor Memory," p. 52, *Digest of Technical Papers, 1972 IEEE International Solid State Circuits Conf.*, Philadelphia, PA.

[2] R. G. Miller, H. Nietsch, B. Rossler, and E. Walter, "Electrically Alterable 8192 Bit n-channel MOS PROM," p. 188, *Digest of Technical Papers, 1977 International Solid State Circuits Conf.*, Philadelphia, PA.

[3] F. Berenga, V. Daniele, G. Ferla, and G. Torelli, "E^2-PROM TV Synthesizer," p. 196, *Digest of Technical Papers, 1978 International Solid State Circuits Conf.*, San Francisco, CA.

[4] H. S. Lee, "A New Approach for the Floating-Gate MOS Non-Volatile Memoroy," p. 475, *Appl. Phys. Let.*, **31**, Oct. 1977.

[5] D. J. DiMaria and D. R. Kerr, "Interface Effects and High Conductivity in Oxides Grown from Polycrystalline Silicon," p. 505, *Appl. Phys. Let.*, **27**, Nov. 1975.

[6] T. W. Sigmon, W. K. Chu, E. Lugujjo, and J. W. Mayer," Stoichiometry of Thin Silicon Oxide Layers on Silicon," p. 105, *Appl. Phys. Let.*, **24**, Feb. 1974.

[7] W. L. Harrington, R. E. Honig, A. M. Goodman, and R. Williams, "Low-Energy Ion-Scattering Spectrometry (ISS) of the SiO_2/Si Interface," p. 644, *Appl. Phys. Let.*, **27,** Dec. 1975.

[8] J. S. Johannessen and W. E. Speicer, "Auger Depth Profiling of Interfaces in MOS and MNOS Structures," p. 849, *J. Vac. Sci. Technol.*, **13,** July/Aug. 1976.

[9] J. S. Johnannessen, and W. E. Speicer," An Auger Analysis of the SiO_2-Si Interface," *J. Appl. Phys.*, **47,** July 1976.

[10] J. S. Johaannessen, C. R. Helms, W. E. Speicer, and Y.E. Strausser, "Auger Depth Profiling of MNOS Structures by Ion Sputtering," p. 547, *IEEE Trans. Elec. Devices*, **ED-24,** May 1977.

Observations of Electron and Hole Transport Through Thin SiO_2 Films

S. T. Hsu

RCA Laboratories, Princeton, NJ 08540

Abstract—The charge transport phenomenon in stacked-gate MOS memory transistors has been investigated. When the oxide field intensity is smaller than the threshold field intensity of gate injection, and the control gate is biased with a negative (or positive) voltage, holes (or electrons) are found to be injected from the substrate to the floating gate. When the oxide field intensity is larger than the threshold field, the gate injection of current carriers becomes possible. Hence electrons (or holes) are injected from the control gate to the floating gate when the control gate is biased with a negative (or positive) voltage.

It was observed that the total number of current carriers trapped at the floating gate injected from the substrate is much smaller than that injected from the control gate.

1. Introduction:

In a previous paper[1] the characteristics of stacked-gate electrically-alterable MOS memory transistors were shown. The floating gate of this type of device can be charged positive or negative when the control gate of the device is biased with large positive or large negative voltages, respectively. No charge transport from the substrate to the floating gate was reported. From the band diagram given by Ref. [1], one expects electrons (or holes) to inject from the substrate to the floating gate when a large positive (or a large negative) voltage is applied to the gate. Experimental data indicate that this is, indeed, the case. These experimental results provide better understanding of the device.

2. Experimental Results

The device used in this study is a stacked-gate MOS/SOS electrically-alterable transistor. The structure of this device is shown in Fig. 1. The thickness of the silicon dioxide between the floating gate and the silicon substrate and that between the floating and the control gate are equal to 150 Å and 700 Å, respectively. The floating gate is a 3000-Å thick layer of undoped polycrystalline silicon. The control gate is made of boron-doped polycrystalline silicon. The silicon film is 6000 Å n type, and the source and the drain are boron implanted P^+ regions.

The memory characteristic of this device is shown in Fig. 2., in which the threshold voltage of the device is plotted as a function of the amplitude of the voltage pulse applied to the control gate with respect to the source and drain. The width of the pulse is equal to 0.6 ms. The threshold voltage is defined as the gate voltage of the device when the drain current is equal to 1 μA at a drain voltage of -2 V. The virgin threshold voltage of the device is equal to -1.3 V. The curve in Fig. 2 was generated as follows. The theshold voltages of the device were measured both before and after the application of a voltage pulse to the control gate. If the theshold voltage of the device was changed by the pulse, it was readjusted to the original value before the next pulse with an amplitude larger than that of the previous one was applied. This procedure was repeated for both positive and negative polarities of pulses to complete the curve.

This result implies that when a large positive voltage is applied to the control gate, the floating gate is charged with positive charges. Therefore, either holes are injected from the control gate to the floating gate or electrons are extracted from the floating gate to the control gate. Simi-

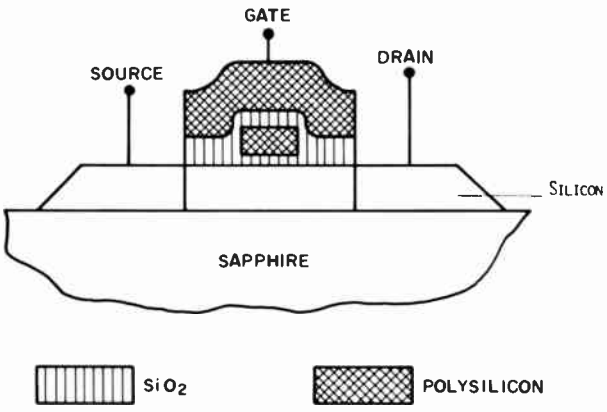


Fig. 1—Stacked-Gate MOS memory transistor used in this study.

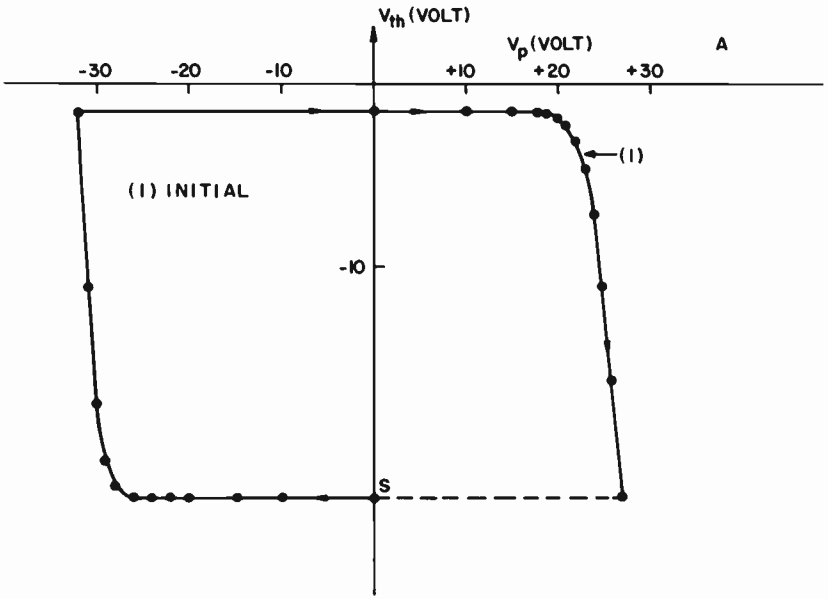


Fig. 2—Characteristic of a typical stacked-gate MOS memory transistor.

larly, when a large negative voltage is applied to the control gate, electrons are stored at the floating gate by either injecting electrons from the control gate to the floating gate or extracting holes from the floating gate to the control gate. These two transport mechanisms cannot be separated by the above result alone.

It can be shown that the charge stored at the floating gate, Q_f, is given by[1]

$$Q_f = C_1 (V_T - V_{TO}), \qquad [1]$$

where C_1 is the unit area capacitance between the floating gate and the control gate, V_T is the threshold voltage, and V_{TO} is the virgin threshold voltage of the device. At $V_T = -22.5$ V the holes stored at the floating gate are approximately equal to 6×10^{13} cm^{-2} or 2×10^{18} cm^{-3}, which corresponds to a charge transport rate of 3.3×10^{21} charges per cm^3 per sec. The electrons or the holes extracted from the floating gate are unlikely to have such a large rate.

The threshold voltage of the device used in Fig. 2 has been measured as a function of the number of pulses applied to the gate. The amplitudes of the pulses were choosen to have field intensities near the threshold field of the gate injection. The results are plotted in Fig. 3. Contrary to the result plotted in Fig. 2, the threshold voltage of the device is increased

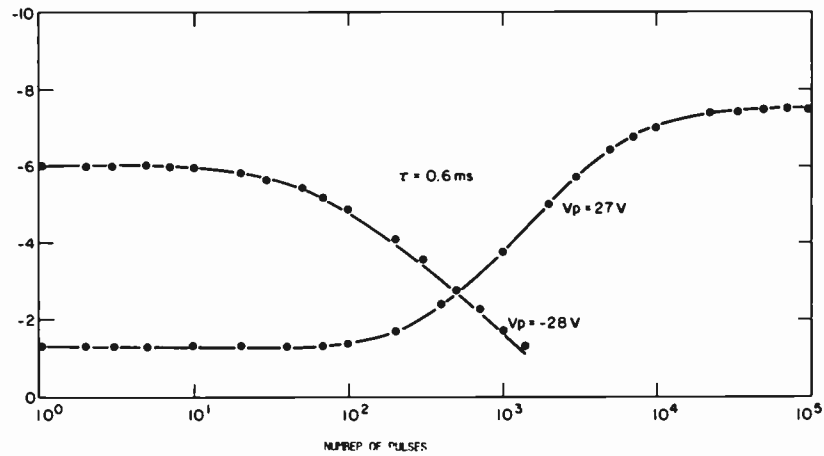


Fig. 3—Theshold voltage as a function of number of pulses for $V_P = -27$ V and $V_P = -28$ V.

when more than one hundred −27 V, 0.6 ms pulses are applied to the control gate. The threshold voltage of the device becomes saturated to −7.5 V when the number of pulse is larger than 5×10^5. The floating gate is, therefore, positively charged with a maximum hole density of 2×10^{13} cm^{-2} or 7×10^{17} cm^{-3}. The possible mechanisms for holes to be trapped in the floating gate when the control gate is biased with a negative voltage are (1) hole injection from the substrate to the floating gate and (2) electron extraction from the floating gate to the substrate. Also plotted in Fig. 3 is the threshold voltage as a function of the number of −28 V, 0.6 ms pulses applied to the control gate. The threshold voltage of the devices decreases with the number of pulses applied to the control gate when the amplitude of the pulses is larger than or equal to −28 V. Therefore, the number of holes trapped at the floating gate decreases with the number of pulses. *Correspondingly*, the density of electrons trapped at the floating gate increases with the number of pulses applied to the control gate. Since the electron density increases when a pulse with amplitude equal to −28 V is applied to the control gate, it is impossible for electrons to be extracted from the floating gate to the substrate at $V_P = -27$ V. Consequently, at $V_P = -27$ V holes are injected from the substrate to the floating gate. At $V_P = -28$ V electrons are injected from the control gate to the floating gate.

In addition the results in Figs. 2 and 3 show that the maximum number of holes stored at the floating gate is much smaller than the number of electrons that can be accumulated at the floating gate when the control gate is biased with negative voltages. When the pulse amplitude is larger than or equal to −28 V, i.e., the oxide field intensity is larger than the

threshold field of gate injection, both hole injection from the substrate and electron injection from the control gate take place. The number of holes stored at the floating gate reaches its saturation value, and the additional holes injected from the substrate flow directly to the control gate. Electrons injected from the control gate are stored at the floating gate. As a result, the floating gate becomes less positively (or more negatively) charged when the number of pulses applied to the gate is increased.

3. Theoretical Considerations

The current carrier transport and storage phenomena discussed above is due to interface effects. It is well known that there is a thin layer, probably of the order of 10 Å, of silicon-rich oxide at the single-crystal silicon–silicon–dioxide interface.[2–5] Hence the energy diagram of a Si–SiO_2 system has a smooth transition from the single-crystal silicon to the silicon dioxide. When the silicon dioxide is grown on the polycrystalline silicon layer, this interface layer is much larger than it is in the single-crystalline silicon silicon-dioxide system.[6] The interface region of a polycrystaline silicon deposited onto a silicon dioxide surface is presumably very abrupt. This is sketched in Fig. 4A.[1] The energy diagram of a stacked-gate MOS system when a large negative voltage is applied to the control gate is sketched in Fig. 4B. Because of the interface transition region, the barrier height at the Si–SiO_2 interface, E_1, is smaller than that at the control gate polycrystalline silicon–silicon dioxide interface, E_4. Holes can be injected from the substrate silicon through E_1 to the floating gate at relatively low fields as compared to electrons injection from the control gate through the energy barrier E_4. The maximum numbers of holes and electrons that can be stored at the floating gate are controlled by the barrier heights E_3 and E_2, respectively. Since E_3 is much smaller than E_2, the maximum number of holes that can be stored at the floating gate is much smaller than the number of electrons. Besides the Schottky barrier lowering effect, the potential barrier E_3 decreases with the bias voltage, while the potential barrier E_2 is almost independent of the bias voltage. Therefore, the maximum number of holes that can be stored at the floating gate decreases as the negative gate-bias voltage is increased. Similarly, when the control gate electrode is biased with positive voltages, electron injection from the substrate occurs at smaller voltages, while hole injection from the control-gate polycrystalline silicon take place at relatively higher voltages. The maximum number of holes that can be stored at the floating gate under the positive gate-bias condition is much larger than the number of electrons.

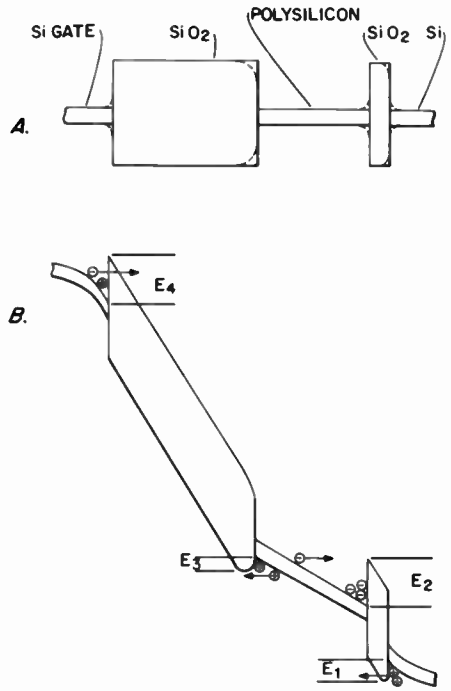


Fig. 4—Energy diagram of a stacked-gate MOS system (A) at equilibrium and (B) with a large negative voltage applied to the control gate.

4. Conclusions

Stacked-gate MOS memory transistors were used as vehicles to investigate the current carrier transport phenomena in an MOS structure. Hole injection from the substrate to the floating gate and electron injection from the control gate to the floating gate have been observed when the conrol gate is biased with a large negative bias voltage. The number of holes that can be stored at the floating gate is much smaller than the number of electrons at negative gate biases. The energy band structure was used to explain these experimental observations.

Acknowledgment

The author wishes to thank M. Jones for her assistance in device fabrication.

References:

[1] S. T. Hsu, "GIMOS a Non-Volatile MOS Memory Transistor," RCA Rev., **42**, p. 424, Sept. 1981.

