HUMAN VISUAL PERFORMANCE AND FLAT PANEL DISPLAY IMAGE QUALITY

LEVEL II ⑫

by

Harry L. Snyder, Ph.D.

Human Factors Laboratory
Virginia Polytechnic Institute
and State University
Blacksburg, VA   24061

DTIC
SELECTE
DEC 0 9 1980
E

July, 1980

AD A092685

DDC FILE COPY

BEST COPY AVAILABLE

80 12 08 020

HUMAN VISUAL PERFORMANCE AND FLAT PANEL

DISPLAY IMAGE QUALITY

by

Harry L. Snyder, Ph.D.

Human Factors Laboratory
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061

July, 1980

Accession For		
NTIS GRA&I		☒
DDC TAB		☐
Unannounced		
Justification		
By		
Distribution/		
Availability Codes		
Dist.	Avail and/or special	
A		

Prepared for:

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A092 685	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) HUMAN VISUAL PERFORMANCE AND FLAT PANEL DISPLAY IMAGE QUALITY.		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER HFL-80-1/ONR-80-1
7. AUTHOR(s) Harry L. Snyder		8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0238
9. PERFORMING ORGANIZATION NAME AND ADDRESS Human Factors Laboratory, Dept. of Industrial Engineering and Operations Research, Virginia Polytechnic Inst. & State Univ., Blacksburg, VA 24061		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 196-155
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Code 455 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE 07/80
		13. NUMBER OF PAGES 437 + xv
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Final		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

visual displays	display design
flat panel displays	visual performance
human factors	visual models
human engineering	contrast sensitivity function
image quality	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is the Final Technical Report of Task I: Human Engineering Survey and Analysis of the subject contract. The task is a survey of the pertinent visual performance, display system capability, and human engineering design requirements for flat panel visual displays, as applied to U.S. Navy Airborne, Shipborne, and Land-Based Systems. The report contains two application examples of the selection of flat panel displays for both airborne and ship-based information systems. It also surveys the current state of the art of flat panel display

DD FORM 1 JAN 73 1473   EDITION OF 1 NOV 65 IS OBSOLETE

technologies, and the manner by which existing visual performance and theory can be applied to the selection and evaluation of flat panel technologies for various applications. The flat panel technologies surveyed are light emitting diodes, electroluminescence, liquid crystal, electrochromic, electrophoretic, and gas (plasma) discharge. A uniform set of performance variables is used to compare the various technologies, both among themselves and with the traditional cathode ray tube. Similarly, these technologies are evaluated against design criteria and current models of image quality which relate human performance to display characteristics. Pertinent human visual performance data are presented as the basis for selecting and applying image quality models. Data gaps and needs are summarized at the end of the report.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

- xi -

Section 1

INTRODUCTION

During the past two decades, there has been a greatly increased amount of activity in research and development of new types of visual displays, many of which have been intended to replace classic incandescent and cathode ray tube (CRT) display devices and systems.

Ten years ago it was predicted by many in the display system business that these new, solid-state (or flat-panel) displays would indeed totally displace the CRT for most if not all applications. The flat-panel displays showed promise of greater ruggedness, finer detail or resolution, less power consumption, better geometric image stability, less image distortion, and greater compatibility with their associated electronic circuitry.

During these ensuing 10 years, some (but not all) of those claims have been met. Simultaneously, the CRT has been improved greatly on many of the same dimensions. As a result, the flat panel display technologies have been used in many applications where CRTs were formerly used, but the replacement of CRTs with other forms has been anything but universal.

Today there exists the same challenge from the solid
state display community. Further, the variety of available
solid state display types continues to increase, due largely
to continued emphasis on the research and development crit-
ical to the electronic interfacing and materials needed for
these devices. As a result, the systems engineer or human
factors engineer concerned only with selecting an appropri-
ate display device for a given application is faced with
many candidates and few criteria for valid selection. This
report is an attempt to provide useful, conveniently pack-
aged information for that systems or human factors engineer
faced with the display selection or evaluation task.

1.1 PURPOSE

The purpose of this report is to provide the critical back-
ground information that the systems or human factors engi-
neer needs to make an intelligent display device selection
for a given application. It is not intended as a designer's
manual, as a complete menu of required future research pro-
jects, nor as an authoritative sourcebook of display systems
engineering.

To make an intelligent and valid display device selec-
tion, or to specify appropriate criteria for display device
fabrication and selection, the human factors/systems engi-
neer requires a working knowledge of (1) the fundamental
physical concepts inherent in each display technology, (2)

functional relationships between display technology types and display parameters pertinent to the human operator's use of the display, and (3) the quantitative relationships between human performance in generic or typical tasks and functional characteristics of the display, irrespective of the display technology.

It is the purpose of this report to provide that information needed by the systems/human factors engineer and to illustrate how that information can be used to arrive at a valid display technology or device selection for any given application.

1.2 ORGANIZATION

Section 2 of this report describes a sequential analysis approach which the systems/human factors engineer can use to (1) select a display technology for a given application, and (2) estimate optimal design characteristics for that display. The approach is described generically at the beginning of Section 2, and then applied to two representative situations---the Navy aircraft cockpit and a shipboard command and control center. The recommended sequential analysis approach is referenced to subsequent sections of this report which provide backup quantitative data on the alternative display technologies, human visual performance data and models, and generic display requirements criteria.

Section 3 of the report presents an overview of current flat-panel display technologies. Each technology is described in terms of its basic physical concepts and materials; its electrical, photometric, and geometric characteristics; its inherent advantages and limitations; and a prognosis for its future development. For purposes of consistency, each technology is described by the same 13 parameters or categories. These parameters or categories are defined at the beginning of Section 3, which concludes with a comparative description of the technologies. Section 3 also includes, using the same 13 categories of description, summary characteristics of the CRT display, for the CRT remains the traditional standard against which all other display devices are compared.

Section 4 describes the spatial, temporal, and chromatic capabilities of the normal human visual system. The approach taken is that of the application of (1) current multichannel processor concepts of the visual system, and (2) linear systems analysis of both the display and the visual system to result in an analytic ability to directly compare the visual system's "requirements" with the information displayed. This analytic approach is developed in Section 5, which also summarizes selected human operator performance data pertinent to the design and evaluation of these flat panel display devices.

There are, unfortunately, substantial gaps in our knowledge of display design requirements for human operator optimal performance. The data gaps, or research needs, are noted in appropriate places throughout the report, as they become pertinent to various subjects of the report.

Finally, Section 6 summarizes the report, examining briefly the current and future flat panel display technologies, their ability to meet the display needs of the human operator, and the display and/or human performance research and development necessary to maximize the selective use of these flat panel display technologies.

Section 2

DESIGN APPROACH AND TECHNOLOGY SELECTION

This section of the report summarizes and gives two examples of a procedure by which the human factors or systems engineer can select one or more flat panel display technologies for a specific application. The procedure can also be used to evaluate specific display designs against system and operator requirements.

The procedure requires knowledge of the information to be displayed, the environment in which the display shall be used, the geometry of the display/observer work station, and any voltage/power constraints. In brief, the procedure takes the generic requirements of the display system, applies to those requirements design criteria from known parametric display/observer research (Section 5), refines the requirements based upon known human visual capabilities (Section 4), and compares the requirements against the range of characteristics of the various display technologies (Section 3). Thus, reference is made to both the display technology literature and the human performance literature in applying this procedure, which is described in greater detail in the following paragraphs.

2.1 GENERALIZED DESIGN APPROACH

The following steps permit the human factors/systems engineer to proceed from a set of display requirements to a selection (or evaluation) of a specific display technology or device. While each application and system will present unique considerations and problems, it is believed that modifications to this procedure, as implemented with reference to the backup data in Sections 3 through 5, will solve most of the selection/evaluation problems.

The overall flow chart of the procedure is illustrated in Figure 1. The individual steps are described as follows.

2.1.1 Display Functional Requirements Definition

The first step in any display selection/evaluation process is the specification of display requirements in terms of the information to be displayed, the environment in which the display will be used, and any particular constraints of relevance to the selection. This step will normally follow such systems engineering activities as mission analysis, mission requirements, function allocation (to operator versus machine), and display/control information requirements analysis. That is, previous activities are assumed to have defined the types of symbols, pictures, etc., to be displayed on the device under consideration.

In this step, it is then necessary to answer the following questions:
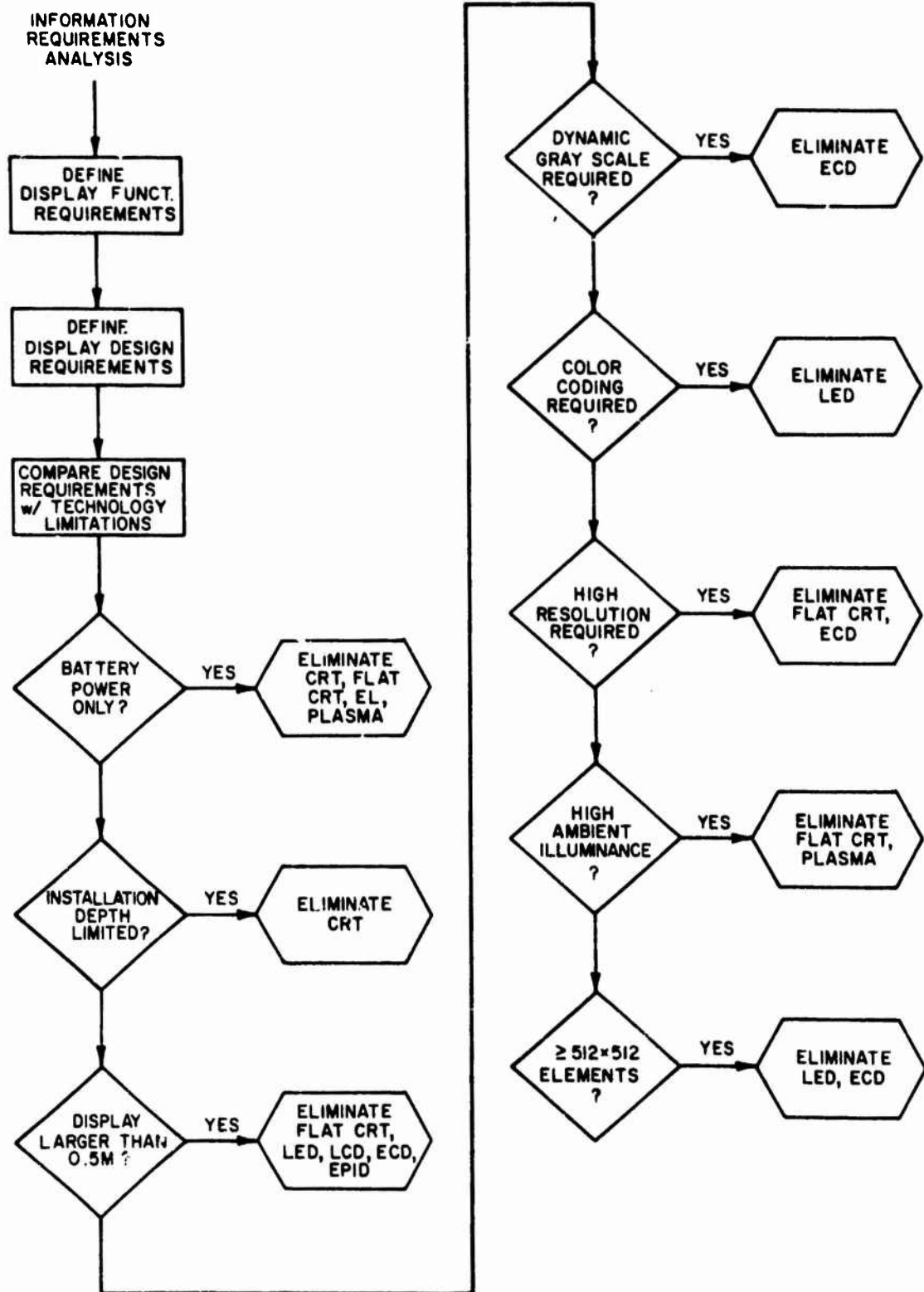
Figure 1: Flow chart for display technology selection