[2] T. W. Sigmon, W. K. Chu, E. Lugujjo, and J. W. Mayer, "Stoichiometry of Thin Silicon Oxide Layers on Silicon," *Appl. Phys. Let.*, **25**, p. 105, Feb. 1974.

[3] W. L. Harrington, R. E. Honig, A. M. Goodman and R. Williams, "Low Energy Ion-Scattering Spectrometry (ISS) of the SiO_2/Si Interface," *Appl. Phys. Let.*, **27**, p. 644, Dec. 1975.

[4] J. S. Johannessen and W. E. Speicer, "Auger Depth Profiling of Interfaces in MOS and MNOS Structures," *J. Vac. Sci. Techn.*, **13**, p. 849, July/Aug. 1976.

[5] J. S. Johannessen, C. R. Holms, W. E. Speicer and Y. E. Strausser, "Auger Depth Profiling of MNOS Structures by Ion Sputtering," *IEEE Trans. Elec. Devices*, **ED-24**, p. 547, May 1977.

[6] D. J. DiMaria and D. R. Kerr, "Interface Effects and High Conductivity in Oxides Grown From Polycrystalline Silicon," *Appl. Phys. Lett.*, **27**, p. 505, Nov. 1975.

A Comparison of p-i-n and Schottky Barrier Hydrogenated Amorphous Silicon, a-Si:H, Solar Cells*

Richard S. Crandall

RCA Laboratories, Princeton, NJ 08540

Abstract—Measurements of the photoconductive response of Schottky barrier and p-i-n solar cell structures made from hydrogenated amorphous silicon are presented to compare the properties that determine their efficiency. It is demonstrated that the photoconductivity determines the fill factor rather than the dark current, as in crystalline silicon cells. Analysis of the current voltage curves of the p-i-n cells are used to show that the electron and hole drift lengths are of the same order in the i layer of the cell.

1. Introduction

A promising material for solar cell applications is hydrogenated amorphous silicon (a-Si:H). Thin-film solar cells with high efficiencies have been fabricated from this material.[1–3] Both Schottky barrier and p-i-n cells give high efficiency. Nevertheless, the p-i-n structure appears to give the highest efficiency for large area cells.[2] In this article we point out some of the reasons why the p-i-n structure may ultimately make the best solar cell. We shall discuss this in the framework of the factors that determine its efficiency. Generally, these can be represented by the fill factor, short-circuit current, and the open-circuit voltage. It is the electron and hole transport that determine the fill factor, essentially the

* Research reported herein was supported by Solar Energy Research Institute, under Contract No. XG-0-9372-1, and by RCA Laboratories, Princeton, N.J. 08540, U.S.A.

shape of the current–voltage curve. These play little role in determining the open-circuit voltage and a lesser role in determining the short-circuit current.

The a-Si:H based cell is different from the usual crystalline solar cell where diffusion of electron-hole pairs from the low-field bulk region to the high-field junction is the main source of the photocurrent. Since the electron–hole pair diffusion length is short in a-Si:H,[4] the collected current comes from those electron–hole pairs that are formed in the high-field junction region. Thus, both the transport in this region and the width of this region determine the collection efficiency. The best cells will be those in which the electric field extends throughout the entire intrinsic region.

In the first part of the article we present measurements of the voltage dependence of the photocurrent for both Schottky-barrier and p-i-n structures. In the second section we compare the results in terms of simplified models of the two structures.

2. Experimental Details

Results of photocurrent measurements on two types of thin-film solar cells produced by dc glow discharge decomposition of silane on heated substrates of stainless steel are presented below. Cells produced by rf discharge on other substrates show similar properties. The Schottky-barrier cell was made by first depositing a few-hundred-Å a-Si:H layer doped with phosphorus followed by an undoped layer.[5] A thin layer of platinum formed the Schottky-barrier contact. The p-i-n structure was grown by first depositing a thin boron-doped layer on the stainless steel, following this with an undoped layer, and finally adding a thin phosphorus-doped layer. The conducting transparent layer to the n-type layer was electron-beam-evaporated ITO.[6]

Photocurrent measurements were made using chopped light at frequencies low enough so that long response times do not limit the measurements. The light source was a krypton laser tuned to 674 nm. The absorption spectrum of a-Si:H is such that light of this wavelength is nearly uniformly absorbed.

Fig. 1 shows the photocurrent as a function of bias applied to a p-i-n solar cell illuminated with uniformly absorbed light through the n-type layer. Here and throughout this paper the photocurrent is the difference between the current under illumination with weakly absorbed light and the current in the dark. For negative voltage, the photocurrent saturates at sufficiently large reverse bias. For positive bias, the photocurrent decreases and changes sign at a bias V_f. The positive photocurrent is usually referred to as a primary photocurrent, because the contacts are

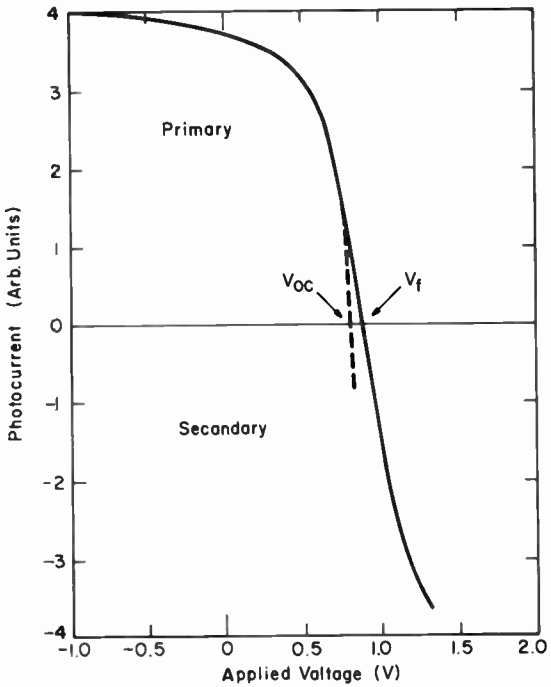


Fig. 1—Photocurrent of a p-i-n solar cell versus voltage for illumination with 674-nm light at room temperature. The dotted curve is the total current.

blocking for the photoexcited carriers, and the negative photocurrent is referred to as a secondary-photocurrent, because the contacts are now ohmic.[7] The dotted curve in Fig. 1 shows the total current (the sum of photocurrent and dark current) for reference purposes. Only at low temperature where the dark current is zero would the total current and photocurrent coincide. It is evident from the figure that the magnitude of the dark current is much smaller than that of the photocurrent except near V_{oc}.

The data in Fig. 1 show that the photocurrent in the vicinity of V_f varies linearly with voltage. This is the region where the photocurrent changes from primary to secondary. An important point is that the magnitude of the slope of the photocurrent-versus-voltage curve is the same for primary and secondary photocurrents. This implies that the transport mechanisms are the same for either type of photocurrent in the vicinity of V_f. The linear slope shows that the photocurrents obey Ohms law.

In Fig. 2 the photocurrent for a Schottky-barrier cell is shown for illumination through the platinum Schottky-barrier contact with weakly absorbed light. The dashed curve in the figure shows the photocurrent

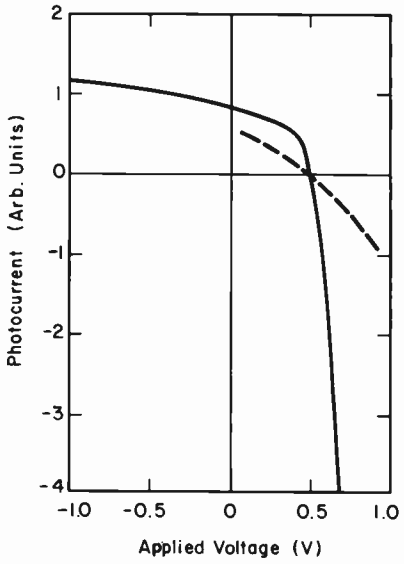


Fig. 2—Photocurrent of a Schottky-barrier solar cell as a function of cell bias. Experimental conditions are the same as in Fig. 1. The dashed curve is for a scale expansion of ten.

on an expanded scale in the vicinity of V_f. In this case, in contrast to the data in Fig. 1 for the p-i-n structure, the primary-photocurrent is not ohmic.

3. Discussion

3.1 Schottky-Barrier Cell

There is simple reason for the nonohmic nature of the primary photocurrent in the Schottky-barrier structure. This becomes apparent in Fig. 3 where the potential energy diagram is sketched. The Fermi level near the Schottky-barrier contact lies well below midgap in the region of the high density of valence band states.[8] In the process of forming this metal–semiconductor contact, the Fermi level must move from near midgap to well below midgap. To accommodate this, electrons leave the i-layer, producing a large density of positive charge in the region within a depletion width of the Pt-a-Si:H interface. This charge leads to a decrease in the electric field with distance from the Pt contact. Since the electron–hole pair diffusion length is low, pairs created in the low-field region will have little chance of being separated before they recombine

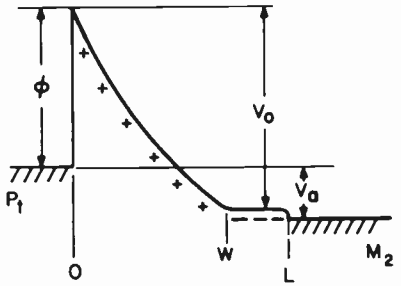


Fig. 3—Electron potential energy diagram for a Schottky-barrier solar cell made of n-type a-Si:H.

and, thus, will not be collected at the contacts. Electron–hole pairs produced in the high-field region have the best chance of being collected at the contacts. Therefore, the photocurrent will be proportional to the width of the high-field region. For the Schottky-barrier cells discussed in this article, the width of the high-field region can be described by the classical depletion width model,[9] in which the depletion width is proportional to the square root of voltage across the sample.

Evidence for this model is shown in Fig. 4 where the reciprocal of the square of the capacitance is plotted versus applied voltage. For a Schottky barrier, the curve should be a straight line as shown in the figure. These measurements were made at 100 Hz, but similar results were obtained between 10 and 5.10^{+4} Hz. The illumination intensity corresponds to 16 mW cm^{-2}.

Fig. 5 shows the photocurrent as a function of the depletion width (W), determined from the capacitance measurements. The conditions are the same as in Fig. 4. Except for the small region near the minimum deple-

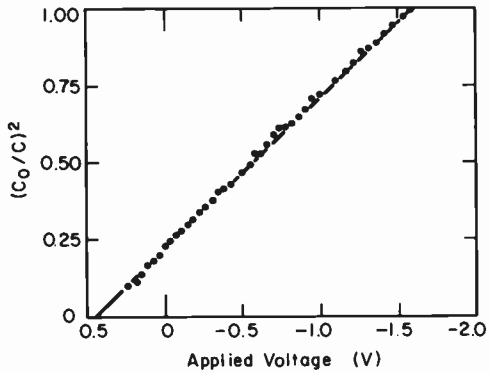


Fig. 4—Reciprocal of the square of the capacitance versus bias for a Schottky-barrier solar cell. The measuring frequency is 100 Hz; the cell is illuminated with 16 mW/cm^2 of 674-nm light.

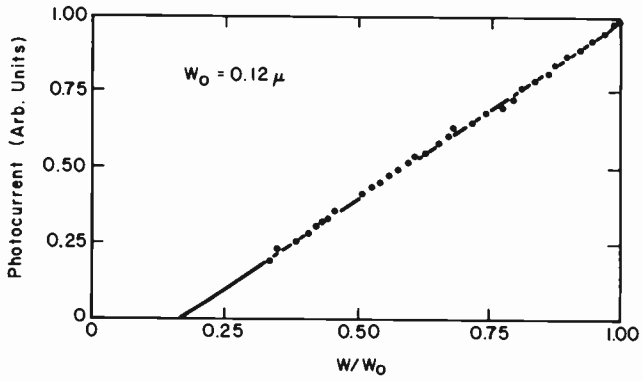


Fig. 5—Photocurrent versus depletion width for the same conditions as in Fig. 4.

tion width, the current varies linearly with the depletion width, consistent with the model that only carriers produced in this region are collected. Since the current varies linearly with the depletion width and the depletion width varies as the square root of the voltage, the current also varies as the square root of the depletion width. This is in contrast to the p-i-n cell where the photocurrent obeys Ohms law near V_f. The fact that the line does not extrapolate through the origin, but rather has an intercept at finite depletion width, has been interpreted in terms of geminate recombination.[10] Perhaps a more reasonable explanation for these data would be a field-dependent trapping of either an electron or hole at a defect. In this case, the effect should be sample dependent and consistent with the failure to find geminate recombination in p-i-n solar cells.

The secondary-photocurrent, on the other hand, is ohmic as expected since for applied voltages greater than V_f, the electric field extends throughout the thickness of the a-Si:H film.

3.2 p-i-n Solar Cell

To describe the transport in the p-i-n structure we refer to the idealized potential energy diagram in Fig. 6. This figure is similar to that for the Schottky barrier structure shown in Fig. 3, with one important exception. That is that the potential varies nearly linearly across the i-layer because there is little space charge in the i-layer. This occurs because of the growth procedure. The solar cell is grown with the p-layer put down first and followed by the i-layer. Because the discharge is not terminated and all the boron purged from the system, there is residual boron in the i-layer. The boron concentration decreases from the p side toward the n side resulting in an i-layer that is graded p to n type. This has an im-

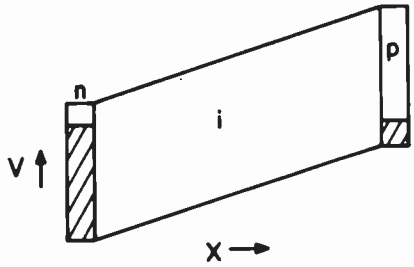


Fig. 6—Potential energy diagram for a p-i-n solar cell. The n and p layers contain the space charge that is necessary to support the uniform field in the i-layer.

portant effect on the behavior of the solar cell. It produces an i-layer that contains much less space charge than the Schottky-barrier cell. The majority of the space charge resides in the depletion widths in the n and p contacts. Thus one can reasonably approximate the field in the i-layer as being uniform, simplifying the analysis of the transport data.

Support for this model of the p-i-n structure is given by capacitance–voltage measurements. For the model of the p-i-n structure with all the space charge contained in the n and p layers, the derivation in Appendix 1 shows that the depletion width (reciprocal of the capacitance) varies linearly with voltage. The capacitance–voltage data presented in Fig. 7 emphasize this behavior. The measurements were made at 250 K to minimize the contribution from the diffusion capacitance in forward bias.[11] The data in the figure clearly exhibit this linear dependence of the depletion width on bias. The depletion width of the p-i-n structure changes by only 3% over the large voltage range, because it is

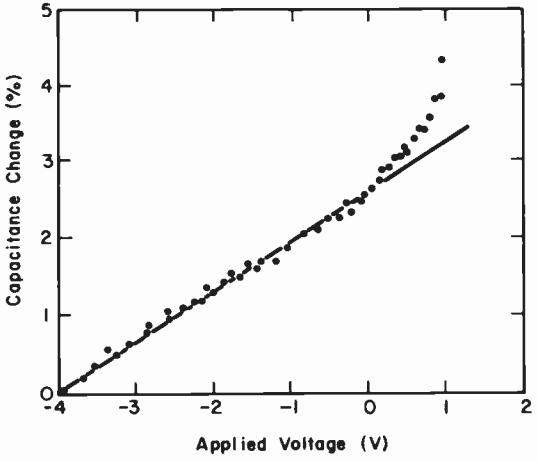


Fig. 7—Fractional capacitance change versus bias for a p-i-n solar cell.

the depletion widths in the n and p contacts (which are a small fraction of the i-layer thickness) that change with voltage. These results are not consistent with recent proposals[12] that the best description of the potential in the i-layer in the p-i-n structure is that of two Schottky barriers in series. For this case, the capacitance should vary as the square root of the voltage and, furthermore, there would be large changes in the capacitance with voltage as for the Schottky-barrier cell.