1. What is the display/observer geometry?

2. What environmental/power constraints exist?

3. What is the nature of the ambient illuminance?

4. What symbology must be displayed?

5. Is dynamic presentation required? If so, at what data rates?

6. Is the displayed information alphanumeric, vector-graphic, or pictorial, or some combination of these?

7. How much of this information must be presented simultaneously, and in what format?

Answers to these questions define the functional requirements of the display. These functional requirements must then be used to generate more specific design or performance requirements.

2.1.2   Design Requirements

The functional display requirements indicate what information is to be displayed, where, when, and how often. The design requirements, on the other hand, specify the exact design variables to which the hardware (and software) must conform to assure adequate operator performance. That is, the functional requirements are compared with our knowledge

- 9 -

of operator/display relationships to generate detailed design requirements for the display hardware.

The variables which must be specified in the design requirements, as a minimum, are listed below. Later sections of this report, indicated in parentheses with these variables, provide the empirical and/or theoretical basis for the specific level of each variable.

* Display size (5.1.5)

* Resolution (4.1, 5.1.1)

* Element Density (5.1.1)

* Element Size, Shape, Spacing (4.1, 5.1.2, 5.1.3)

* Element Uniformity (5 1.4)

* Geometric Linearity (5.1.4)

* Noise or Signal/Noise Ratio (5.1.4, 5.4)

* Rise, Fall Times (4.2, 5.2.1)

* Persistence (4.2, 5.2.2, 5.2.3)

* Refresh Rate (4.2, 5.2.2)

* Color (4.4, 5.3)

* Luminance and Dynamic Range (5.4)

* Viewing Angle

* Symbol, Character Fonts, Sizes (5.5.1)

* Graphics Format (5.5.2)

* Power, Voltage Limitations (3.1.2)

* Installation Constraints (3.1.1)

2.1.3 Technology Selection

The display design requirements, once specified, can be compared directly with the capabilities and limitations of the alternate technologies to eliminate those technologies incapable of meeting specific design requirements. Often this process, if applied rigidly, will result in the elimination of all candidate technologies, since our requirements are often derived under worst-case conditions and there is no all-purpose, ideal display technology. Thus, tradeoffs are inevitable, and there is no universal format for conducting these tradeoffs, although the two applications of Sections 2.2 and 2.3 provide examples.

Nonetheless, Figure 1 and the following steps will often result in logical elimination of totally unacceptable technologies or devices, and will permit the identification of variables suitable for inclusion in parametric tradeoff analyses. At each of the following steps, design variables are used to eliminate candidate technlogies/devices based upon design requirements. Data for these decisions are contained

in the technology summaries of Section 3, and quantitative summary tables that can be referred to directly are given in parentheses. A more generalized summary table, which qualitatively compares technologies with design requirements, follows as Table 1.

2.1.3.1   Power/voltage

The CRT requires 10 to 15 KV and about 100 W to operate, and cannot therefore be driven by batteries for sustained periods. Similarly, the flat CRT, EL, and plasma displays require over 100 mW/cm^2 of power and are not battery compatible (Table 5).

2.1.3.2   Installation depth

The prime reason for finding an alternative to the conventional CRT is to reduce the display depth. If only a few centimeters of depth are available, the CRT must be eliminated (Table 4).

2.1.3.3   Large display size

Multioperator requirements for displays larger than 0.5 m eliminate from consideration the flat CRT, LED, LCD, ECD, and EPID (Table 4). However, projection systems using an LCD light valve may be in competition for some limited ambient illuminance situations.

TABLE 1. Qualitative Comparison of Technologies by Design Variables

Technology	Size	Power/ Voltage	Color Capability	Luminance Capability	Resolution	Dynamic Range	Uniformity	Matrix Addressing	Cost
Cathode-Ray Tube (CRT)	Miniature to large projection	high	yes	low to high	high	yes	fair	yes	low
Flat CRT	small	medium	yes	low to medium	medium	yes	fair to good	yes	high
Light-emitting diode (LED)	small	low	limited	low to very high	high	yes	good	no	low
Electroluminescent (EL)	small to large	medium to high	limited	low to high	high	yes	fair	yes	high
Plasma	small to medium	high	possible	medium	medium	yes	good	yes	high
Liquid Crystal (LCD)	small to medium	low	limited	n/a	medium	yes	good	yes	low to medium
Electrochromic (ECD)	small to medium	low	discrete	n/a	unknown	no	good	no	low
Electrophoretic (EPID)	small to medium	low to high	discrete	n/a	medium	yes	good	probably	low

2.1.3.4 Dynamic gray scale

Dynamic shades-of-gray video requirements (e.g., television) will eliminate from consideration the ECD (Table 10).

2.1.3.5 Color coding

Flexible and variable color coding eliminates the LED, and is compatible with the EL, plasma, LCD, ECD, and EPID under limited circumstances only (Table 6). Typically flexible, dynamic color coding will require selection of the CRT or flat CRT.

2.1.3.6 High resolution

While various applications demand different resolution limits, if one assumes "high resolution" to imply element densities of at least 3.25/mm, then the flat CRT and ECD are usually eliminated from consideration (Table 9).

2.1.3.7 High ambient illuminance

Although contrast-enhancing filters can be very useful, some technologies cannot meet minimum contrast requirements under high (e.g., 5000 lux) ambient illuminance. Those eliminated for this reason will typically include the flat CRT and plasma displays, and often the CRT and EL (Table 10).

2.1.3.8 Large element arrays

Displays requiring at least 512 x 512 element arrays elimi-
nate from consideration the LED and ECD (Table 13).

Other considerations, such as cost or compatibility with
raster addressing, may also serve to eliminate certain tech-
nologies or display units. As in all systems design activi-
ties, however, the selection or elimination of specific
items should be made on the basis of quantitative tradeoffs
and analyses. The items described above constitute only a
"shopping list" and general approach. Detailed analyses
should be based upon known charaacteristics of the various
display technologies (summarized in Section 3) and empirical
human operator requirements and capabilities (summarized in
Sections 4 and 5).

The following two application examples illustrate how
this approach and the supporting data can be applied to rep-
resentative Navy information display situations.

2.2 AIRBORNE DISPLAYS

It is neither feasible nor possible to include a complete
airborne display requirements analysis and evaluation within
this report. Such an approach would require a complete mis-
sion definition, crew size assumption, aircraft
flight/weapons capability definition, and ultimately a veri-
fying task analysis. Such a procedure is beyond our level
of concern and intent. In fact, such efforts have been par-

tially or wholly implemented under other programs, such as the Navy AIDS (Advanced Integrated Display System), the Air Force DAIS (Digital Avionics Information System), and a recent effort under Air Force sponsorship (Mills, Grayson, Loy, and Jauer, 1978).

However, within the context of this report, it is instructive to compare the characteristics of the surveyed technologies against a generic set of airborne display requirements. These generic requirements have been synthesized by the author from several sources, and are considered to be representative of the cockpit display requirements of a high performance Naval aircraft of the 1980s. (The reader is cautioned that the following representative requirements are only illustrative, and that specific requirements must be set for each system.)

2.2.1 Generic display requirements

For purposes of this analysis, the generic required displays will be (1) a head-up display, (2) a vertical situation display, (3) a horizontal situation display, and (4) an auxiliary display. The purpose of each and its displayed information are described below (Hitchcock, 1973).

2.2.1.1 Head-up display (HUD)

The HUD provides basic flight control and weapons delivery information to the pilot, while it simultaneously per-

mits the pilot to view the outside world. The see-through display is located above the top of the instrument panel, between the leading edge of the panel and the windscreen. Information to be displayed on the HUD includes the following:

* Aircraft reticle -- the frame of reference for the other symbology

* Horizon -- an artificial horizon for reference during times of degraded visibility

* Pitch ladder -- lines depicting pitch attitude in quantitative increments

* Magnetic heading -- digital heading readout

* Velocity vector -- point of predicted impact of current aircraft trajectory

* Target symbol -- computed line-of-sight to expected target position

* Boresight reference -- line-of-sight of the armament datum line

* Armament legend -- status and selection of armament

* Steering symbol -- commanded flight path symbol

* Target range data -- present range and weapons effectiveness range

* Mode advisory legends -- alphanumeric legends to indicate system operating mode

* Altitude scale -- scale to indicate either barometric altitude or radar altitude

* Bomb fall line -- superimposed prediction of weapons delivery line on outside world

* Vertical speed scale -- scale showing rate of ascent or descent

* Angle-of-attack error -- symbol indicating error between actual angle of attack and commanded angle of attack

From this listing, it is clear that the HUD must display both vectorgraphic and alphanumeric information, and that this information is dynamically changing. In fact, the positional rate of change of some of the information can be as great as 200 deg/s, requiring an update rate of 300 times/s (Hitchcock, 1973).