If we take the above capacitance-voltage results as evidence that the p-i-n structure can be described by a uniform field throughout the i-layer, then the voltage dependence of the photocurrent can be understood in terms of a simple model of the transport in this cell. For the linear portion of the current-voltage curve shown in Fig. 1, both the primary (positive) and secondary (negative) photocurrents must be described by the Ohms law. As shown in Appendix 2, the photocurrent can be approximated by

$$J = e(nv_n + pv_p) = eG(l_n + l_p), \qquad [1]$$

where e is the magnitude of the electron charge, n the density of photo-excited electrons per unit voilume, and v_n the electron drift velocity; the quantities p and v_p have similar definitions for the holes. The quantity G is the number of electron–hole pairs produced per unit volume per unit time. The second expression follows from the definitions of the drift length and lifetime. The electron and hole drift lengths (mean distance travelled in the direction of the field before recombination) are, respectively,

$$l_n = v_n\tau_n; \quad l_p = v_p\tau_p. \qquad [2]$$

The times τ_n and τ_p are, respectively, the lifetimes of the electron and hole before recombining. The free electron and hole densities, in the steady state, are given by

$$n = G\tau_n; \quad p = G\tau_p. \qquad [3]$$

That the current voltage curves have the same slope on either side of V_f implies that the transport parameters do not change as the electric field in the i-layer changes sign. Since the contacts do not determine the transport at low fields, it is determined by recombination in the bulk. The photocurrent given by Eq. [1] is just the usual photocurrent that would be measured on a bulk film with ohmic contacts. This type of measurement is often used as a guide to the quality of the photoconductor and to the nature of the recombination centers.[8,13,14] In a-Si:H one usually finds that the recombination centers are distributed in energy.[13,14] The evidence for this is[13,14] that the photocurrent exhibits a power-law variation with light intensity where the power is between 0.5

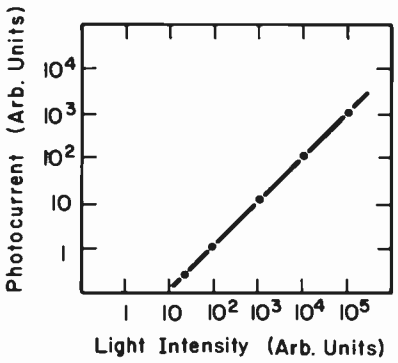


Fig. 8—Logarithm of the photocurrent versus light intensity using 674-nm light. Measurements were made on a p-i-n cell in the Ohms-law region near V_f. The maximum light intensity is about 120 mW/cm^2 incident on the cell.

and 1.[7] Typical values for a-Si:H are between 0.7 and 0.8.[13,14] The photocurrent in the p-i-n structure, on the other hand, shows a linear variation of photocurrent with light intensity typical of a single recombination center. The photocurrent versus light intensity for the p-i-n structure is shown in Fig. 8. Here the logarithm of the photocurrent is plotted versus the logarithm of the light intensity. The slope of the line through the data points is unity.

As the electric field increases, the sum of the electron and hole drift lengths will increase until they are on the order of the i-layer thickness (L). Then the contacts, which are blocking for this primary-photocurrent, will limit the current flow reducing it below the Ohms-law value. When the sum of these drift lengths significantly exceeds the i-layer thickness the current saturates at a value eGL.

This concept is aided by reference to Fig. 9 which is a sketch of the electron and hole densities in the i-layer for the condition where $l_p + l_n < L$. In this case, strong recombination in the bulk of the i-layer keeps the carrier density constant in this region. The electron and hole densities are thus given by Eq. [3]. The lifetimes are constants, independent of position. Near the blocking contacts, this approximation breaks down

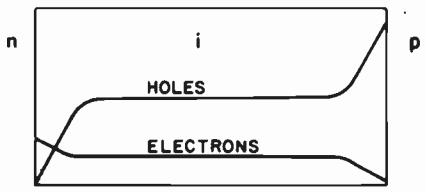


Fig. 9—The electron and hole density as a function of distance in the i-layer of a p-i-n cell.

within a distance on the order of the minority-carrier drift length. Near the n-i interface, this length is the drift length of the minority-carrier *holes* and at the i-p interface it is the drift length of the minority-carrier *electrons*. Since there is little recombination in these regions, the electron and hole densities increase linearly away from the interfaces.

To make an exact calculation of the detailed shape of the photocurrent-voltage curve in the transition region between the Ohms-law and saturated-current regions requires a numerical solution of the continuity equation as well as a detailed knowledge of the spatial dependence of the electron–hole recombination times. Nevertheless, arguments are given in Appendix 2 for an approximate solution to the continuity equation that agrees quite well with experiment. However, even though a simple expression can be obtained that connects the Ohms-law region with the transition region, no new physical information is obtained, since the only parameter in the expression for the current (Eq. [23]) is the sum of the electron and hole drift lengths, which can also be determined without a detailed knowledge of the transport, as will be outlined below.

From the ohmic and saturated regions of the primary-photocurrent–voltage curve, one can determine the important transport parameters, namely the drift lengths. The saturated value of the photocurrent determines the magnitude of the generation rate G independent of any knowledge of the absorption constant or transmission coefficient of the contacts. Dividing G, determined in this manner, into the photocurrent in the ohmic region gives the sum of the electron and hole drift lengths. We cannot separate the sum into its constituents, but this is not a drawback for solar cell operation because it is the sum of these two that determines the operating parameters for the cell.

Analysis of the data in Fig. 1 shows that $\overline{\mu\tau} = \mu_p\tau_p + \mu_n\tau_n = 7 \times 10^{-9}$ $\text{cm}^2\,\text{V}^{-1}$. This number can be compared with estimates of the $\mu\tau$ products determined by other means. Moore[15] has made measurements of the hole diffusion length using the photo-electromagnetic effect on thin films of a-Si:H. He finds a diffusion length of about 0.1 μm. From his data he determined $\mu_p\tau_p = 3.2 \times 10^{-9}$ $\text{cm}^2\,\text{V}^{-1}$. Substituting this value into the sum of the $\mu\tau$ products determined above, we find that $\mu_n\tau_n \approx 3.8 \times 10^{-9}$ $\text{cm}^2\,\text{V}^{-1}$, on the same order as the $\mu\tau$ product for holes. This is a surprising result since it is generally believed that the hole drift length is much shorter than the electron drift length. Also, Moore[15] found that $\mu_n\tau_n = 7.8 \times 10^{-8}$ $\text{cm}^2\,\text{V}^{-1}$, about 20 times the value found for the p-i-n cell. The difference may be a property of the p-i-n cell in that the i-layer is presumably different from the material studied by Moore because of the residual boron in the i-layer of the p-i-n cell. This may reduce the electron lifetime.

In Fig. 10, current-voltage curves for two p-i-n cells are shown to compare the more detailed expression (Eq. [23]) with experiment. The data are dots and circles and the curves are Eq. [23] with an appropriate choice of $\mu_p\tau_p + \mu_n\tau_n = \overline{\mu\tau}$. The agreement between theory and experiment is good showing that one parameter is sufficient to explain the data. Of course this parameter can also be determined from the slope of the curve in the Ohms-law region. For curve A, $\overline{\mu\tau} = 9.5 \times 10^{-9}$ cm^2 V^{-1} and for curve B, $\overline{\mu\tau} = 1.7 \times 10^{-9}$ cm^2 V^{-1}. The difference between the two is presumable due to the different preparation conditions.

4. Summary

The results presented in this article indicate that the p-i-n structure should make a better solar cell than the Schottky-barrier cell. This is supported by reports that show the p-i-n structure to give the highest efficiency. One of the reasons that the p-i-n structure is better is that the density of defects is low in the i-layer. In this article we have compared the photovoltaic behavior of p-i-n and Schottky-barrier solar cells. The main difference between the two types of cells is that the electric field in the p-i-n cell is nearly uniform across the i-layer, whereas the field is nonuniform in the Schottky-barrier cell. This is because the Schottky-barrier cells trap holes under illumination, which causes a significant decrease in the depletion width.[14] Even for cells that may be fully depleted at zero bias in the dark, there is a collapse of the depletion width to a fraction of the cell thickness under one sun illumination. For p-i-n structures, hole trapping is not a problem under strong illumination. DLTS measurements show little hole trapping. The reasons for the

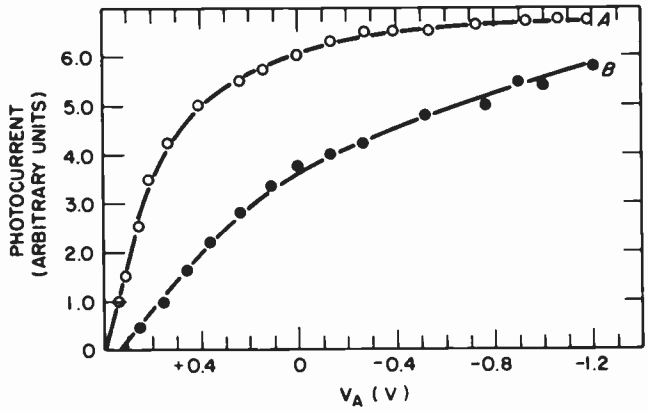


Fig. 10—Photocurrent as a function of applied bias for two p-i-n cells illuminated with 674-nm radiation. The dots and circles are data; the curves are Eq. [23].

superior behavior of the p-i-n cell in this respect is not understood at present but may have something to do with the different growth procedures for the two types of cells.

A model calculation of the photocurrent for the p-i-n cell showed that the parameter that determines the photovoltaic behavior is the sum of the electron and hole drift lengths. This result, that both carriers play a significant role in the transport, is due to the fundamental property that the electrons and holes must recombine in pairs. Comparisons of the electron and hole drift lengths determined from the current-voltage curve with independent measurements of the hole drift length led to the conclusion that the electron drift length in the p-i-n structure is lower than that in a bulk film. This is presumable due to the residual boron in the i-layer of the cell.

Acknowledgment

I am indebted to Dave Carlson for supplying me with the solar cells.

Appendix 1

The electric fields in the p-i-n structure depicted in Fig. 6 can be readily derived by standard methods. We make the assumption that the i-layer thickness is fixed at a length L and the n and p layers are thick enough so that they do not become fully depleted under reverse bias. The n-layer contains positive space charge of density ρ_+; the p-layer contains negative space charge ρ_-. The depletion widths in these layers are W_+ and W_- for the n and players, respectively.

Because of the growth procedure for the p-i-n cell, the i-layer is expected to be graded p to n type from the p-i interface. Therefore, it is reasonable to represent the space charge in the i-layer by negative charge decreasing away from the i-p interface. The space-charge densities are assumed uniform in the p and n layers.

Poissons equations in the three regions are thus

$$\frac{dE_n}{dx} = \frac{\rho_+}{\epsilon} \qquad [4]$$

$$\frac{dE_p}{dx} = -\frac{\rho_-}{\epsilon} \qquad [5]$$

$$\frac{dE_i}{dx} = -\rho_- \left(\frac{x}{L}\right)^{\alpha} \frac{1}{\epsilon} + \frac{\rho_s}{\epsilon} \qquad [6]$$

where the subscripts on the electric field E refer to the different layers. The coordinate system is chosen with $x = 0$ at the i-n interface and $x =$

L at the i-p interface. A power law with exponent α was chosen for the space-charge gradation in the i-layer. A uniform positive space charge of density ρ_s is assumed present in the i-layer.

Integration of Eqs. [4]–[6] gives

$$E_n = \frac{\rho_+(x + W_+)}{\epsilon} \qquad [7a]$$

$$E_p = -\frac{\rho_-(x - W_-)}{\epsilon} \qquad [7b]$$

$$E_i = -\frac{\rho_- L(x/L)^{\alpha+1}}{(\alpha + 1)\epsilon} - \frac{\rho_s x}{\epsilon} + E_0. \qquad [7c]$$

The depletion widths and the uniform part of the field in the i-layer are found by matching the fields at the boundaries. Integrating Eqs. [7a]–[7c] and again using the boundary conditions permits all the constants to be expressed in terms of V_o the potential difference across the structure.

After some algebraic manipulation, the constant field in the i-layer can be expressed in terms of the potential and the space-charge densities. However, the expression is rather unwieldy in its general form. Since the experimental conditions are usually such that the space charge gradient is large in the i-layer and, furthermore, the impurity density ρ_s is much less than the space-charge densities in the p and n layers, E_o can be approximated by

$$E_o = \frac{V_o}{(1 - b)/L} - \frac{L\rho_s}{\epsilon}, \qquad [8]$$

where $b = \rho_s/N_A \ll 1$. This expression shows that as long as the average field, V_o/L, in the i-layer is greater than the surface charge field, $L\rho_s/\epsilon$, due to the space charge in the i-layer, the field in the i-layer is approximately uniform.

The depletion width W for this structure is the sum of W_+ and W_-. It is given by

$$W = \xi\left[1 + \frac{4}{\xi^2 L^2}\left(\frac{\xi L^2}{(\alpha + 1)(\alpha + 2)2\epsilon} + V_o\right)\frac{\epsilon}{2e}\left(\frac{1}{N_D} + \frac{1}{N_A}\right)\right]^{-1/2}, \qquad [9]$$

where $\xi = \alpha/(\alpha + 1)$ and N_D and N_A are the donor and acceptor densities in the n and p layers, respectively. In the limit of small α, the space charge is nearly uniform in the i-layer, and $W \propto \sqrt{V_o}$. In this case the p-i-n cell would behave like a Schottky-barrier cell. In the opposite limit, where $\alpha \gg 1$ and the doping density decreases rapidly away from the p-i interface, the depletion width is

$$W \approx L\left[1 + \frac{\epsilon}{e}\left(\frac{1}{N_D} + \frac{1}{N_A}\right)\frac{V_o}{L^2}\right], \qquad [10]$$

which shows a linear change in depletion width with voltage in agreement with the data in Fig. 7.

Appendix 2

An exact analytical solution of the coupled continuity and Poissons equations may not be possible for the p-i-n structure. Apart from the difficulty resulting from the coupling of Poissons' equation to the continuity equations through the free-carrier space charge, the problem of the strong spatial dependence of the free-carrier lifetimes makes an exact solution difficult. This spatial dependence of the lifetimes results from the fundamental condition that the electrons and holes must recombine in pairs so that their respective recombination rates are equal at all points in space. Stated mathematically this condition is

$$\frac{n(x)}{\tau_n(x)} = \frac{p(x)}{\tau_p(x)} = R. \qquad [11]$$

Reference to Fig. 9, which is a sketch of the electron and hole density in the i-layer, shows that near the blocking contacts n and p have strong variations with distance from the contacts. Because of the above condition on the recombination rates, this will result in a strong variation in the lifetimes.