Figure 2 illustrates a representative HUD symbology for air-to-air attack, and Figure 3 illustrates one for air-to-ground weapons delivery, as seen by the pilot.
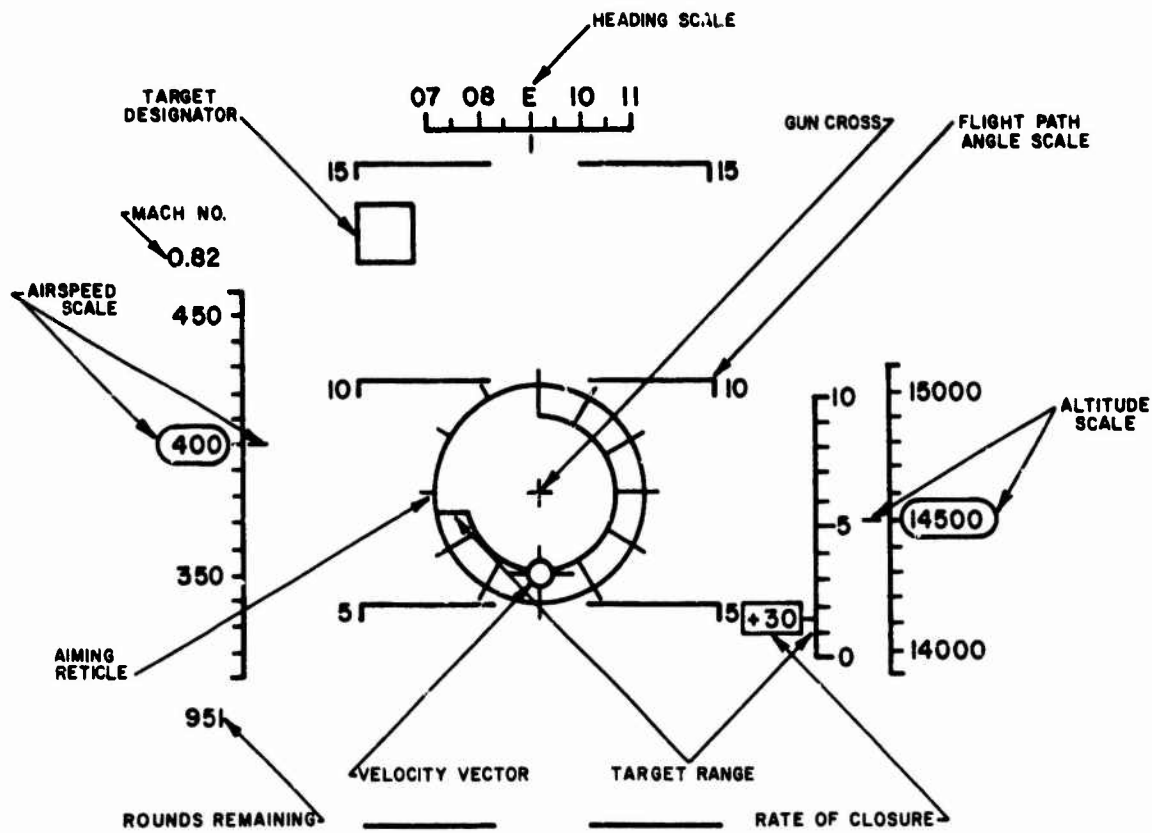
Figure 2: HUD symbology for air-to-air attack mode

The HUD must simultaneously accomplish three functions. First, it must present the required information accurately to the pilot, with sufficiently exact positioning and in an unambiguous, yet uncluttered, format. Second, it must have a field of view, or angular coverage, adequate for all functions in which displayed information (e.g., target position) must be overlaid on the real world image. Third, the transmission through the HUD must be sufficient so as not to reduce pilot visibility through the display.

These functional requirements can be translated into a set of design or performance requirements, which will of
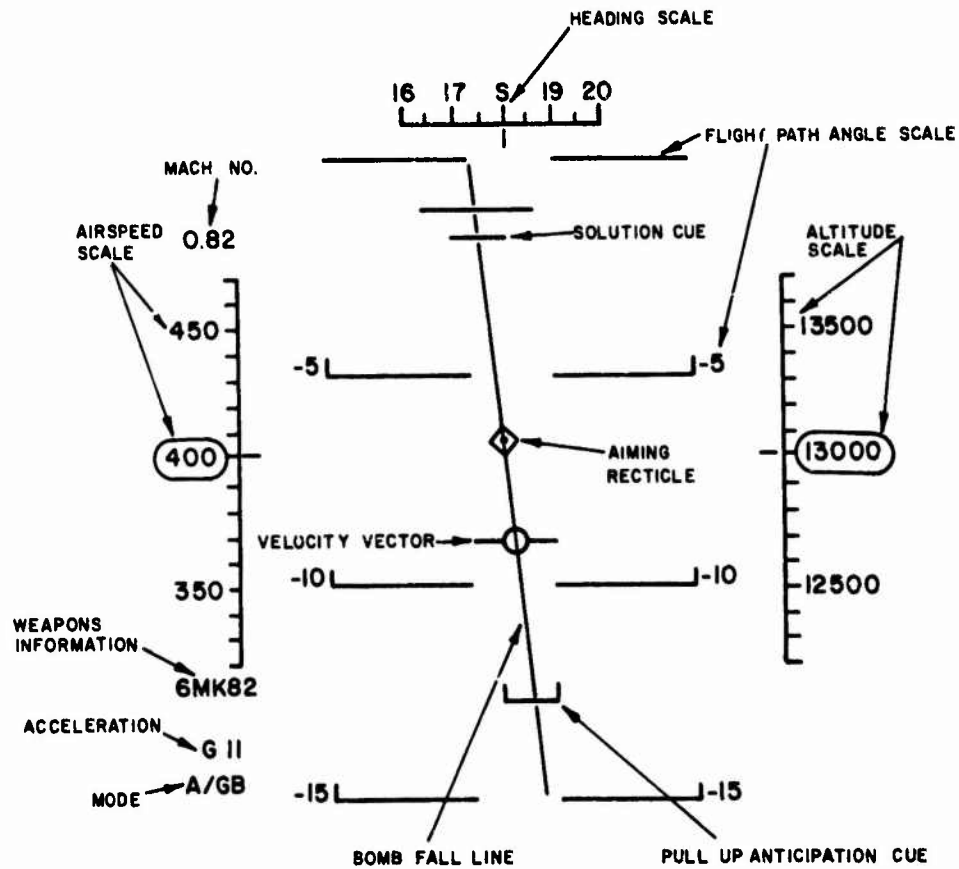
Figure 3: HUD symbology for air-to-ground weapon delivery mode

course vary with the mission and aircraft. For our pur-
poses, we shall adopt the following generic design require-
ments:

* Contrast ratio -- $\geq$ 1.4:1, preferably 9:1

* Transmittance -- $\geq$ 80%, preferably $\geq$ 90%

* Binocular disparity -- $\leq$ 1 mr ($\leq$ 3.4 arcmin)

* Field of view -- $\geq$ 20 deg (vertical) by 30 deg
  (horizontal)

* Line width -- 1.0 mr (3.4 arcmin)

* Character height -- 7 mr (24 arcmin)

* Eye/display distance -- 56 cm

* Luminance nonuniformity -- $\leq$ 20%

* Linearity -- < 2%

* Character font -- 5 x 7 dot matrix, Huddleston

* Color -- monochrome

* Image luminance -- $\geq$ 10,000 cd/m^2

The contrast ratio of 9:1 follows from Section 5.5, which indicates that a modulation of 80% leads to asymptotic character legibility. The minimum contrast ratio is set at 1.4:1 because this is still well above threshold for the symbol angular subtense (3.4 arcmin). A line width of 3.4 arcmin (8.8 cyc/deg) requires a modulation of 0.015 at threshold; the minimum recommended modulation of 0.17 (CR = 1.4:1) is over 11 times the threshold value (Section 4.1), which should be adequate for 95% detection, especially when the trained pilot knows the position of the needed information. While less modulation would be required if the displayed line widths were larger (in the region of 2-3 cyc/deg, or 10-15 arcmin), the amount of information required for display (Figures 2 and 3) would cause a clut-

tered appearance. Thus, this combination of size and modulation is considered an effective tradeoff result.

Similarly, the binocular disparity is within easy fusion capability of healthy observers, yet compatible with 1 element of display resolution, the minimum possible error. The field of view is arbitrarily set by operational mission considerations, as the eye/display distance is constrained by cockpit geometry and ejection envelope requirements. The luminance nonuniformity, while possibly detectable, is not critical for a symbolic display. Linearity of 2% is achievable with any technology, but is governed by pilot control capabilities in conjunction with mission requirements. No color coding is needed, although color contrast might improve display visibility. The image luminance is, as noted, determined by the background luminance, optical characteristics, and contrast ratio.

The font is specified to be a 5 x 7 dot matrix Huddleston, in keeping with the results of Section 5.4.3. Assuming that two line widths (2 mr) are maintained between characters, vertically and horizontally, then a maximum of 39 characters can be placed on the screen vertically, and 75 characters horizontally. Of course, this density would yield a cluttered, impractical display, but the information is useful in estimating the appearance of the formats depicted in Figures 2 and 3.

2.2.1.2 Vertical situation display

The VSD presents the basic data required for aircraft control. It contains a display of current flight information, such as attitude, altitude, rate of climb, heading, airspeed, etc. It may also present information about the outside world ahead of the aircraft, such as radar, infrared, or television images (Hitchcock, 1973). The symbols and alphanumerics displayed on the VSD are likely to include:

* Pitch attitude -- the displacement of the horizon line from the aircraft symbol

* Roll attitude -- rotation of the horizon line

* Aircraft heading -- magnetic bearing of aircraft velocity vector

* Mach/airspeed -- aircraft speed through air

* Command airspeed -- commanded airspeed, plus display of deviation (error) from indicated airspeed

* Altitude -- radar and barometic altitude

* Command altitude -- programmed "ideal" altitude

* Command heading -- programmed "ideal" heading

* Steering command -- programmed "ideal" flight path

- 23 -

* Aircraft symbol -- fixed reference in center of display

* Sensor imagery -- television, infrared, or radar dynamic imagery, usually presented in raster scan

* Breakaway command -- discrete symbol to indicate steering change due to weapon release

* Command angle of attack -- programmed "ideal" pitch angle minus velocity vector

* Velocity vector -- velocity vector of aircraft

The VSD must therefore display TV-type information, as well as alphanumerics and vectorgraphics. Of all cockpit displays, the VSD is perhaps the most critical for flight safety. It is simultaneously the display that requires the most information integration and content variability.

Representative VSD display formats are shown for air-to-ground weapon delivery (Figure 4) and air-to-air attack (Figure 5). It must be remembered that video information (i.e., TV) can be superimposed on this symbology at all times.