Nevertheless, we can make an approximate solution of the system of Poissons and the continuity equations that can give physical insight into the transport in the p-i-n structure. Surprisingly enough, this approximate solution agrees well with experiment. First, we simplify the problem by taking the electric field across the i-layer to be uniform. This was discussed in the text and shown to be a good assumption. Second, we restrict the solutions to those for uniformly absorbed light. Under these conditions, diffusion can be neglected.[15] This simplifies the algebra somewhat. Under these conditions, the set of equations to be solved is

$$-v_p \frac{dp(x)}{dx} + G - \frac{p(x)}{\tau_p(x)} = 0 \qquad [12]$$

$$v_n \frac{dn(x)}{dx} + G - \frac{n(x)}{\tau_n(x)} = 0 \qquad [13]$$

The boundary conditions for these equations are that the minority carriers vanish at the blocking contacts. This means that the hole density vanishes at the n-i interface, which is blocking for holes, and similarly the electron density vanishes at the p-i interface. Therefore,

$$p(o) = n(L) = 0 \qquad [14]$$

As discussed above, these equations cannot readily be solved because

of the lack of knowledge of the x dependence of the lifetimes. An expression for the lifetime that might apply to the p-i-n structure is that derived from the Shockley-Read[16] recombination theory. However, the solutions of Eqs. [12] and [13] that result using this theory are mathematically complex and it is difficult to obtain physical insight from them. Therefore, we choose to make the following approximations to $\tau_n(x)$ and $\tau_p(x)$ that permit straight-forward solutions of Eqs. [12] and [13]. Actually the Shockley-Read recombination theory shows that these approximate expressions for the lifetimes are reasonably valid where they are applied.

We make a constant lifetime regional approximation in which the i-layer is separated into two regions. In each of these regions R is approximated by a linear function of n or p. Let region 1 be between $x = 0$ and $x = x_c$. In this region, approximate R by $R = p/\tau_p^0$. Region 2 is between $x = x_c$ and $x = L$. Here let $R = n/\tau_n^0$. The point x_c is the cross-over point where the recombination changes from hole to electron dominated. At this point, $p/\tau_p^0 = n/\tau_n^0$. The recombination times τ_p^0 and τ_n^0 are constants independent of x and the carrier density.

In region 1 near $x = 0$ where the hole density is zero, R can be neglected with respect to G. Thus Eq. [12] becomes

$$v_p \frac{dp(x)}{dx} = G. \qquad [15]$$

With the boundary condition (Eq. [14]) the solution of this equation is

$$p(x) = Gx/v_p, \qquad [16]$$

which is a hole density that increases linearly from the n-i interface. Current continuity requires that the electron density in this region will also have a linear variation with x reaching a maximum at $x = 0$.

Similar considerations apply in region 2 near $x = L$ at the p-i interface, where $R \ll G$ so that the approximate solution of Eq. [13] is

$$n(x) = G(L - x)/v_n. \qquad [17]$$

If $R \approx G$ at some region of the i-layer, the drift term in the continuity equations can be neglected and the electron and hole densities are given by the usual expressions in photoconductors.

$$n = G\tau_n; \quad p = G\tau_p. \qquad [18]$$

In this region, the lifetimes are constant independent of x so that $\tau_n(x) = \tau_n$ and $\tau_p(x) = \tau_p$. The current will thus be given be given by the usual Ohms-law expression. If on the other hand $R < G$ throughout the i-layer, the drift term in Eqs. [12] and [13] dominates and the current saturates.

The above arguments reproduce the general features of the current-voltage curve in Fig. 1. However, it is desirable to connect the two limits of strong recombination and ohmic current flow with the regime of a saturated photocurrent.

Using the constant lifetime approximation permits straight-forward solutions of the continuity Eqs. [12] and [13]. The solution of Eq. [12] valid in region 1 is

$$p = G\tau_p^0[1 - \exp(-x/l_p)]. \tag{19}$$

This expression for p can be combined with Eq. [3] to determine n in region 2. Similarly,

$$n = G\tau_n^0[1 - \exp\{(x - L)/l_n\}] \tag{20}$$

is the solution of Eq. [13] valid in region 2. Eq. [19] shows that if $x < l_p$ then p increases linearly with x. There is a similar increase in n away from the p-i interface at $x = L$. Eqs. [19] and [20] and the condition that $p/\tau_p^0 = n/\tau_n^0$ at $x = x_c$ are sufficient to determine x_c:

$$x_c = \frac{L}{1 + (l_n/l_p)} \tag{21}$$

which shows that the crossover from hole- to electron-dominated transport is determined by the ratio of the drift lengths.

The photocurrent is found by substituting Eqs. [20] and [21] into the first integral of the continuity equations,

$$J = e[v_n n(x) + v_p p(x)] \tag{22}$$

and evaluating the expression at $x = x_c$. The result is

$$J = eG[l_n + l_p][1 - \exp\{-L/(l_n + l_p)\}]. \tag{23}$$

It now remains to show that the photogenerated carriers do not cause the electric field in the i-layer to significantly depart from uniformity. To show this we calculate the field distortion due to the free-carrier space charge using Poissons equation,

$$\epsilon \frac{dE(x)}{dx} = e[p(x) - n(x)] \tag{24}$$

Instead of calculating the field distortion in general, we shall calculate it for the worst case. This is when there is no recombination and the mobility of one carrier is much larger than that of the other. Since, in this regime, the space charge is inversely proportional to the drift velocity, the slow carrier contributes the majority of the space charge. If $\mu_p \ll \mu_n$, then

$$\Delta E(x) = \frac{eGx^2}{2\epsilon v_p}, \tag{25}$$

where $\Delta E(x)$ is the field distortion. As the condition on the amount of field distortion we calculate the fractional change in voltage due to this distortion,

$$\epsilon_r = \int_0^L \frac{\Delta E(x)dx}{V_o} = \frac{1}{6}\frac{eGL^4}{\epsilon\mu_p V_o^2}. \qquad [26]$$

As an example, let the absorbed photon intensity correspond to a short-circuit current of 0.012 A, the i-layer be 5×10^{-5} cm, $\mu_p = 10^{-1}$ cm^2 V^{-1} s^{-1}, and $V_o = 1$ V. Then $\epsilon_r = 0.0025$, which is indeed small. Even though the error increases as V_o decreases, it will not become as large as Eq. [26] predicts because the space charge decreases under conditions of strong recombination.

References:

[1] D. E. Carlson, "Recent Developments in Amorphous Silicon Solar Cells," *Solar Energy Materials*, **3**, p. 503 (1980).

[2] Y. Tawada, T. Yamaguchi, L. Nonomura, S. Hotta, H. Okamoto, and Y. Hamakawa, 2nd Photovoltaic Science and Eng. Conf. in Japan (1980).

[3] A. Madan, J. McGill, W. Czarbatyj, J. Yang, and S. R. Ovskinsky, "Metal-Insulator Solar Cells Using Amorphous Si:F:H Alloys," *Appl. Phys. Lett.*, **37**, p. 826 (1980).

[4] D. L. Staebler, "Hole Diffusion Length Measurements in Discharge-Produced a-Si:H," in Proc. 8th Int. Conf. on Amorphous and Liquid Semiconductors, Cambridge, MA, Ed. by W. Paul, in *J. Non-Cryst. Solids*, **36**, p. 387 (1980).

[5] C. R. Wronski, D. E. Carlson, and R. E. Daniel, "Schottky Barrier Characteristics of Metal-Amorphous Silicon Diodes," *App. Phys. Lett.*, **29**, p. 602 (1976).

[6] Indium-Tin-Oxide transparent contact.

[7] A. Rose, *Concepts in Photoconductivity and Allied Problems*, Interscience, New York (1963).

[8] H. Fritsche, *Solar Energy Materials*, **3**, p. 447 (1980).

[9] S. M. Sze, *Physics of Semiconductor Devices*, p. 370, Wiley-Interscience, New York (1969).

[10] R. S. Crandall, R. Williams, and B. Tompkins, "Collection Efficiency Measurements on a-Si:H Solar Cells," *J. Appl. Phys.*, 50, p. 5506 (1979).

[11] See Ref. [9], p. 107.

[12] A. Rothwarf, "Model for Field and Light Dependent Effects in a-Si Solar Cells," 15th IEEE Photovoltaic Specialists Conf. Kissimmee, Florida, 12 May 1981; V. Dalal, "Analysis of Amorphous Silicon Solar Cells," *Solar Cells*, **2**, p. 261 (1980).

[13] C. R. Wronski and R. E. Daniel, "Photoconductivity, Trapping, and Recombination in Discharge-Produced, Hydrogenated Amorphous Silicon," *Phys. Rev. B*, **23**, p. 794 (1981).

[14] R. Williams and R. S. Crandall, "Carrier Generation, Recombination, and Transport in Amorphous Silicon Solar Cells," *RCA Rev.*, **40**, p. 371, Dec. 1979.

[15] J. Reichman, "Collection Efficiency of Low-Mobility Solar Cells," *Appl. Phys. Lett.*, **38**, p. 251, 15 Feb. 1981.

[16] W. Schockley and W. T. Read, Jr., "Statistics of the Recombinations of Holes and Electrons," *Phys. Rev.*, **87**, p. 835 (1952).

Field Nonuniformity Due to Photogenerated Carriers in a p-i-n Solar Cell*

Richard S. Crandall

RCA Laboratories, Princeton, NJ 08540

Abstract—The changes in the electric field due to the free-carrier space charge in a p-i-n solar cell are calculated. A criterion for significant field reduction caused by space charge is presented. It is suggested that there will be significant electric field lowering for hydrogenated amorphous silicon, a-Si:H, solar cells much thicker than 1 μm under 1 sun illumination.

Introduction

There is considerable interest in the effects of space charge on the operation of a-Si:H solar cells.[1–5] The concern has been mainly with the trapped charge arising from the presumed large density of gap states of a-Si:H.[6] In n-type Schottky-barrier solar cells there is ample evidence that trapped charge controls the operation of the cell.[4,5] Nevertheless, in some p-i-n cells the space charge is insignificant in the absence of illumination.[7] Furthermore, the space charge under illumination is not associated with trapped charge, but rather with the free carriers themselves.[8] In a solar cell the photogenerated carriers are not replenished at the contacts. This produces, under the action of an applied field, a displacement of holes to one side and electrons to the other side of the cell. This space charge has been shown to give a measurable contribution to the capacitance.[8]

In this article we use a simplified model of a p-i-n solar cell to calculate the transport of photocarriers by solving the coupled continuity and Poissons equations. From this we draw definite conclusions as to the

* Research reported herein was supported by Solar Energy Research Institute, under Contract No. XG-0-9372-1, and by RCA Laboratories, Princeton, N.J. 08540, U.S.A.

operating conditions of practical cells. We find that the a-Si:H cell is limited primarily by the hole space charge, which predominates because of the low hole mobility.[9]

Simplified Model

A simple criterion to determine when space-charge considerations are important is that the field due to the hole space charge is of the order of the applied field. The hole space charge is roughly equal to the product of the short-circuit current and the hole transit time. To get a more accurate criterion and to see how the field and transport are affected by the space charge, one must solve the continuity equations along with Poissons equation. In general, this set of equations can only be solved numerically. However, there are situations of physical interest where an approximate solution can be obtained. One of these is when diffusion and recombination can be neglected. This is useful for finding a criterion as to when space-charge considerations are important since diffusion and recombination will tend to reduce the space charge below the value found without them.

It is convenient to define dimensionless densities, fields, and lengths. If G is the generation rate of electron-hole pairs per unit volume per unit time, q the elemental charge, ϵ the dielectric constant, L the i-layer thickness, μ_N the electron mobility, μ_P the hole mobility, E the electric field, P the hole density, N the electron density, X the position coordinate measured from the n-i interface, and $a = \mu_P/\mu_N$ then the normalized field, hole, and electron densities are $e = E/E_c$, $p = P/P_c$, and $n = N/N_c$, respectively where

$$E_C = L\sqrt{qG/\epsilon\mu_P}; \quad P_c = \sqrt{\epsilon G/q\mu_P}; \quad N_c = \sqrt{\epsilon G/q\mu_N}. \qquad [1]$$

In what follows, we shall assume that the light is absorbed uniformly throughout the i-layer. This permits the continuity equations to be written in the compact form

$$d(ep)/dx = 1; \quad d(en)/dx = -\beta \qquad [2]$$

and the Poisson equation becomes

$$\frac{de}{dx} = p - \beta n \qquad [3]$$

where $\beta = \sqrt{a}$. With the boundary condition that photogenerated charges are not replenished at the contacts ($p = 0$ at $x = 0$ and $n = 0$ at

$x = 1$), Eqs. [2] and [3] can be readily integrated. The first integrals of Eq. [2] are

$$ep = x; \quad en = \beta(1 - x). \qquad [4]$$

Substitution of Eq. [4] into [3] gives a differential equation that can be readily solved for the electric field as a function of position. This field is

$$e^2 = e_s^2 + (1 + a)\, x^2 - 2\, ax. \qquad [5]$$

The field at $x = 0$, the surface field, e_s, is found by integrating the electric field across the i-layer and setting the result equal to the voltage across the i-layer, V.

The field has a minimum at $x = a/(1 + a)$, which for equal electron and hole mobilities occurs in the center of the i-layer. However, for a-Si:H, it is more reasonable to take $a \ll 1$,[9] which shifts the minimum toward the n-i interface. Conversely, for $a \gg 1$ the minimum of the field is shifted toward the i-p interface.

Since the exact expression for e_s is cumbersome, it will not be reproduced here. However, there are two limiting forms. For $a \ll 1$, we find in the limit of $e_s > 1$ that $e_s = V/(E_c L)$. Therefore the surface field ($E_s = E_c e_s$) is V/L and the field in the i-layer is nearly uniform and equal to V/L. In the limit $e_s < 1$, $e_s = 2a\ V/(E_c L)$ and the surface field is $2a\ V/L$ which is much less than the average field V/L. In this case, the field is nonuniform in the i-layer. Thus the meaning of the field E_c is now clear. It separates the high-field region with no field distortion from the low-field region with significant field distortion. The above implies that the field decreases rapidly once the critical field is reached.

Conclusions

The rapid onset of field distortion can be seen in Fig. 1 where the normalized field e is plotted across the i-layer for different values of e_s for the case of equal electron and hole mobilities ($a = 1$). It is apparent from the figure that once the average field drops below the critical value, E_c, the field in the cell distorts significantly from the average value V/L. The three curves are for a change of only 30% in V/L.

Fig. 2 shows the more representative situation for a-Si:H, where $a = 0.1$.[9] The small hole-to-electron-mobility ratio moves the minimum in the electric field close to the n-i interface. However, it also has the effect of making the onset of the field distortion more gradual. This can be seen by the large range of surface fields represented in the figure. Even though the field, e_s, at the n-i interface decreases, the field at the p-i interface increases. This will have an important consequence if recombination of

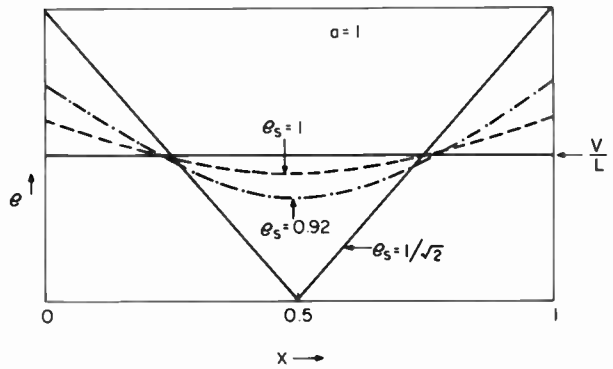


Fig. 1—Normalized electric field as a function of normalized distance across the i-layer for $a = 1$. The average field, V/L, is the solid horizontal line in the center of the graph. The parameter labeling each curve is e_s. The field goes to 0 at the center of the cell when $e_s = 1/\sqrt{2}$.

minority carriers at an interface is important, because diffusion of the minority carriers toward the interface increases as the field decreases.

Of course the field distortions cannot be as large as shown in the figure, because recombination and diffusion, which have been neglected, will tend to ameliorate the field distortion. Both these effects will become important because the carrier densities and their gradients increase as the field decreases. This is shown in Fig. 3 where the normalized hole density is plotted as a function of x for $e_s = 2$ and $e_s = 0.1$. When $e_s \gg 1$, the field is uniform and p increases linearly with x reaching the maximum value of 1 at $x = 1$. However, for small e_s, when the field decreases to nearly 0 at $x = 0.1$, there is a large increase in p and dp/dx.

It is useful to have a criterion to determine when the field distortion will be significant. One such criterion would be when the minimum field has decreased to zero. Of course the field would not actually drop to zero because, at low fields, recombination will become important as well as

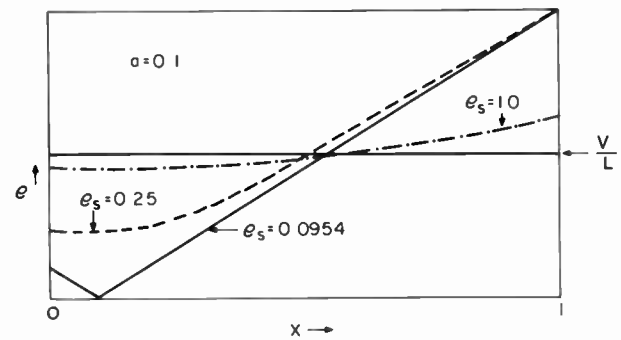


Fig. 2—Same plot as Fig. 1. Here $a = 0.1$ and the minimum $e_s = 0.0954$.

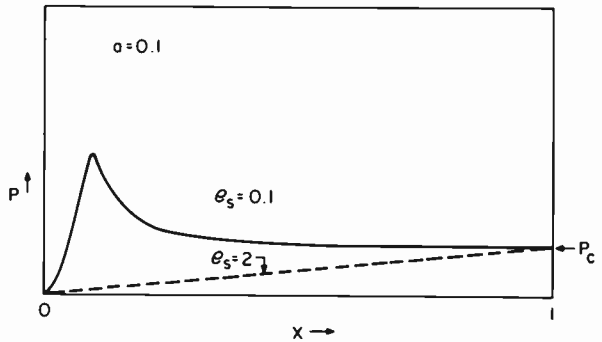


Fig. 3—Normalized hole density plotted as a function of normalized distance across the i-layer for $e_s = 2$ and $e_s = 0.1$ for the condition $a = 0.1$. The hole density is P_c at $x = 1$. The electric field is zero in the i-layer at $e_s = 0.0954$.

diffusion and remove the singularity. However, setting the minimum field to zero gives the condition that $E_c = 2V/L$, which gives meaning to the normalizing field E_c. For a-Si:H the condition becomes

$$L < (4V^2 \epsilon \mu_P / I_{sc})^{1/3}, \qquad [6]$$

where eGL has been replaced by the short circuit current, I_{sc}. For a typical a-Si:H solar cell, $I_{sc} = 0.012$ a, $V = 1$ V, and $\mu_P = 0.1$ cm^2/V-s.[9] For these conditions, L must be less than about 1 μm to avoid significant field distortion. Actually this is just the region where solar cell efficiency begins to decrease as a function of thickness.[10]

References:

[1] V. Dalal, "Analysis of Amorphous Silicon Solar Cells," *Solar Cells*, **2**, p. 261 (1980).

[2] T. Tiedje, C. R. Wronski, B. Abeles, and J. M. Cebulka, "Electron Transport in Hydrogenated Amorphous Silicon: Drift Mobility and Junction Capacitance," *Solar Cells*, **2**, p. 301 (1980).

[3] A Madan, "Localized State Distribution in Amorphous-Silicon Based Alloys Using the Field Effect Technique," *Solar Cells*, **2**, p. 277 (1980).

[4] B. T. Debney, "Model for Amorphous Silicon Solar Cells," *Solid State Electron Devices*, **2**, S-15 (1978).

[5] R. Williams and R. S. Crandall, "Carrier Generation, Recombination, and Transport in Amorphous Silicon Solar Cells," *RCA Rev.*, **40**, p. 371 (1979).

[6] A. Madan and P. G. Lecomber, *J. Non-Cryst. Solids*, **20**, p. 239 (1976).

[7] R. S. Crandall, "A Comparison of p-i-n and Schottky-Barrier Hydrogenated Amorphous Silicon, a-Si:H, Solar Cells," *RCA Rev.*, **42**, p. 441, Sept. 1981 (this issue).

[8] D. E. Carlson, R. S. Crandall, J. Dresner, B. Goldstein, J. J. Hanak, J. I. Pankove, H. Schade, D. L. Staebler, H. A. Weakliem, and R. Williams, Quart. Report No. 3, SERI/PR-0-8254-3.

9 J. Dresner, private communication.

[10] Y. Hamakawa, H. Okamoto, and Y. Nitta, "Optimum Design and Device Physics of the Horizontally Multilayered High Voltage Solar Cells Produced by Plasma Deposited Amorphous Silicon," 14th IEEE Photovoltaic Specialists Conf., San Diego, CA, 7 Jan. 1980, p. 1074.

Computer Simulation of Horizontal Transient Response of the NTSC Color-TV System

S. S. Perlman

RCA Laboratories, Princeton, NJ 08540

Abstract—A computer simulation program to faithfully calculate the overall horizontal transient response of an NTSC color TV system from light-image input to light-image output is described. The program processes two horizontal lines of picture data in either the time domain or the corresponding, Fourier transformed, frequency domain. Electronic processing, linear and nonlinear, is directly simulated in a serial fashion as the signal passes through the entire system. Program results are presented as software generated plots and tables or displayed subjectively on a TV monitor. Microprocessor controlled video hardware converts the digital software program results into corresponding sets of analog RGB signals that drive the kine inputs of the display monitor. A split-screen capability permits easy simultaneous subjective comparisons of up to four different program results corresponding to changes made anywhere within the system.

1. Introduction

The quality of a color television system is frequently measured on the basis of its response to a variety of test patterns. Measurements of horizontal transient response for standard patterns like the composite, multiburst or color-bar test signals, have long been recognized as suitable vehicles for evaluation.* Compuer-aided evaluations for certain luminance (black-and-white) transitions have been employed successfully. The effects of various combinations of vestigial sideband (VSB) filters, FCC predistortion filters and receiver RF/IF processing were demon-

* These test signals are often included during the vertical interval, lines 18 and 19, of a standard NTSC transmission.

strated many years ago to greatly influence the errors and distortions in the received luminance waveshapes.[1]

The Light-to-Light (L-to-L) programs extend the computer-model concept to color, in addition to the luminance signals; they also simulate directly and trace these signals as they pass through a complete NTSC television system. Vertically equivalent TV signals are treated in the simulation; software-generated plots can be directly compared to experimental oscilloscope traces at corresponding locations anywhere within the system. Signals are traced from the camera's light-image input to the receiver's light-image output. A special display capability permits subjective viewing of the simulated responses in order to enhance evaluation.

The L-to-L programs were developed as an aid for system and circuit designers in their quest to analyze the overall performance effects of signal processing changes made anywhere within the TV systems. Complicated design problems, like those related to chroma-quadrature or color-edge distortion, for example, can now be directly evaluated via computer-simulation techniques without the need to construct complicated test fixtures.

2. Programming Techniques

A television system can conveniently be divided into three basic portions: the camera, the transmitter, and the receiver. The L-to-L computer simulation treats each of these independently and combines them with other portions, as shown in Fig. 1. The programs simulate real TV-like signals and process them in a serial fashion as they pass through the system from a light-image input to a light-image output. Each portion of the system is divided into a series of programmed boxes, each containing appropriate mathematical models to represent the actual signal processing involved. Both linear and nonlinear processing are treated.

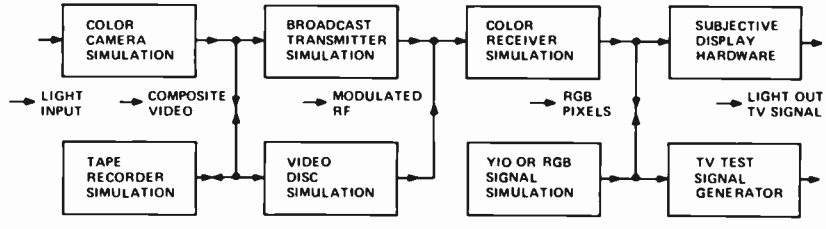


Fig. 1.—Block diagram for the light-to-light computer simulation.

Software-generated plots and data tables are designed to compare with experimentally obtainable oscilloscope traces or data anywhere within the system.

Essentially, all of the processings in the camera, transmitter, and receiver portions of the system have been successfully programmed. Simulations for the tape recorder and VideoDisc portions are currently under consideration and not yet fully programmed. The subjective display hardware and the test-signal generator are fully operational. These portions of the simulation convert the red-green-blue (RGB) or composite video signals into sampled digital picture-element (pixel) data and then to equivalent analog TV signals in an NTSC format.

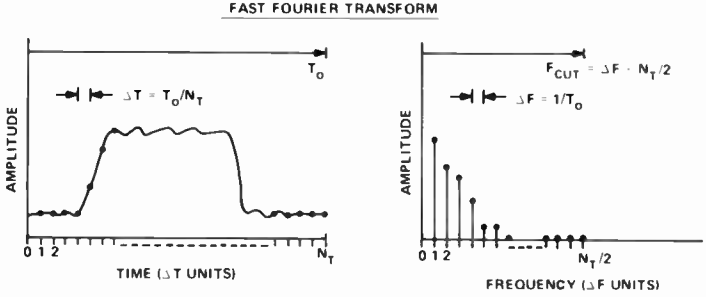
Also included as part of the L-to-L simulation is a special portion designed to aid in research related to the evaluation of psychological effects or human visual perception of color transitions. Test patterns with arbitrarily defined hue saturation and luminance transitions can be displayed and used subjectively to evaluate viewer preferences.

Output signals from each portion of the simulation are designed to be totally compatible and acceptable as input signals to the next or other portions. A special provision is provided via a jump-and-alter subroutine which allows the user to change signals and then to jump to any other box location within the program or to go to the start of another program for a different portion of the system. Possible alter options include saving a signal on disc file, inputting a new signal from a disc file, changing domain (time or frequency) of a signal, applying an MTF (filter) to the signal, adding two signals, multiplying two signals, copying another signal, combing-out odd or even harmonics from a signal, amplitude modulating with a carrier, adding impulse noise, applying nonlinear clipping, and plotting resulting signals on a graphic terminal[2] or flat-bed plotter.[3] The jump routine provides an easy method for cycling through one of the boxes in the program in order to evaluate effects of various design changes.

3. Programming Limitations

The L-to-L simulation programs have been written in Fortran IV language using the CMS interactive mode on an IBM/370 computer. The simulated TV signals are each stored as a pair of arrays (real and imaginary vector parts). A trade-off between array size and sample rate or cutoff (Nyquist) frequency had to be established. Fig. 2 reviews some fundamentals of the Fast Fourier Transform (FFT) used to convert between the time and frequency domain representations of the signal. Variables of interest are

T_o = Time Window (or periodicity),
N_T = Number of Samples (or array size),



N_T	T_o 63.555 μs = 1 H LINE			T_o = 127.11 μs = 2 H LINES			T_o = 254.22 μs = 4 H LINES		
	ΔT (ns)	ΔF (kHz)	F_{CUT} (MHz)	ΔT (ns)	ΔF (kHz)	F_{CUT} (MHz)	ΔT (ns)	ΔF (kHz)	F_{CUT} (MHz)
1024	62.07	15.73	8.06	124.1	7.87	4.03	248.3	3.93	2.01
2048	31.03		16.11	62.07		8.06	124.1		4.03
4096	15.52		32.22	31.03		16.11	62.07		8.06
8192	7.76		64.45	15.52		32.22	31.03		16.11
16384	3.88		128.9	7.76		64.45	15.52		32.22

Fig. 2—Fundamental parameters for the Fast Fourier Transform. Only the signal periodicity, T_o, and the number of data samples, N_T, need be chosen to get the required bandwidth, F_{cut}, and the proper harmonic content, ΔF.

F_{CUT} = Highest Harmonic Frequency (or Nyquist frequency),
ΔF = Lowest Harmonic Frequency,
ΔT = Sample Time.

These are interrelated by the equations in the figures such that only two, T_o and N_T, need be chosen. In addition, the FFT algorithm requires N_T to be an integer power of two.

The best choice for the L-to-L programs is to make T_o equal two horizontal TV lines (2*63.55 μsec $\equiv 2H$) and $N_T = 2^{12} = 4096$. For these values, $F_{CUT} = 16.11$ MHz and $\Delta F = 15.734/2$ kHz (half the horizontal line rate) and $\Delta T = 0.03101$ μsec. Lower N_T values lead to inadequate F_{CUT} or bandwidth while lower T_o leads to improper harmonic content to express a composite NTSC video signal. Proper harmonic content is maintained, provided T_o is expanded in multiples of $2H$ lines; however, larger T_o values were considered impractical due to the corresponding larger array sizes needed. For example, to represent a four-field stationary TV image, $T_o = 1050H$ is required with a corresponding array size, $N_T = 2^{21} = 2{,}097{,}152$ to keep similar bandwidth. This little exercise serves to demonstrate the enormous quantity of information contained in a TV scene. Clearly, a compromise had to be established, and $T_o = 2H$ was selected as the periodicity of the simulated TV signals. This compromise restricts the simulation to vertically equivalent scenes, like

color-bar patterns, and to the analysis of horizontal transient response only.

Any time-domain TV signal can be simulated by sampling at the selected rate of 0.031 μsec or 2048 per H line. The data can then be entered into the computer and stored as the desired signal. When transformed to the frequency domain, a series of $N_T/2$ harmonics of one-half the horizontal line rate (15.734/2 kHz) will result. The sum of these waves with the corresponding amplitudes and phase angles will yield the original time-domain signal, provided the high-frequency content of the wave remains below the 16.11-MHz Nyquist frequency selected.[4]

Simulation of the various electronic processes encountered in the TV system can usually be accomplished by simple mathematical or computer manipulation of the individual time- or frequency-domain samples corresponding to the TV signal. For example, modulation-transfer functions (MTFs) or electronic filters can be applied by a frequency-domain-vector multiplication of the MTFs' amplitude and phase characteristics with the corresponding amplitude and phase properties of the signal. This process is generally easier than the equivalent time-domain convolution with the impulse response of the MTF. In the L-to-L programs, MTFs can be described as data sets of frequency amplitude and phase angle in the range of interest. Interpolation techniques are employed to evaluate the MTF at the 15.734/2 kHz harmonic frequency's need. The programs also accept MTFs expressed as poles and zeros, and calculate the amplitude and phase responses at the harmonic frequencies. Other electronic processes can be handled in a similar fashion. Generally speaking, nonlinear functions, like clipping, are performed by manipulation of time-domain signals.

4. Program Example

4.1 Camera Portion

Use of the L-to-L simulation begins with the choice of an input light pattern corresponding to an arbitratry color image. Horizontal $2H$ time-domain traces of the R, G, and B signals are selected. A choice of three built-in test signals is provided: EIA color bars, composite, and multiburst. Other signals have to be input from the terminal or a disc file. Processing then proceeds as outlined in Table 1.