As in the case of the HUD, the detailed derivation of display functional requirements is beyond our purposes. However, a representative set of VSD design criteria is given as follows:
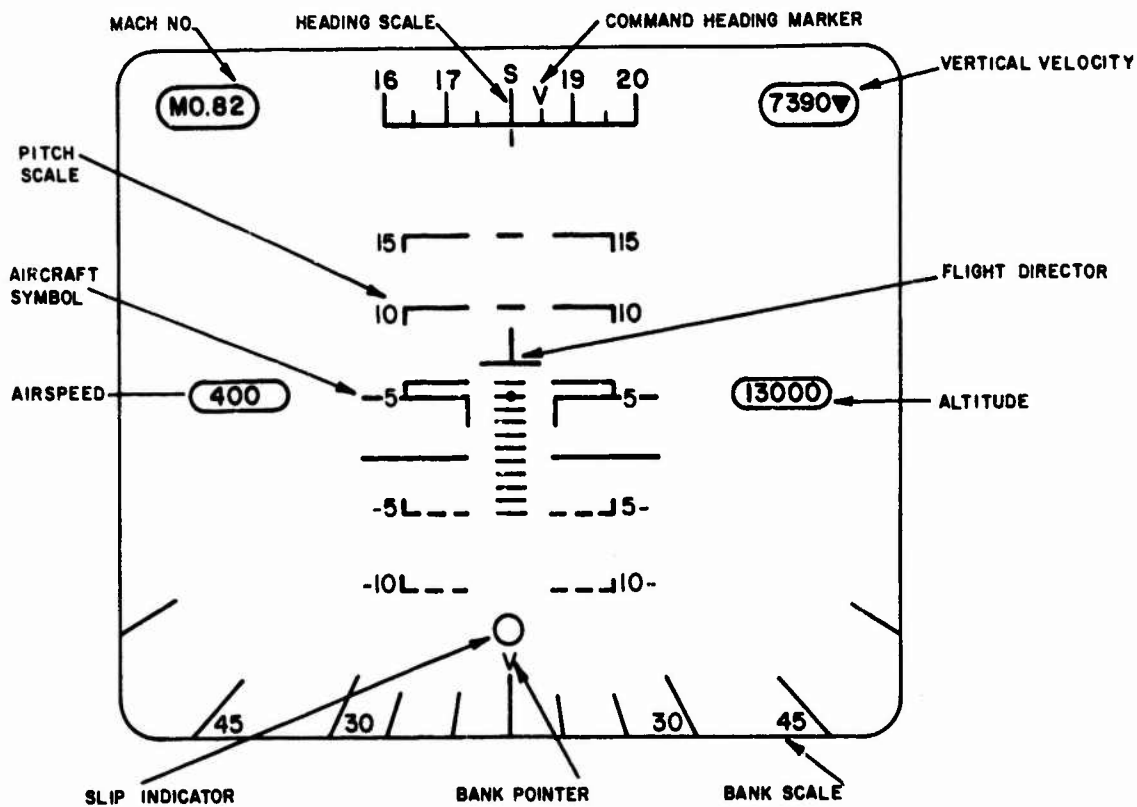
Figure 4: VSD symbology for air-to-ground weapon delivery mode

* Contrast ratio -- $\geq$ 1.4:1, 9:1 preferred for symbology; 32:1 for video

* Gray scale variability -- 10 discrete levels, minimum modulation = 0.94

* Resolution -- 3.15 TVL (or elements) per mm

* Viewing distance -- 71 cm

* Viewing angle -- $\pm$ 15 deg

* Refresh rate -- $\geq$ 50 Hz
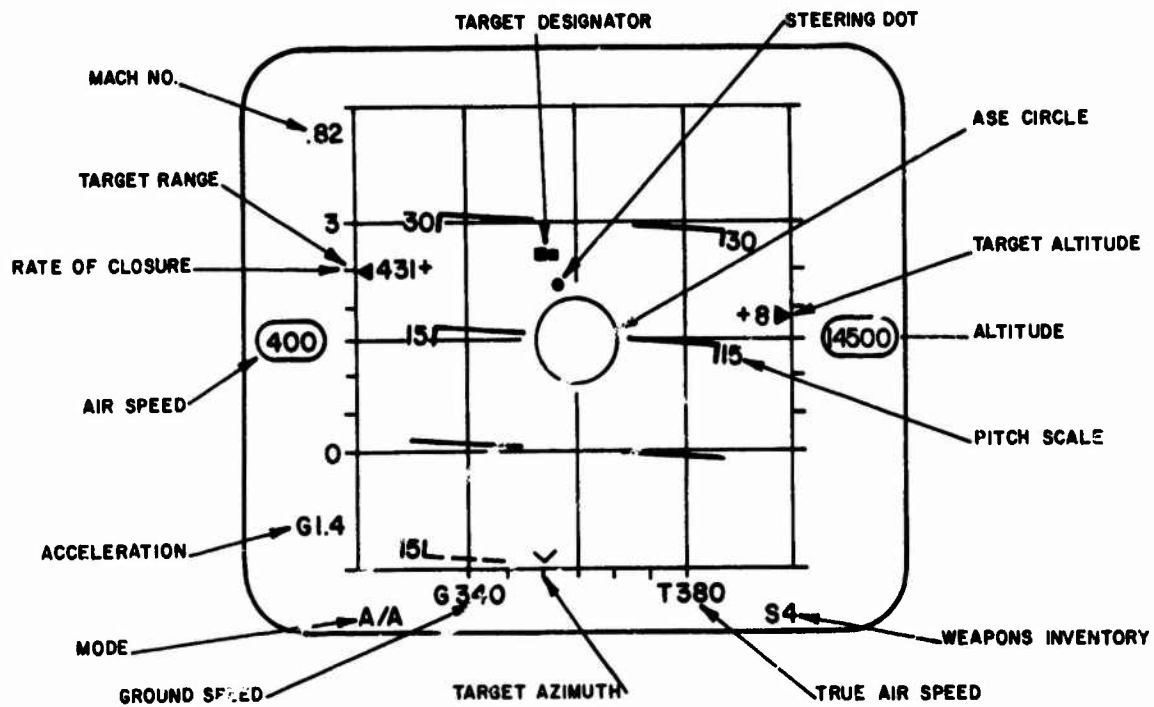
Figure 5. VSD symbology for air-to-air attack mode

* Display size -- 20 x 25 cm

* Power -- $\leq$ 400 W

* Nonuniformity -- $\leq$ 20%

* Line width -- 0.9 mr (2 elements wide)

* Linearity -- < 2%

* Jitter -- < 0.5 element (or TVL)

The reasoning behind contrast ratio and line width requirements is the same as for the HUD. In the case of the

VSD, however, there is no "see-through" requirement, and the display can therefore be covered with a glare-reducing, contrast-enhancing filter.

The element size, or resolution, of the display subtends 1.54 arcmin, for a spatial frequency of 19.5 cyc/deg. To avoid visible "rastering" or detection of the resolution elements, the no-image element modulation should be less than 6% for the fundamental spatial frequency.

Display size is based on the mission analysis and display symbology requirements (Figures 4 and 5) and upon available cockpit space. With a resolution element size of 0.32 mm, there are 630 elements vertically and 780 horizontally on the display. When a TV-type image is presented, requiring typically 490 x 612 elements, the excess elements around the edges can be used for critical symbology which is not superimposed on the TV image.

2.2.1.3 Horizontal situation display (HSD)

The HSD presents the information the pilot requires to position his aircraft accurately over the earth's surface, that is, to navigate. His general analog of the HSD is a map with the aircraft's heading toward the top. In fact, the VSD might have an aeronautical chart projected on its surface, with symbols superimposed on the map image.

The information requirements for the HSD are:

* Navigation symbols -- aircraft position, waypoints, destination, latitude and longitude, all in plan view

* Time-to-go -- calculated time to selected waypoint or destination

* Course -- ground track, in plan view

* Radar -- imaging display of ground map (PPI sector scan) radar

* Terrain avoidance/clearance -- outline of all terrain at or above clearance altitude, computed from radar image and displayed symbolically

* Heading -- current aircraft magnetic heading

* Wind velocity and bearing -- computed data

* Threat symbol -- ground position of any sensed or known threat

Representative HSD display formats are shown for air-to-ground weapon delivery (Figure 6), cruise (Figure 7), radar terrain avoidance (Figure 8), and landing (Figure 9).

Representative functional display design requirements for the HSD are very similar to those for the VSD. They are:

* Contrast ratio -- 9:1 for video and symbology, 1.4:1 minimum for symbology

- 28 -
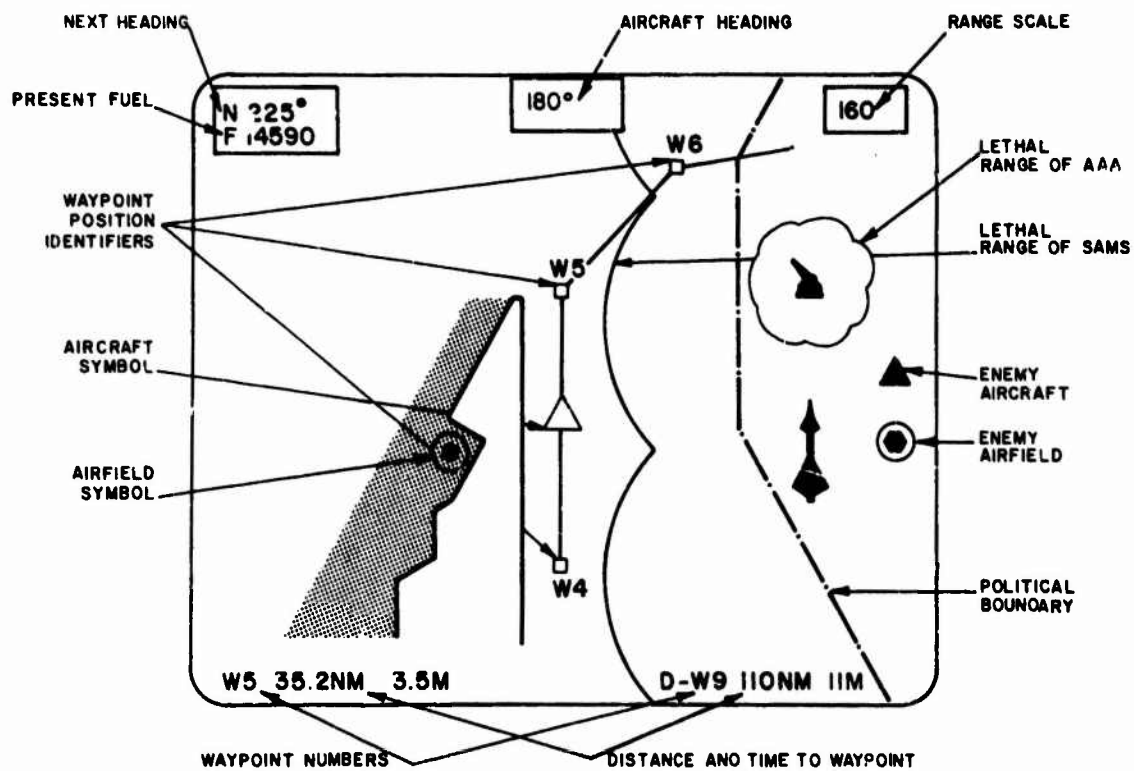
Figure 6: HSD symbology for air-to-ground weapon delivery