As an example, consider a program run using a standard EIA color-bar test signal (100-percent saturated colors at 75-percent amplitude). A standard camera design, meeting NTSC specifications, is also assumed. After processing with appropriate MTFs and matrix elements, a composite video signal is obtained and saved at the end of the camera portion

Table 1—Summary of Camera Simulation

A.	Select light input time signals for R, G, and B.[a]
B.	FFT to frequency domain.[b]
C.	Apply optical MTF to R, G, and B.[b]
D.	Apply pick-up device MTF to R, G, and B.[b]
E.	FFT to time domain[b]
F.	Apply nonlinear gamma correction to R, G, and B.[b]
G.	Apply encoder matrix to form Y, I, and Q.
H.	FFT to frequency domain.
I.	Apply different MTF to Y, I, and Q.
J.	Modulate color subcarrier with I and Q.
K.	Form composite signal.
L.	FFT to time domain.
M.	Add SYNC to signal.
N.	Convert to IRE units.
O.	FFT to frequency domain.
P.	Save composite camera output signal.

[a] Choices include EIA color bars, composite and multiburst signals, or arbitrary R, G, and B signals stored in disc file.
[b] This step omitted for standard test pattern signals, like EIA color bars.

of the program. The frequency-domain representation of the signal is shown in Fig. 3. Both luma and chroma information are presented but can be separated easily by viewing only the even harmonics (luma) or the odd harmonics (chroma), as shown in Fig. 4. The time-domain equivalent of the camera output signal is given in Fig. 5. Only the first 25 μsec of the $2H$ trace is shown. Of course, the camera I and Q modu-

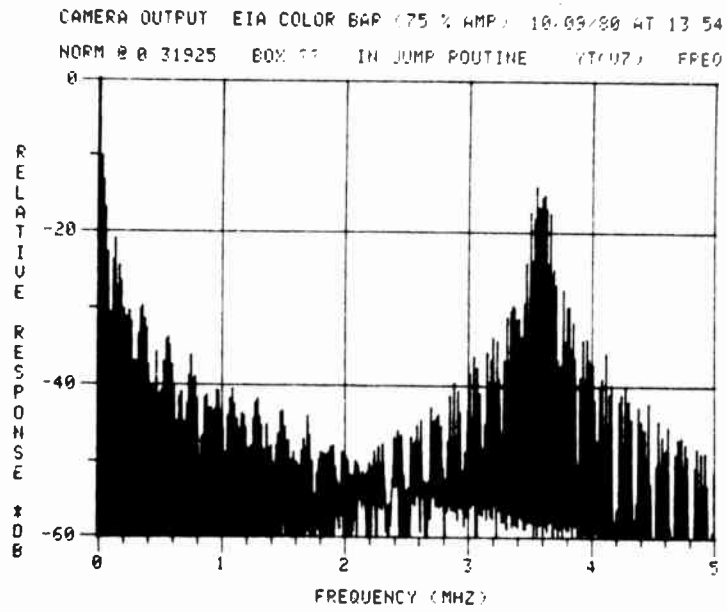


Fig. 3—Simulated camera output signal in the frequency domain for the example program run showing both odd and even harmonics of half the horizontal line rate.

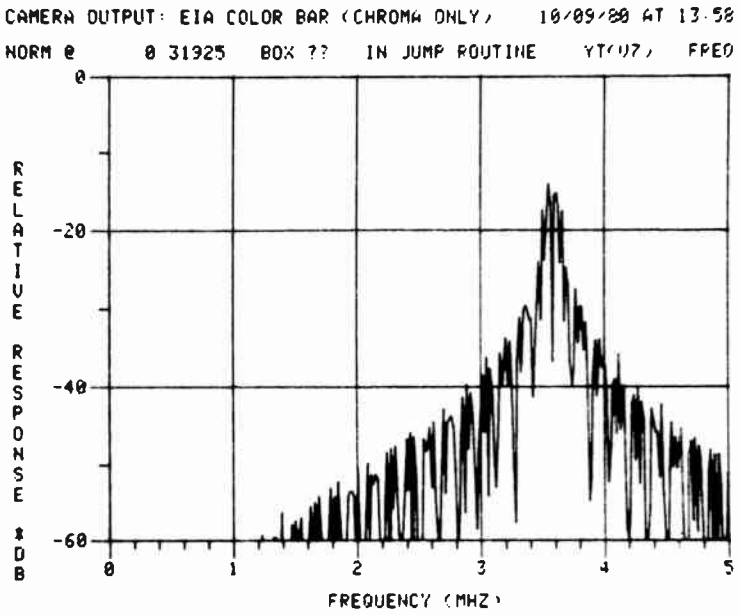


Fig. 4—Simulated camera output signal in frequency domain for the example program run showing only odd harmonics of half the horizontal line rate.

lators set the amplitudes and phases of the chroma signals for the different colors of the signal to their proper values relative to the burst.

The camera output signal represents a baseband input signal that can be used in essentially any of the other portions of the L-to-L programs.

4.2 Transmitter Portion

The program run continues with the assumption of a standard NTSC transmitter (Channel 4 with picture carrier at 67.25 MHz). The processing is outlined in Table 2. Again, appropriate MTF characteristics and process variables are selected and the signal leaving the transmitter is as shown in Fig. 6. The spectrum actually shown, however, has the transmitter output signal (66 to 72 MHz) plus two adjacent channel signals (at −10dB) that were added in the beginning of the receiver portion of the program. The time-domain equivalent for the modulation envelope of the transmitter output signal is complicated by the presence of the FM-modulated sound carrier at 67.25 + 4.5 MHz. On location, the signal is demodulated with special equipment and viewed directly on an oscilloscope. A comparable capability is provided in the L-to-L programs by a simulation of the response of an ideal demodulator. The RF/IF section provides flat-band response plus the proper Nyquist slope

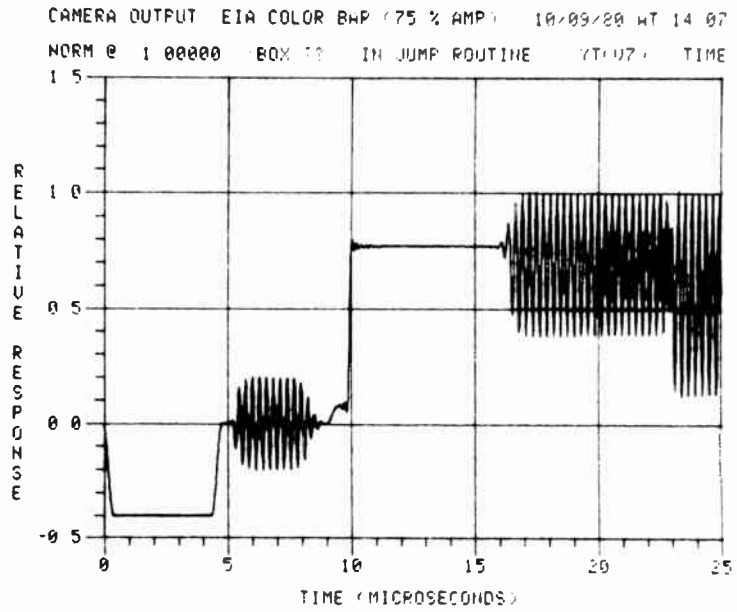


Fig. 5.—Simulated camera output signal in the time domain for the example program run showing only the first 25 μsec of the first horizontal line.

region. It also provides equalization for the predistortion and rejection characteristics for co-channel sound and adjacent-channel signals. Fig. 7 shows the resultant demodulated and synchronously detected transmitter output signal. Consumer TV receivers need not match such ideal front-end characteristics. The cost of implementation makes these designs impractical. Generally, less than ideal characteristics are tolerated with an attempt made during baseband processing to correct for any deficiency.

Table 2—Summary of Transmitter Simulation

A.	Input composite baseband signal.
B.	Apply predistortion MTF.
C.	Apply phase equalization.[a]
D.	Apply low-pass MTF.
E.	FFT to time domain.
F.	Adjust video gain for % modulation.
G.	Modulate RF carrier.
H.	FFT to frequency domain.
I.	Apply power amplifier gain and MTF.
J.	Apply vestigial sideband MTF.
K.	Add frequency-modulated sound signal (single tone).
L.	Save RF transmitter output signal.[b]

[a] Simulation of RCA TTS-1B Delay Equalizer System can be used. User selects desired switch positions.
[b] Signal corresponds to the positive half of the modulation envelope.

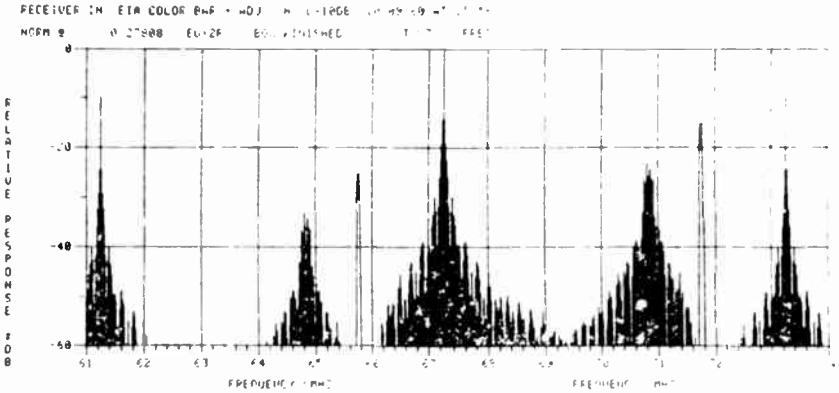


Fig. 6—Simulated transmitter output signal in frequency domain for the example program run showing main-channel information (67.25 MHz) plus adjacent-channel signals at −10 dB amplitude.

4.3 Receiver Portion

Our example program run now proceeds to the receiver portion. An outline of the processing included is given in Table 3. We shall assume a typical consumer-TV receiver design. The front-end RF and IF sections are lumped-element designs. Envelope detection is employed, followed

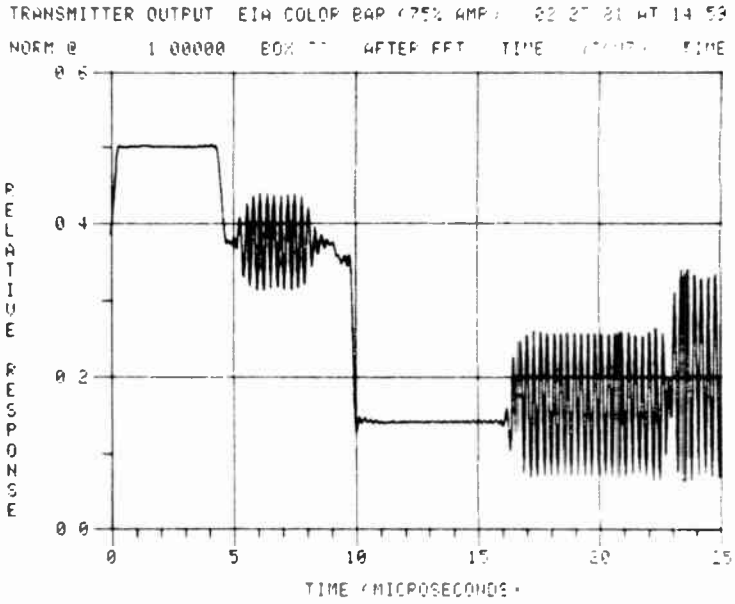


Fig. 7—Simulated transmitter output signal, in time domain, after demodulation and detection using high-quality studio equipment.

Table 3—Summary of Receiver Simulation

A.	Input RF signal.[a]
B.	Add reflections (echos).
C.	Add adjacent channel signals.
D.	Apply RF gain and MTFs.
E.	Modulate to IF frequencies.
F.	Apply MTFs and IF gain.
G.	Detect to baseband signal.[b]
H.	Apply detector MTF and nonlinearity.
I.	Apply sound, alias, and/or equalization MTFs.[c]
J.	FFT to time domain.
K.	Apply AGC and convert to IRE units.
L.	FFT to frequency domain.
M.	Apply CCD or power splitter.
N.	Apply 3.58 chroma trap to luma signal.[c]
O.	Apply luma-phase adjust.[c]
P.	Apply CCD-clock MTF.[c]
Q.	Add vertical restoration signals to luma.[c,d]
R.	Apply luma peaker and delay adjust.
S.	Apply chroma peaker and delay adjust.
T.	FFT to time domain.
U.	Apply automatic chroma gain adjust.
V.	Recover chroma-burst phase for demodulator.
W.	Demodulate chroma with desired gain and angles.
X.	Apply color-difference MTFs.
Y.	Apply decoder matrix for R, G, and B.
Z.	Apply kine-driver MTF to R, G, and B.
Za.	Save R, G, and B kine signals for display.
Zb.	FFT to time domain.
Zc.	Evaluate color-error information.[e]
Zd.	Adjust brightness and contrast.
Ze.	Apply kine gamma correction.
Zf.	Calculate light output versus distance.

[a] RF and IF processing can be bypassed if input is composite baseband signal.
[b] Envelope or synchronous detection available.
[c] Omit if inappropriate for design.
[d] Vertical restoration in the L-to-L Program introduces only chroma noise (dot crawl components) into the luma signal.
[e] Quantitative numbers can be calculated to express color errors experienced during a color transition.[8]

by a power splitter, and luma/chroma separation filters. Chroma-demodulation angles and gain values are designed to match the specific primary phosphor colors of the kinescope.*

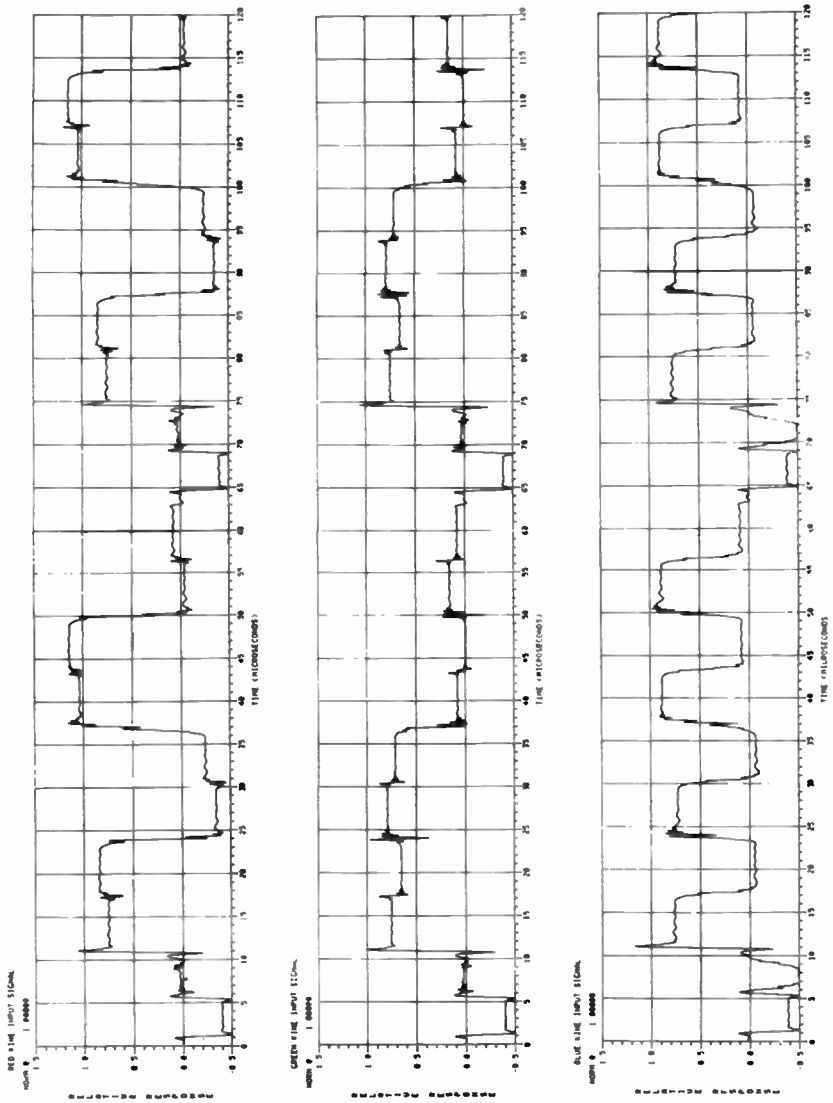


Fig. 8—Simulated RGB kine signals in the time domain for the example program run showing the entire two horizontal lines considered.

* For example, "field" phosphors are used in RCA kinescopes. These differ from the NTSC standard phosphors, mainly in the color of the green phosphor.

When the simulation is performed with the receiver controls (like tint, chroma, peaking, etc.) set at their nominal values, the resultant RGB kine-input signals are as shown in Fig. 8. A careful examination of these waveforms demonstrates the presence of various effects typically observed in experimental measurements made on such receiver designs. For example, we see (1) dot crawl, or the presence of odd harmonics in the RGB signals, (2) luma peaking, or presence of preshoot and overshoot on luma transitions, (3) color-matrix adjustments, or the change in primary color amplitudes required to obtain different saturated colors. Although software-generated plots contain all of the important information necessary to evaluate horizontal transient response of a TV system, the task is greatly simplified by the use of the subjective display capability of the simulation.

4.4 Subjective Display Portion

Since the L-to-L signals are direct simulations for two horizontal lines of video information, they can be converted to equivalent analog waveshapes and injected into the appropriate line positions to form a four-field interlaced raster corresponding to a stationary vertically equivalent TV scene. This is possible for both composite video signals, like those leaving the camera portion, or RGB video signal leaving the receiver.

Hardware, operating at video rates, had to be designed and constructed to perform the D/A conversions of the software signals and to properly insert these waveshapes into an NTSC TV signal. A block diagram of the subjective display hardware is presented in Fig. 9. Pixel values are obtained by software-interpolation techniques using the chosen video-display rate of 14.32 MHz (four times the color subcarrier). This rate is generally accepted as a standard video rate in NTSC broadcast studio equipment. The pixels are stored as 8-bit integers with a total of 1820 needed for each $2H$-long signal. Once the pixel values have been computed, they are transferred via a modem to the display microprocessor and stored on a floppy disc for later use.[5]

In order to display RGB signals, 3 times 1820 pixels—or roughly 6K of RAM—is needed. The microprocessor memory map was designed to contain 24K of high-speed (video rate) RAM. Thus, four complete sets of RGB pixels can be stored and accessed by the microprocessor and the video-control circuits. Each set of RGB pixels can correspond to a different program run and, therefore, to different design changes in the TV system.

Control circuits were designed to simultaneously address corresponding RGB pixel values at the video rate and to route this data to the

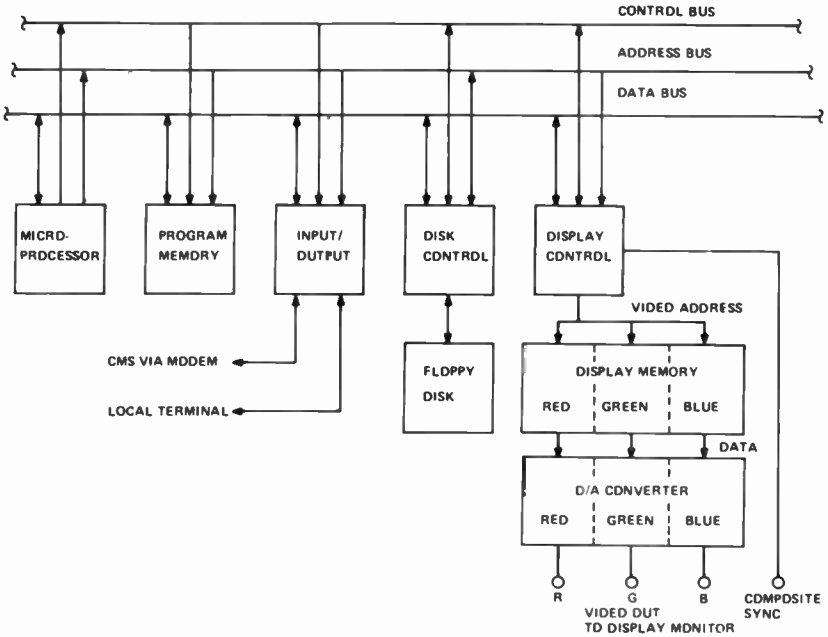


Fig. 9—Block diagram of the subjective display hardware.

three digital-to-analog (D/A) converters (only one D/A converter is needed when composite video test signals are being generated). A sync generator chip provides the basic timing for a standard NTSC TV signal.[6] Microprocessor software was developed to provide trigger information for the control circuits so that pixel data could be properly inserted and interlaced into the NTSC format. The processor is operated in an interrupt mode or on a time-shared basis so that normal processor functions can be performed even when the display is turned on. Since sufficient RAM is provided to store four sets of pixels, it is possible to divide the TV raster into bands and simultaneously display different program results.

The RGB signals from our test run (Fig. 8) were processed as outlined above and the resulting TV signals were displayed on an RGB monitor. An RGB monitor is needed that has broadband kine inputs (greater than the receiver design being simulated). The kinescope should be the same as, or at least have the same phosphor colors as, the receiver being simulated. Fig. 10 shows the monitor screen.[7] Band 1, at the top of the screen, represents the scene for our test run of the program; this corresponds to the TV system with the camera, transmitter, and receiver controls set to nominal design values. The other bands correspond to other program runs where everything was kept constant except: Band 2, chroma gain

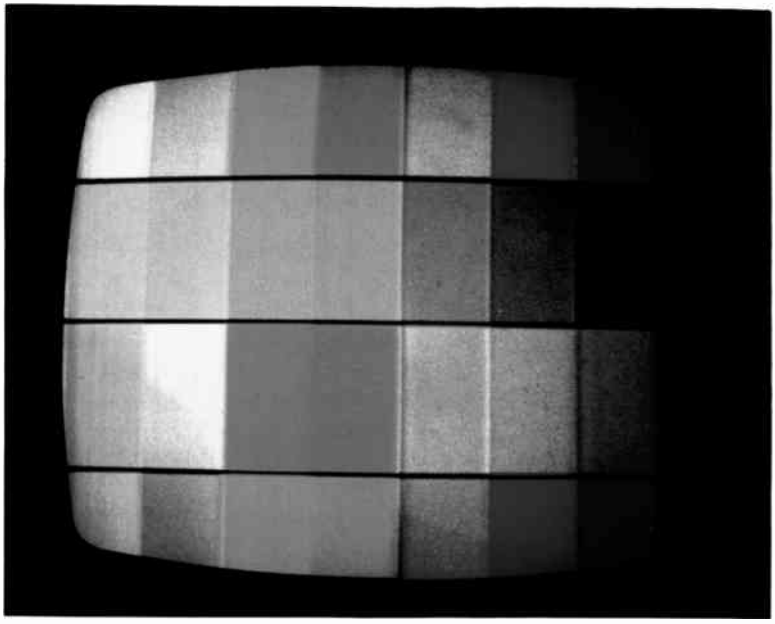


Fig. 10—The subjective display portion of the light-to-light computer simulation for the horizontal transient response of an NTSC color television system. Band 1 at the top shows results of the program run and subsequent bands shows results of runs with different changes in receiver design parameters.

in the receiver reduced a factor of ten below nominal; Band 3, tint control in the receiver adjusted so that chroma demodulation was +30 degrees from nominal; and Band 4, tint control in the receiver adjusted so that chroma demodulation was at −30 degrees from nominal. These other program examples correspond to rather obvious changes in the TV system, and are presented merely to demonstrate the effectiveness of the subjective display capability. Analyses of these simulated RGB waveshapes without the use of the subjective display would be much more difficult and less satisfactory. Of course, the L-to-L simulation programs are intended to treat the more interesting cases where results of system design changes are not immediately obvious or even predictable without the aid of a simulation program.

5. Summary

A new computer tool for evaluations of horizontal transient response of the NTSC color television system is described and demonstrated. Vertically equivalent television signals are simulated and mathematically processed in a serial fashion to exactly copy actual linear and nonlinear

electronic processing in current camera, transmitter, and receiver designs. Color-bar signals with arbitrarily defined color transitions can be processed and results can be presented on a subjective TV monitor display.

Simultaneous display and evaluation of up to four signals is possible. Each can correspond to a slight modification of any of the processing steps within the entire TV system. Consequently, selection of the best modification or circuit design change can be easily made without breadboarding all of the possibilities.

Comparable simulations for VideoDisc and videotape recorders and players are not fully operational, but, when available, should ultimately prove an excellent design aid. Subjective and simultaneous comparisons of TV systems using these components with others employing the normal camera, transmitter, and receiver designs should be possible. This unique display capability gives the L-to-L simulation program a special advantage over the more traditional evaluation techniques.

Acknowledgments

The author gratefully acknowledges many discussions with numerous researchers within RCA working on specific portions of the TV system. D. H. Pritchard and A. C. Schroeder deserve particular mention for being instrumental in suggesting the project and providing overall information on the NTSC TV system. Special thanks is also given to J. P. Bingham for suggesting the addition of a subjective display capability, and to H. G. Lewis, A. L. Greenberg, and H. M. Kern for their assistance in the design and construction of the necessary hardware.

References:

[1] J. P. Bingham and C. B. Patel, "Effects of Vestigial Sideband Passband Characteristics Upon Television Transient Response," *IEEE Trans. Broadcasting*, Vol. BC-22, No. 3, pp. 68–72 (Sept. 1976).

[2] Tektronix computer graphics terminals, such as Model 4006-1, display terminal.

[3] Tektronix computer graphics terminals, such as Model 4662, interactive digital plotter.

[4] R. W. Ramirez, "The Fast Fourier Transform's Errors Are Predictable, Therefore Manageable," *Electron.*, Vol. 47, No. 12, pp. 96–102 (June 13, 1974).

[5] Intel 8085 microprocessor plus SBC80 diskette system.

[6] Fairchild 3262B TV SYNC generator.

[7] Conrac color television monitor, Model RHM. The phosphors of this monitor do not agree with RCA "field" phosphors. They are closer to the NTSC standard phosphors.

[8] D. H. Pritchard and T. M. Wagner, "Color Signal Inphase and Quadrature Distortion Measurement and Evaluation," *RCA Rev.*, Vol. 38, pp. 31–32 (March 1977).

Patents Issued to RCA Inventors—Second Quarter 1981

April

L. Abbott Cancelling Cross Modulation in Two Color TV Signals Passed Through Non-Linear Path (4,264,919)
A. A. Ahmed Regulated Current Source Circuits (4,260,945)
A. A. Ahmed Current Amplifier With Regenerative Latch Switch (4,260,955)
F. Aschwanden Automatic Deviation Limit Control Circuit for Secam Encoders (4,263,609)
F. Aschwanden Automatic Chroma Gain Control Circuits Useful in Secam Coders (4,264,918)
B. Astle Amplitude Limiter With Automatic Duty Cycle Control for Use in a Phase-Locked Loop (4,263,565)
B. Astle Voltage Controlled Oscillator (4,263,567)
V. S. Ban Radiation Heated Reactor for Chemical Vapor Deposition on Substrates (4,263,872)
J. C. Bleazey Track Skipper for VideoDisc Player (4,262,174)
T. W. Branton Electron Beam Influencing Apparatus (4,261,017)
B. J. Chang Article Transfer Apparatus (4,264,254)
P. Datta and N. V. Desai Method for the Manufacture of Multi-Color Microlithographic Displays (4,263,386)
M. S. Deiss Search Type Tuning System (4,264,977)
A. G. Dingwall Voltage Comparator (4,262,221)
J. B. George RFI Shield for an IR Input Circuit (4,262,365)
W. G. Gibson, F. C. Liu, and M. W. Muterspaugh Television Signal Processing System (4,263,611)
W. G. Gibson and R. D. Thompson Comb Filter Equalization Circuit (4,263,612)
C. Grilletto and F. C. Vowinkel Measurement of a Gas Constituent by a Mass Spectrometer (4,260,886)
W. E. Ham Semiconductor-on-Insulator Device and Method for Its Manufacture (4,262,299)
F. G. Hammersand Magnetron Filament Assembly (4,264,843)
J. R. Harford Tempertaure Compensating Bias Circuit (4,260,956)
W. Hinn Burst Insertion Apparatus for Secam-PAL Transcoder (4,263,608)
W. Hinn Automatic Kinescope Biasing System (4,263,622)
G. W. Hunka Signal Detection System (4,263,555)
A. C. Ipri and D. W. Flatley Method of Making Radiation Resistant MOS Transistor (4,259,779)
A. C. Ipri Method of Manufacturing Short Channel MOS Devices (4,263,057)
R. H. Isham and C. J. Petrizio Oscillator Circuit (4,260,960)
P. A. Levine CCD Gain Control (4,262,217)
L. D. Miller and F. W. Peterson Pickup Tube Having Mesh Support Electrode Aligning Means (4,264,841)
F. R. Nyman and T. E Smith Information Record Stampers (4,262,875)
G. H. Olsen and A. V. Caflero Single-Crystal Hexaborides and Method of Preparation (4,260,525)
T. R. Pampalone and E. J. Gavalchin Positive Resist Medium and Method of Employing Same (4,262,073)
T. R. Pampalone, N. V. Desai, and E. S. Poliniak Positive Resist for Electron Beam and X-Ray Lithography and Method of Using Same (4,262,083)
T. R. Pampalone Method for the Manufacture of Multi-Color Microlithographic Displays (4,263,385)
D. Sakarya and J. B. George Regulator Arrangement Useful in an Infrared Remote Control Transmitter (4,264,896)
D. Sakarya Drive Circuit for an Infrared Remote Control Transmitter (4,264,982)
O. H. Schade, Jr. Differential Amplifier Circuit (30,587)
O. H. Schade, Jr. Bandgap Reference (4,263,519)
L. N. Schiff System for Limiting Intermodulation Distortion of Talkspurt Signals (4,262,355)
R. L. Shanley, 2nd, L. A. Harwood, and E. J. Wittmann Controlled Output Composite Keying Signal Generator for a Television Receiver (4,263,610)
S. Shwartzman Two Step Method of Cleaning Silicon Wafers (4,261,791)
W. W. Siekanowicz and J. R. Fields Modulator With Variable Launch Conditions for Multi-Electron Gun Display Devices (4,263,529)
S. A. Sieckler Low Distortion Signal Amplifier Arrangement (30,572)
G. E. Theriault Double Trapping of Adjacent Channel Sound (4,263,619)
M. Van Renssen and M. H. Wardell, Jr. Kinescope Screen Center Locator (4,261,013)
L. E. Waller and G. S. Zorbalas Dual Standard Color Framer (4,261,008)
C. C. Wang and R. F. Bates Method of Depositing an Abrasive Layer (4,260,647)
H. R. Warren Headwheel Servo Lock Verification with Stationary Head (4,263,625)