* Gray scale variability -- 0.80 minimum modulation, 6
  discrete levels

* Resolution -- 3.15 TVL (or elements) per mm

* Viewing distance -- 71 cm

* Viewing angle -- ± 15 deg

* Display size -- 20 x 25 cm

* Power -- ≤ 150 W

* Nonuniformity -- < 20%

- 29 -
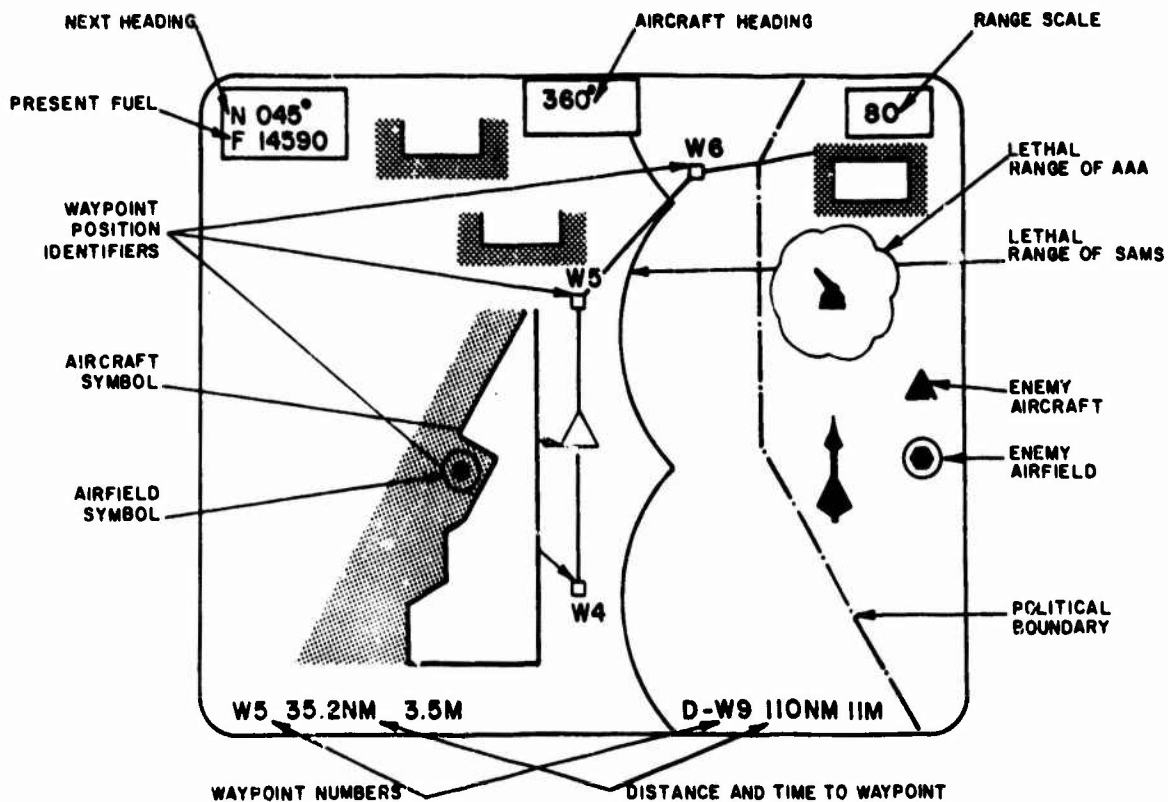
Figure 7: HSD symbology for cruise, north up

* Line width -- 0.9 mr

* Linearity -- < 2%

* Jitter -- < 0.5 element (or TVL)

The requirements in size and element density are the same for the VSD and the HSD, as are other image related parameters except the contrast ratio and addressable gray levels. No TV or infrared information will be presented on the HSD; however, radar information will be required. As a result, the depictable gray levels must correspond to the number of usable levels in the radar signal. Six are assumed for our purposes.
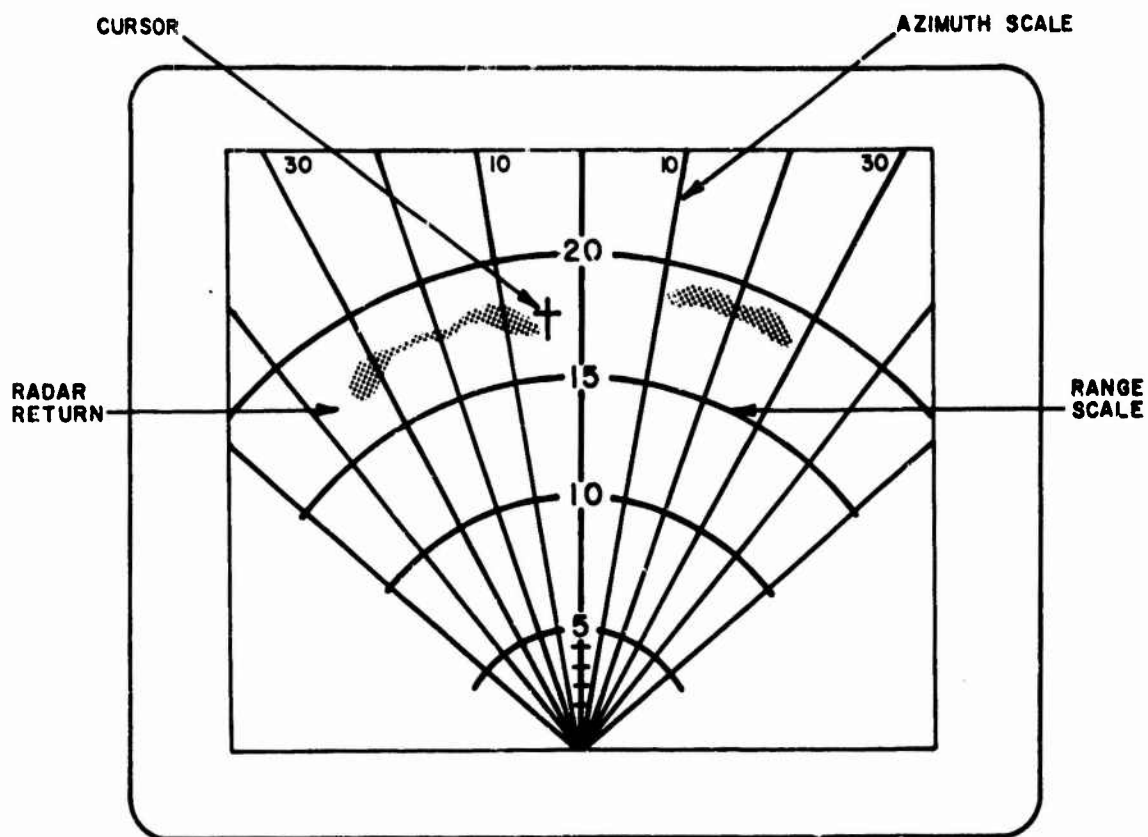
Figure 8: HSD display for radar terrain avoidance

2.2.1.4 Auxiliary display (AD)

The auxiliary display (of which there may be two) is used to present checklists, energy management, engine status, and weapons status information. It (or they) basically serves a variety of information needs of the pilot because of its flexibility for presentation of alphanumeric data, vector-graphic information, and warning symbols. Representative display contents include:

* Time to altitude -- predicted time to reach prese-
  lected altitude
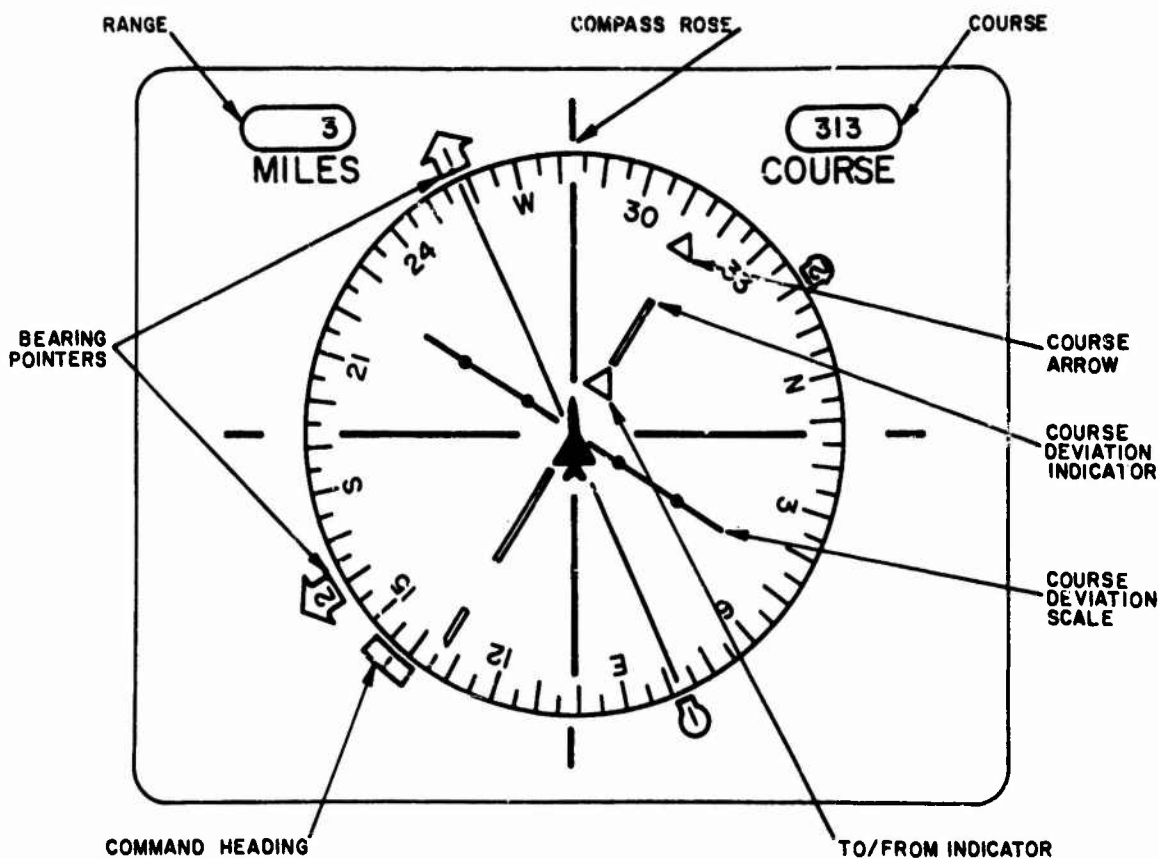
Figure 9: HSD display for landing

* Time for minimum fuel climb -- climb time for mini-
  mum fuel consumption

* Time for minimum time climb -- climb time for mini-
  mum time consumption

* Energy level -- current percent of maximum aircraft
  kinetic energy

* Engine status -- satisfactory/unsatisfactory

* Fuel flow rate -- rate of fuel consumption

* Oil pressure -- as measured

* Stores status -- weapon stores by station

* Checklists -- prestored checklists for all flight
  segments

* Reference information -- profiles, missed approach
  procedures, approach plates, etc.