C. E. Weitzel and J. H. Scott, Jr. Planar Semiconductor Devices and Method of Making the Same (4,263,709)
F. S. Wendt High Frequency Ferroresonant Transformer (4,262,245)
C. F. Wheatley, Jr. Reference Voltage Circuit Using Nested Diode Means (4,260,946)
D. H. Willis Color Television Degaussing Circuit (4,262,232)
C. M. Wine Touch Switch Arrangement Useful in a Television Receiver (4,263,618)

May

A. A. Ahmed Relaxation Oscillator Having Switched Current Source (4,270,101)
A. R. Balaban and S. A. Steckler Differential Amplifier Current Repeater (4,266,245)
A. E. Bell Information Record (4,270,132)
D. W. Breithaupt Remote Control TV Subcarrier Phase Shifter System (4,266,197)
T. W. Burrus Stylus Lifting/Lowering Actuator With Improved Electromagnetic Motor (4,266,785)
B. L. Compton and P. B. Pierson Optical Scanner With Variable Scan Line Angle (4,265,524)
J. W. Daniel, Jr. Phaselocked Receiver With Orderwire Channel (4,270,221)
C. A. Deckert and G. L. Schnable Silicon Nitride and Silicon Oxide Etchant (4,269,654)
T. W. Edwards and R. S. Pennypacker Manufacture of Thinned Substrate Imagers (4,266,334)
E. V. Fitzke, D. B. Wenner, and M. J. Polak Temperature Insensitive Filter for Kinescope Envelopes (4,269,616)
L. A. Goodman Method for Manufacturing a Self-Aligned Contact in a Grooved Semiconductor Surface (4,268,537)
C. P. Jindra and J. H. Atherton Small Signal Memory System With Reference Signal (4,270,190)
I. Ladany Etching of Optical Fibers (4,265,699)
C. J. Petrizio and R. H. Isham Relaxation Oscillator (4,267,527)
E. S. Poliniak and N. V. Desai Method for Forming a Shallow Surface Relief Pattern in a Poly(Olefin Sulfone) Layer (4,267,257)
W. F. Reichert Method of Making a Schottky Barrier Field Effect Transistor (4,266,333)
R. L. Rodgers, 3rd Peak-Response Controller for Average-Responding Automatic Iris (4,268,866)
O. H. Schade, Jr. Operational Transconductance Amplifiers With Non-Linear Component Current Amplifiers (4,267,519)
A. Schwarzmann Phase Equalizer in Microwave Transmission Line (4,270,104)
W. W. Siekanowicz Electron Current Collector for Flat Panel Display Devices (4,266,159)
G. E. Thornberry Radio Frequency Interference Suppression Apparatus (4,267,528)
L. A. Torrington VideoDisc Caddy Having Disc Entrapment (4,266,784)
R. M. Unetich and W. Shibler Regulated Filament Supply for High-Power Tubes (4,267,487)
J. A. Weisbecker Display System (4,270,125)
C. T. Wu System for Ascertaining Magnetic Field Direction (4,267,640)

June

R. L. Angle Method for Making a Closed Gate MOS Transistor With Self-Aligned Contacts With Dual Passivation Layer (4,272,881)
R. L. Angle Method for Making a Closed Gate MOS Transistor With Self-Aligned Contacts (4,274,193)
S. Berkman, R. Metzl, R. E. Novak, and D. L. Patterson Heat Radiation Deflectors Within an EFG Crucible (4,271,129)
K. R. Bube Glazing Paste for Bonding a Metal Layer to a Ceramic Substrate (4,273,822)
P. Datta VideoDisc Processing (4,275,100)
R. H. Dawson and G. L. Schnable Passivating Composite for a Semiconductor Device Comprising a Silicon Nitride (Si_3N_4) Layer and Phosphosilicate Glass (PSG) Layer (4,273,805)
C. B. Dieterich PCM Detector (4,275,416)
R. M. Evans Method for Tuning a Filter Circuit (4,272,743)
F. C. Farmer, Jr. and D. P. Knight Precision Cathode Current Regulator (4,275,347)
A. W. Fisher Method of Laying Out an Integrated Circuit With Specific Alignment of the Collector Contact With the Emitter Region (4,272,882)
W. V. Fitzgerald, Jr. Service Switch Apparatus (4,272,777)
J. S. Fuhrer Defect Compensation for Color Television (4,272,785)
R. A. Gange and F. J. Marlowe System for Compensating for Cathode Variations in Display Devices Utilizing Line Cathodes (4,271,377)
J. J. Gibson VideoDisc Playback Apparatus With Non-Linear Aperture Correction (4,272,786)
A. P. Gilson and B. W. Siryj Protective Cartridge for Optical Discs (4,273,342)

M. B. Goldman and G. I. Morton A-C Rectifier Circuit for Powering Monolithic Integrated Circuits (4,276,592)
J. J. Hanak Tandem Junction Amorphous Silicon Solar Cells (4,272,641)
J. R. Harford Gain Controlled Amplifier Using a PIN Diode (4,275,362)
L. A. Harwood and R. L. Shanley, 2nd Color-Difference Signal Processing Circuits (4,272,778)
C. H. Hubert Orientation of Momentum Stabilized Vehicles (4,275,861)
A. C. Ipri CMOS SOS With Narrow Ring Shaped P Silicon Gate Common to Both Devices (4,271,422)
W. K. Knapp Resettable Bistable Circuit (4,275,316)
K. Knop Apparatus and Method for Measuring the Ratio of Two Signals (4,272,197)
A. L. Leidich Amplifier Circuit (4,271,394)
M. E. Malchow Differential FM Detector With Series Tuned Filter (4,272,726)
E. A. Miller Centering Support for a Rotatable Wafer Support Susceptor (4,275,282)
A. M. Morrell Color Picture Tube With Screen Having Light Absorbing Areas (4,271,247)
R. P. Parker Circuit for Inhibiting Radio Frequency Interference in a Television Receiver (4,276,566)
K. J. Phillips Nutation Damping in a Dual-Spin Spacecraft (4,272,045)
O. H. Schade, Jr. Switched Current Source for Current Limiting Complementary Symmetry Inverter (4,274,014)
A. Schwarzmann Phase Shifter (4,275,366)
B. W. Siryj and A. P. Gilson Optical Disc Player System (4,271,489)
B. W. Siryj and L. D. Moore Apparatus for Converting Rotary Motion to Linear Motion (4,274,294)
G. E. Theriault Saw Filter Preamplifier (4,271,433)
L. A. Torrington VideoDisc Player Having Record Extracting Mechanism (4,272,083)
C. E. Tracy and W. Kern Bulk Glass Having Improved Properties (4,273,828)
C. C. Wang, L. Ekstrom, T. C. Lausman, and H. Wielicki VideoDisc Lubricants (4,275,101)
C. M. Wine Receiver With a Channel Swapping Apparatus (4,271,532)
H. A. Wittlinger Differential-Input Amplifier Circuit (4,272,728)

AUTHORS

Jay J. Brandinger received a BEE from Cooper Union in 1951. In 1962 he received an MSEE and in 1968 a Ph.D. in Electrical Engineering from Rutgers University. He joined RCA Radio Research Laboratories in Riverhead, Long Island, New York, as a member of the technical staff in 1951. In 1959 he was appointed group head of Communications Research at RCA Laboratories in Princeton, New Jersey. From 1959 to 1974 he had responsiblity for communication systems, display and television systems research. In 1974 he was appointed Division Vice President, TV Engineering, responsible for the developoment and design of all RCA Consumer Electronics television products. Dr. Brandinger was named Division Vice President, RCA "SelectaVision" VideoDisc Operations in Indianapolis, Indiana, in 1979 with overall responsibility for player and disc products. In his most recent assignment he was given responsibility for rapidly expanding RCA's VideoDisc manufacturing capability as the new "SelectaVision" VideoDisc Division's Vice President and General Manager. He retains overall responsiblity for the RCA CED "SelectaVision" VideoDisc system.

Jon K. Clemens received a B.A. degree in Physics from Goshen College in June, 1960, a B.S. and M.S. degree in Electrical Engineering from MIT in June 1963, and a Ph.D. degree in Electrical Engineering from MIT in September, 1965. His Ph.D. thesis research was in optical character recognition for reading machines for the blind. He joined RCA Laboratories in 1965 to work on high density video recording for consumer use. During the years 1965 through 1970 he worked on various approaches to video disc systems and conceived the capacitive pick-up video disc system used by RCA today. The basic U.S. patent on this system entitled "Information Records and Recording/Playback Systems Therefore" was granted to him for this invention. During the next years his work included all aspects of system development, including disc format and signal encoding systems and the responsibility for specifying the RCA VideoDisc standards. In 1975 he was appointed Head, Signal System Research, with the primary responsiblity for developing the VideoDisc signal system for both mastering and player design, and has recently been appointed Director, VideoDisc Systems Research Laboratory. He has received three RCA Laboratories Outstanding Achievement Awards and the David Sarnoff Award for outstanding technical achievement for his work on the CED VideoDisc System. He was also a recipient of the 1980 Edward Rhine Prize for major improvements in television systems. He is a member of the IEEE, Eta Kappa Nu, Sigma Xi, and Tau Beta Pi Societies.

Richard S. Crandall received a B.S. degree in physics from Cornell University in 1960, an M.S. in physics in 1962, and a Ph.D. in physics from the University of Illinois. He joined RCA Laboratories in 1964 and has worked in the fields of Electron and phonon transport in insulators and semiconductors, electron surface states on liquids, electrochromics, macro ions in solution, and is now engaged in studies of amorphous silicon solar cells. Dr. Crandall received an RCA Achievement Award for work on the electron-photon interaction and spent the year from 1971 to 1972 at the Zurich Laboratories of RCA. He is a member of The American Physical Society and the IEEE.

Marvin Freeling received the B.A. degree in Physics from New York University in 1953 and an M.S. degree in Applied Physics from Harvard University in 1954. In 1954 he joined the Technical Staff at Bell Telephone Laboratories where he worked on the design and development of reflex klystrons for radio-relay systems. Mr. Freeling joined RCA in 1957, engaged in the design and development of traveling-wave tubes for electronic countermeasure systems and for communication systems. In 1963, he received the RCA Electronic Components and Devices Engineering Achievement Award for participation in the design and development of the traveling-wave tube for Relay, RCA's first communication satellite. Mr. Freeling later joined RCA Americom in 1974. He has since been engaged in communications systems studies and design for RCA's Satcom satellites. He is also responsible for technical support of RCA Americom's legal department in dealings with governmental regulatory agencies, such as the FCC. Mr. Freeling has received a U.S. Patent as the co-inventor of recirculation memory system having a diplexed feedback loop.

Sheng T. Hsu received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1958, the M.S.E.E. degree from National Chiao-Tung University, Hsienchu, Taiwan, in 1960 and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 1966. From 1966 to 1970, he was with Fairchild Semiconductor Research and Development Laboratory, Palo Alto, California, as a Member of the Technical Staff. From 1970 to 1972, he was an Assistant Professor of the Department of Electrical Engineering, University of Manitoba, Winnipeg, Man., Canada. He is now a staff member of RCA Laboratories, Princeton, N.J., working on semiconductor devices and integrated circuit technologies.

Richard C. Palmer received the BEE degree from the University of Virginia in 1943 and did subsequent graduate work in Physics at the Stevens Institute of Technology. After one year with the General Electric Co., Schenectady, NY and two years with the Tennessee-Eastman Corp., Oak Ridge, Tenn., he joined the Allen B. DuMont Laboratories where he worked in the fields of television studio equipment, television cameras, color television, phototubes, and television display devices. He joined the RCA Laboratories as a Member of the Technical Staff in 1960, and has pursued research in the fields of television displays and VideoDisc systems.

Mr. Palmer served on Panel 12 of the National Television Systems Committee and on various committees of the IRE and has been granted over 20 patents. He has received three RCA Laboratories Outstanding Achievement Awards and is a member of IEEE, Sigma Xi, and Tau Beta Pi Societies.

Stuart S. Perlman joined RCA Laboratories in 1962 as an Engineering Consultant and a year later as a full-time Member of Technical Staff. Dr. Perlman's research efforts have included work on semiconductor heterojunctions, piezoferroelectric analog memory devices, and microsonic (SAW) TV-IF filters; for the past few years, he has worked on television system analysis employing computer simulation techniques.

Dalton H. Pritchard received the BSEE degree in Electronics in 1943 from Mississippi State University. In 1946, Mr. Pritchard joined RCA Laboratories as a Member of the Technical Staff at Princeton, New Jersey. There he has been engaged, for a number of years, in research in many aspects of color television systems development, receivers, color kinescopes, transmitting encoders, cameras, and magnetic recording of TV. This work included the planning and testing of systems and circuits proposed for adoption by the National Television Systems Committee (NTSC). At present, Mr. Pritchard is engaged in the development of video processing circuitry for color TV receivers in the areas of colorimetry, decorder matrix methods, and the development of analog techniques employing CCD devices for TV applications.

In the period between 1952 and 1980, Mr. Pritchard received nine RCA Laboratories Achievement Awards. He has been granted 36 patents in the fields of Color Television and Information Display. In 1975, he was appointed Fellow of the Technical Staff, RCA Laboratories. In June, 1977, he was awarded the Vladimir K. Zworykin Award for "Significant Contributions to Color Television Technology" at the IEEE Consumer Electronics Spring Conference in Chicago. In September, 1981, he was among nine co-recipients of the international Eduard Rhein Prize 1980, presented in Berlin. He was cited for numerous contributions in the field of video techniques and particularly as a leader in the development of the Dynamic Detail Processor employed in RCA ColorTrak receivers. He is a Fellow of the IEEE, Fellow of SID, and is a member of Sigma Xi, Tau Beta Pi, and Kappa Mu Epsilon.

Michael D. Ross received a B.S. and M.S. degree from Drexel University in 1957 and 1962, respectively. In 1970 he received an M.S. degree in Engineering Management from Drexel University. He joined the RCA Government and Commercial Systems organization in 1957 where he worked on various television projects for the government and NASA. In 1969 he joined RCA Laboratories, Princeton, New Jersey, to work on the RCA VideoDisc system. He has received two RCA Laboratories Outstanding Achievement Awards and has been granted several patents in the areas of signal encoding and player circuitry for the RCA VideoDisc System.

Leonard Schiff received the B.S.E.E. degree, magna cum laude, from City College of New York in 1960, the M.S.E.E. degree from New York University in 1962, and the Ph.D. degree in Electrical Engineering from the Polytechnic Institute of Brooklyn, New York, in 1968. From 1960 to 1967 he was employed by Bell Laboratories. There he participated in the development of the #1 ESS Switching System, including device development of the "ferreed" cross- point, design of the central processor, and support of Western Electric personnel in placing the system on-line. Dr. Schiff later contributed to the design of the experimental #1 ESS-ADF, a large store-and-forward message switch. His work was largely concerned with the design, analysis, and construction of the basic software architecture. On joining RCA laboratories in 1967, Dr. Schiff worked on the analysis of a variety of communications systems, both analog and digital, and on traffic and queueing systems. Many of these projects were in support of cellular mobile radio systems, especially techniques for making these systems more spectrally efficient and more efficient in their traffic-carrying capacity. His work on both mobile radio systems and automatic vehicle location systems has military as well as commercial applications.

Dr. Schiff is presently Head of the Communications Analysis Research group. His recent work has been in support of the development and operation of various satellite communications systems. He has been devising cost-effective methods of terrestrial interconnect and tradeoff studies for various methods of multiple access to the satellite. Dr. Schiff is a member of Eta Kappa Nu, Tau Beta Pi, Sigma Xi, and IEEE.