Figure 10 illustrates a preflight checklist, while Figure
11 indicates the landing approach and status, in profile and
plan views.

CHECKLIST

AFTER ENTERING COCKPIT
1. RUDDER PEDALS - ADJUSTED
2. SHOULDER AND HIP HARNESS - CONNECTED
3. ALL PILOT'S EQUIPMENT - ATTACH AND SECURE
4. OXYGEN SYSTEM - CHECK (PRICE)

LEFT CONSOLE
1. VENT AIR CONTROLS - SET
2. DOUBLE DATUM SWITCH - OFF
3. ADF/ AUXILIARY UHF RECEIVER - SET AND OFF
4. RHAW - PWR OFF
5. INTER - SET
6. IFF - SET AND OFF
7. UHF RADIO - SET AND ON
8. EMER FLAP SWITCH - NORM
9. FLAP HANDLE - FLAP UP

Figure 10:   AD with checklist
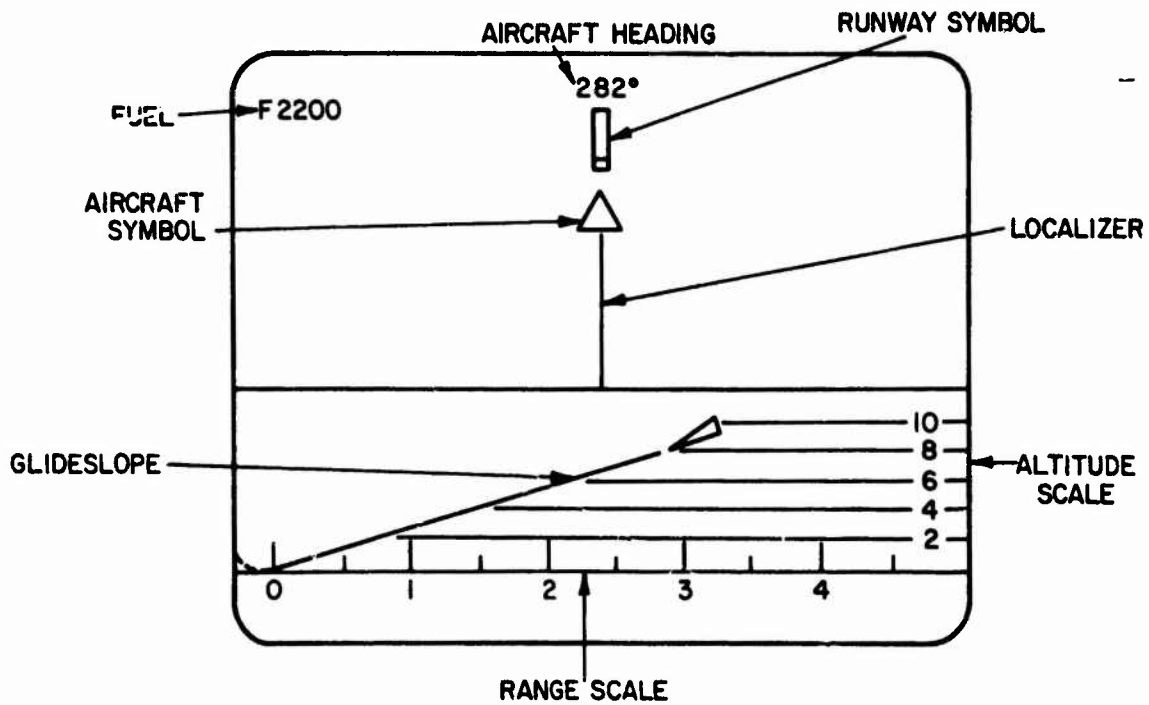
Figure 11: AD illustrating glideslope and localizer
            information for landing

        Representative functional display requirements for the AD

are:

   * Contrast ratio -- > 1.4:1, 9:1 preferred

   * Resolution -- 2.4 elements/mm

   * Viewing distance -- 71 cm

   * Viewing angle -- ± 15 deg

   * Refresh rate -- ≥ 50 Hz

   * Display size -- 20 x 25 cm

* Power -- $\leq$ 150 W

* Nonuniformity -- < 20%

* Line width -- 1.17 mr (2 elements wide)

* Linearity -- < 2%

* Jitter -- < 0.5 element

The AD requirements are similar to those of the VSD and HSD. Since no imagery will be displayed, however, no gray scale is needed and display modulation need only be adequate for symbology and vectorgraphics. Similarly, the amount of information to be placed on the AD is less, and the total display resolution will be adequate at 480 x 600 elements. With each element subtending 2.02 arcmin, for an equivalent spatial frequency of 14.9 cyc/degree, the flat-field element modulation, in the "on" or "off" condition, must be less than 0.012, which is a difficult criterion to meet.

2.2.2   Technology Selection

2.2.2.1   Head-up display

The technology choices for the HUD can initially be reduced by addressing the (1) luminance and (2) resolution requirements.

First, the displayed image luminance must be at least 10,000 cd/m^2. However, the image source in a "see-through" display is projected onto the combining glass as indicated in Figure 12. The luminance at the source is given by:

$$L_S = L_C/R \qquad\qquad (1)$$

in which

L_S = source luminance,

L_C = displayed luminance, and

R = combiner reflectance.

But,

$$L_C = L_B(CR - 1)(T), \qquad\qquad (2)$$

in which

L_B = background luminance,

CR = contrast ratio, and

T = transmittance of background luminance through
the combiner.

Therefore,

$$L_S = (L_B)(CR - 1)(T)(R)^{-1}. \qquad\qquad (3)$$

Assuming that the scene luminance is 30,000 cd/m^2, CR = 1.4, T = 0.9, and R = 0.8, the source luminance must equal:

$$L_S = 30,000 \text{ cd/m}^2(1.4 - 1)(0.9)(0.8)^{-1}$$
$$= 13,500 \text{ cd/m}^2.$$

As indicated in Section 3.3, only the CRT and LED displays can meet this requirement, while the EL comes close. However, if the ambient illuminance is used as a light source for a light modulating display, then one can also consider LC, EC, and EPID technologies.

Secondly, the resolution requirements of 1 mr at a distance of 56 cm and over a 30 deg field of view may permit us to narrow down the possibilities. The 30 deg field of
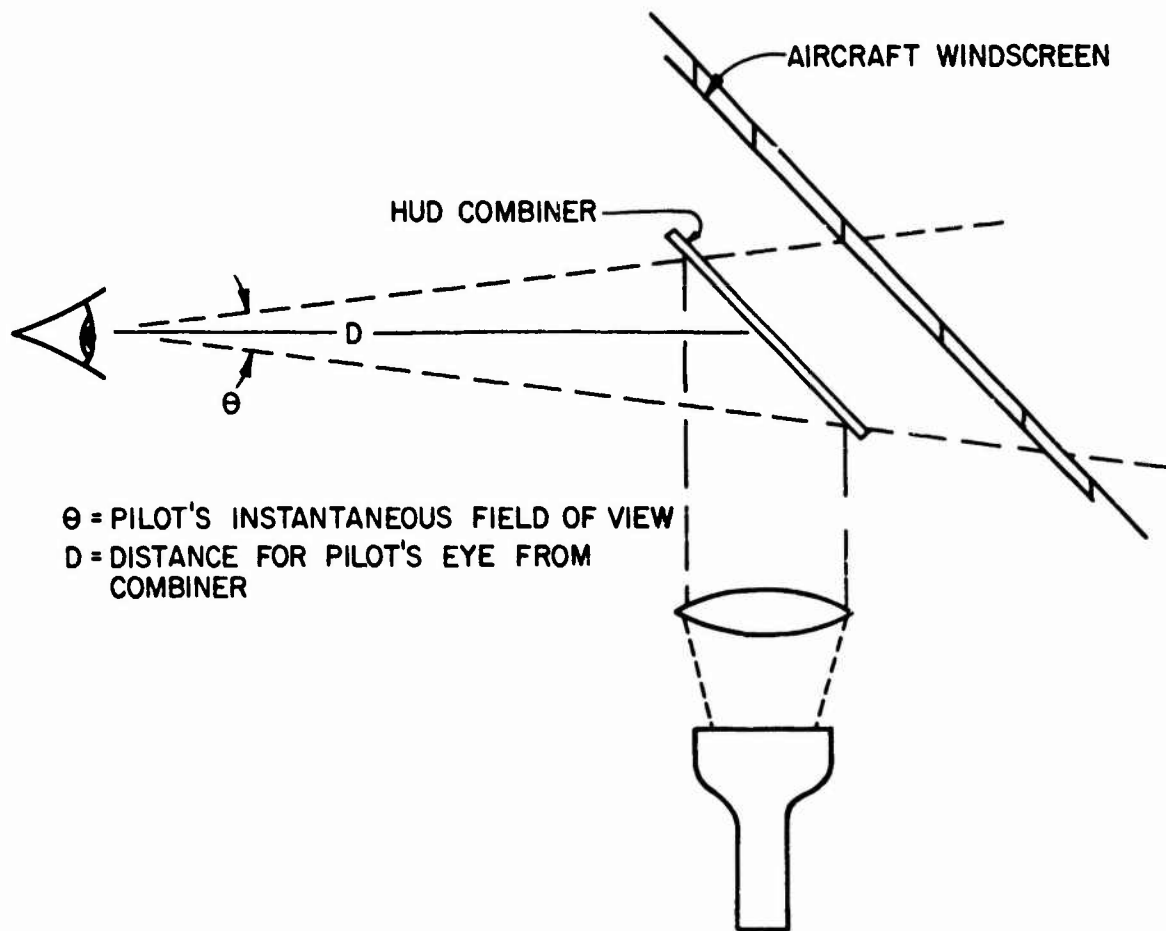
Figure 12: Traditional HUD optical schematic

view is equal to 524 mr. Thus, the display source must have
at least 524 resolution elements. If we assume that the
source must be conveniently small (e.g., 12 x 8 cm) compared
to the optical combiner, which must be 30 x 20 cm at the
56-cm viewing distance, then the source must have an element
density of 524 elements/12 cm, or 4.4 elements/mm. Refer-
ring to Table 9, only the CRT, LED, and EL technologies meet
these requirements. However, if the display source size
were increased slightly, the LCD would also be competitive.

At this point, other tradeoffs become pertinent. Specif-
ically, the conventional reflective optics are incapable of

providing both a transmission of 90% and a reflection (R) of the displayed image of 80%. In addition, collimating optics required with a reflective combining glass must be equal in diameter to the diagonal of the combining glass, or 36 cm. There is little room for a lens of this diameter inside the panel. Therefore, refractive optics, using a holographic lens, are recommended. It is feasible to obtain such a refractive combiner with a reflectance of over 80% and a transmission of over 90%.

Using these optics, one can choose a CRT with adequate resolution and luminance. Its useful life, at 13,500 cd/m^2, is likely to be on the order of 3000 hours, and it would draw about 7.5 W of power. Alternates would include an LED array which could produce over 60,000 cd/m^2. However, the required refresh rate, to prevent image breakup or "strobing" at the image movement rates required, would be in excess of 500 Hz. As will be noted later, it is not feasible to address a large LED array at this high rate.

The EL display could provide the necessary luminance, element density, display size, and refresh rate. It is definitely a meaningful candidate.

Lastly, one should consider light modulating displays. EC and EPID displays are possible candidates, but neither has yet been fabricated into arrays that would be suitable for this purpose. The LC array, however, has been developed into 350 x 350 element arrays. With additional development,

the design objective of a 524 x 350 element array might be achieved. The element density obtained thus far (3.93/mm) is not far from the required 4.4/mm, and the power consumption is very attractive. Further, the acceptability of the LC array will depend on its rise/fall time improvement to be compatible with the rates of movement of some display elements, e.g., horizon bar, airspeed, and magnetic heading.

At this time, Hughes Aircraft has developed a prototype refractive HUD using an LC source. Its specifications (Ernstoff, 1978) are as follows:

* Total field of view: 20 deg (horizontal) by 14.5 deg (vertical)

* Instantaneous field of view: 16 deg x 12 deg

* Binocular disparity: $\leq$ 1 mr

* Combiner transmission: 95%

* LC display resolution: 350 x 350 elements, 1 element/mr

* Illumination source power: 50 W

* Symbol luminance: adjustable between 1 cd/m^2 and 9600 cd/m^2

* Contrast ratio against 102,780 cd/m^2: 1.3:1

* Contrast ratio against dark background: 32:1

A schematic drawing of this device is shown in Figure 13.
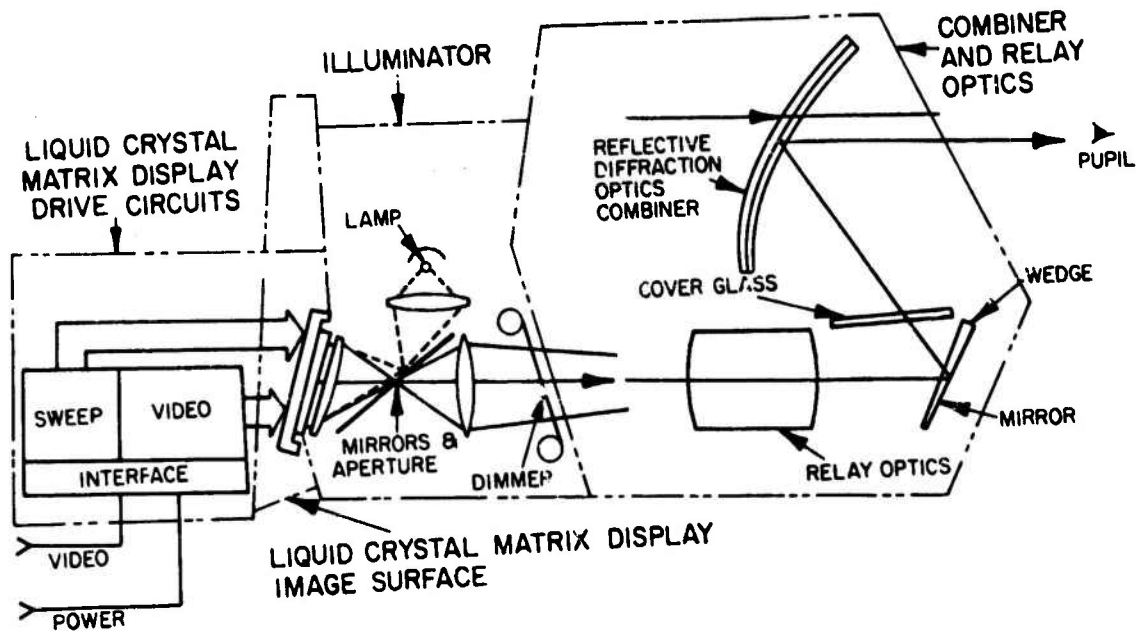
Figure 13: Conceptual drawing of Hughes's HUD with LC
display, from Ernstoff (1978)

At the present time, the only proven choice for a HUD is
the CRT with either reflective or refractive optics. How-
ever, with progress in LCD development, a more attractive
choice seems to be imminent.

2.2.2.2   Vertical situation display

The major considerations in technology selection for the
VSD are the display size (20 x 25 cm), the element density

(3.15 elements/mm), the gray scale (M = 0.94), and its addressing/driving interface requirement to accept video information.

The size requirement can be met by the CRT, LED, EL, plasma, and LC technologies. Further, each of these can meet the element density requirement, as well as the modulation requirement under no ambient illuminance. However, as the ambient increase and contrast-enhancing filters are necessary to maintain image modulation, the mean image luminance is reduced. The precise nature of this tradeoff between decreased image luminance and increased contrast depends on the type of filter and the emission spectrum of the display.

Figures 14 and 15 illustrate this tradeoff for two display luminance levels for a P43 CRT phosphor and a wavelength "notched" filter. Since our required contrast ratio is 32:1 (M = 0.94), it is seen that the CRT contrast ratio falls far short of the required value, even for a narrow notched filter (T = 0.1) and a high luminance output (3,425 cd/m^2). If a much greater luminance level were used for the CRT (to a maximum of 30,000 cd/m^2), then it is possible for a filtered CRT to meet the requirements. The CRT luminance output would have to be about 7000 cd/m^2, which is quite feasible, although there might be substantial spot size increase. Specific CRTs could be compared for luminance/spot size tradeoffs, and the effect on resolution can be determined from well-known relationships.
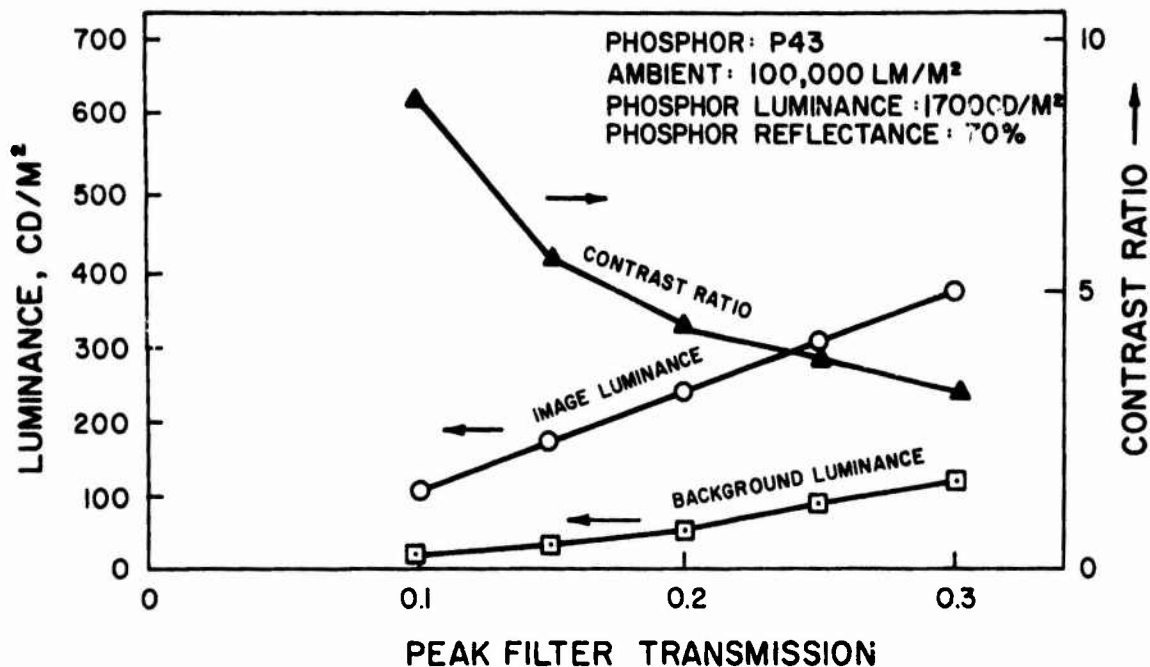
Figure 14: Contrast ratio and image luminance as functions of peak filter transmission

The LED display can meet luminance, resolution, and size requirements, but is not amenable to matrix addressing for the video requirements.

The EL display will meet luminance, size, and resolution requirements, and will barely meet the modulation requirement. Its narrow spectral emission (525, 585 nm) also makes it a good candidate for narrow notch filtering to enhance contrast. Further, it can nicely meet the addressing requirements of a video signal.

While the plasma panel has much to offer, its limited luminance output will result in a very low contrast under
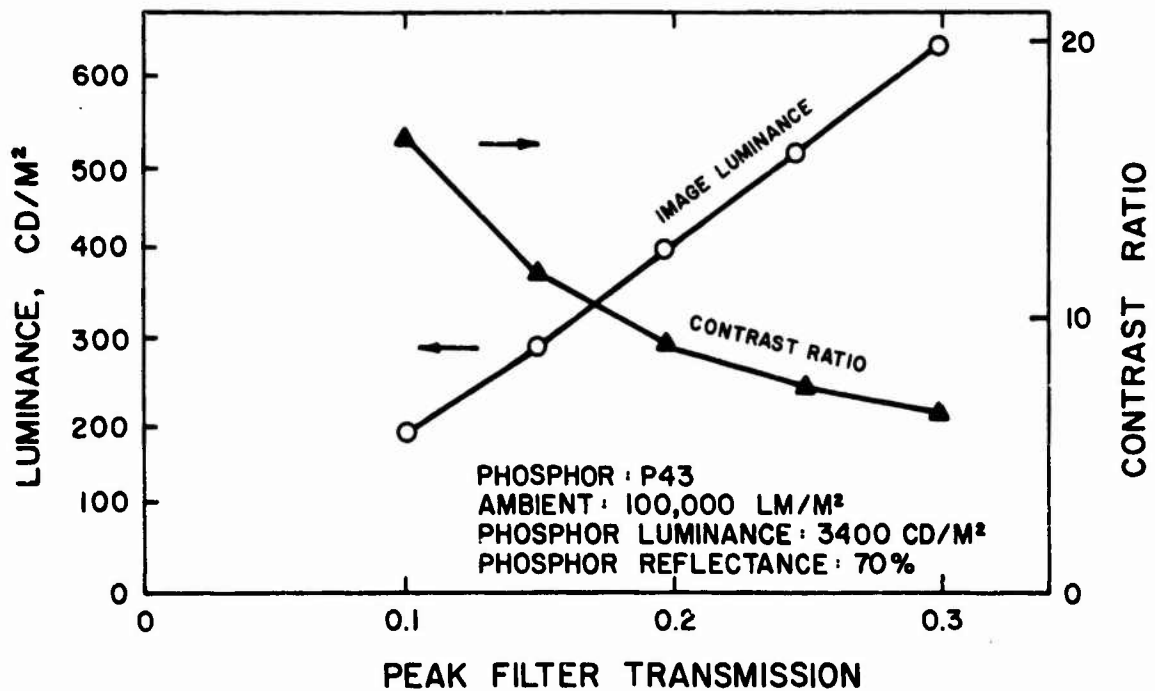
Figure 15: Contrast ratio and image luminance as functions of peak filter transmission

the high cockpit ambient illuminance. For this reason, it will not be an acceptable candidate.

Lastly, the LC display meets all requirements, but may present problems due to rise/fall times. While these problems are being attacked, the current 50 ms rise/fall time is not acceptable for dynamic video presentation or for rapidly moving symbology, such as the horizon line or altitude indicator on the VSD.

Thus, the remaining candidates are the high intensity CRT and the EL. The EL has significant advantages in expected life, thickness (or depth), voltage (600 versus 15,000),

jitter, and geometric stability. However, its low luminous efficiency (19 lm/W), compared to that of the CRT (65 lm/W), causes it to require much greater power. For the 500 cm^2 area in this application, its power consumption would be about 600 W, well in excess of the 100 W required by the CRT. Thus, at the present time, the CRT remains the prime candidate for the airborne VSD application.

2.2.2.3   Horizontal situation display

As indicated in Sections 2.2.1.2 and 2.2.1.3, the HSD requirements are essentially the same as those for the VSD, with the exception of the modulation required for gray scale rendition. As a result, the same candidate technologies (CRT, EL, plasma, and LC) should be considered. However, the plasma can again be eliminated due to its low luminance output (and hence low modulation under high ambient illuminance).

The EL, while meeting other requirements, will draw about 600 W of power, which is well in excess of the allowable amount and would require significant heat dissipation.

The LC display is a meaningful candidate for the HSD. It can provide the needed modulation under high illuminance (and can use an internal light under low ambient conditions), meets resolution requirements, and will draw less than 1 W of power. Because the information rate to the HSD, for both radar video and symbology, is lower than that for

the VSD, the 50 ms rise/fall times of the LC may not be a problem, although a more detailed analysis would be required.

Thus, the CRT and the LC display technologies are attractive candidates for the airborne HSD application.

2.2.2.4 Auxiliary display

The AD requirements are generally similar to those of the HSD, except that no gray scale is required and the resolution is somewhat less, at 2.4 elements/mm.

The candidate technologies which can meet the resolution, size, and modulation requirements are the CRT, LED, EL, plasma, and LC displays.

The EL can be eliminated for power consumption reasons, as it was for the VSD and HSD.

The plasma panel can be eliminated because its modulation will be too low under high ambient illuminance levels.

The LED becomes moderately attractive for this application because no gray levels are required, the amount of information to be simultaneously displayed is within its addressing/driving limits, it requires little depth (e.g., 10 mm), and its color flexibility is large. However, it would draw about 1 mW/element, or about 288 W if all elements were on. Because the number of "on" elements at a time will be less than 35% (typically), then the total power consumption would be about 100 W, which is within our allowable power budget.

The LC display is quite attractive for this requirement because (1) it will provide better contrast than either the CRT or LED under high ambient illuminance, (2) it requires less power (1 W) and weighs less (about 10 pounds) than the CRT, (3) it has a potentially longer lifetime, and (4) it requires less panel depth.

On balance, the LC display appears to be the best choice for the Auxiliary Display function.

2.3   DISPLAYS FOR AMPHIBIOUS TASK FORCE COMMAND AND CONTROL

To illustrate the application of the content of this report to command and control, the amphibious command ship has been selected. (Again, the reader is cautioned that this is an illustrative example, not a detailed design recommendation.)

"The amphibious command ship, and specifically the LCC 19 Class, is designed to provide the total command and control environment necessary to successfully execute an amphibious operation. . . . The mission of the LCC 19 Class ship is to serve as the command ship for the Amphibious Task Force Commander, Landing Force Commander and Air Control Group Commander during amphibious operations."   (NAVSHIPS, 1976, p. 1-1). To meet the mission objective, the LCC 19 is configured to have an amphibious control center with the following major areas:

* Flag Command

* Landing Force Command

* Supporting Areas Coordination

* Military Operations/Ship-to-Shore/Logistics

* Air Operations

* Surface/Subsurface

* Detection and Tracking

* Electronic Warfare

While each of these areas presents a complex challenge to optimize display, control, and workspace layout design, the Flag Command Area has been chosen as representative for the purposes of this report. The Flag Command must monitor, coordinate, and control all movements of the amphibious task force during all amphibious operation phases. Because the amphibious operation can develop and change rapidly, it is critical that the Flag Command be in constant communication with all functional components of the operation.

"The Flag Command Area is used to display all collected and processed data on the real-time tactical displays, vertical plots, and status boards. The displayed data is evaluated by cognizant personnel and then disseminated as required, both internally and externally" (NAVSHIPS, 1976, p. 2-1).

The Flag Command Area is manned by 17 persons under maximum manning conditions, and is supervised by the Commander/Chief of Staff. Its layout is given in Figure 16. The various communications devices, controls, and displays include the following:

* Status boards

* CRT Typewriter (QUEST)

* Data Display Console (Amphibious Flagship Data System, AFDS)

* Television viewer

* Closed-circuit teletypewriter

* Pneumatic tube

* Navy Tactical Data System (NTDS) circuits

* Radio circuits

* Secure telephones

* SP telephones

* Dial telephones

* Multichannel circuits (MC)

The Commander Amphibious Task Force (CATF) (2110, Figure 16) and the CATF Chief of Staff (2111, Figure 16) monitor
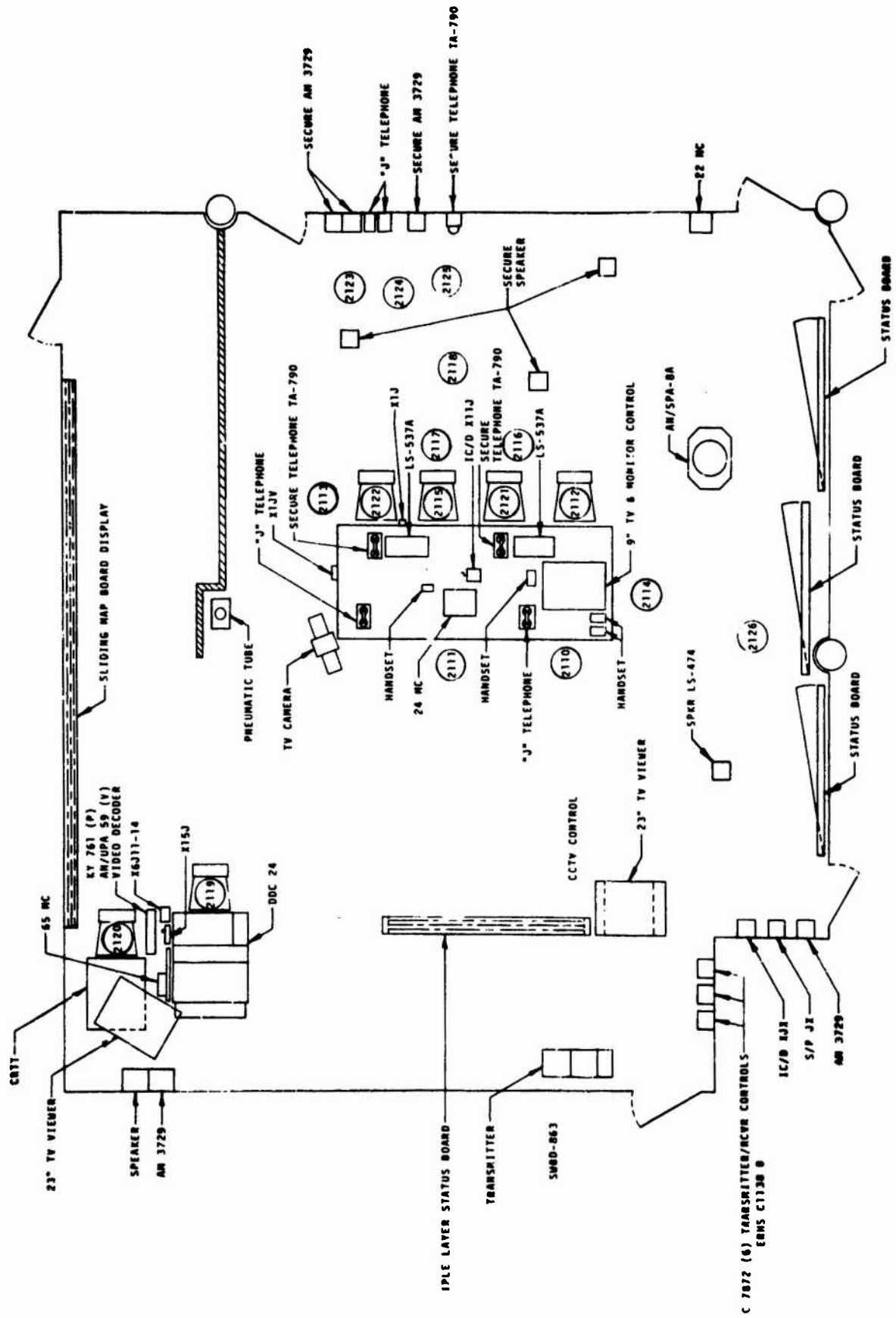
Figure 16: Layout of LCC 19 Flag Command Area

- 49 -

the overall situation by directly viewing the Multiple Layer Status Board and the adjacent TV Viewer. They obtain supplemental verbal information from the Flag Operations Officer (2112), the Flag Intelligence Officer (2113), the Tactical Air Officer (2114), The Flag Meterological Officer (2115), the Flag Command Watch Officer (2116), and the Assistant Flag Command Watch Officer (2117). The Watch Officers are assisted by Tactical Communications (2121, 2122).

The CATF has direct access to dial telephone, interphones, and other handsets, as does the CATF Chief of Staff. Both are in direct phone contact with various units through the NTDS telephone system and secure telephones. Similarly, the Flag Operations Officer, Flag Intelligence Officer, Tactical Air Officer, and Flag Meterological Officer have direct access to secure and nonsecure phone systems. The only visual displays of direct concern to them are the Status Boards and the TV monitors.

The Flag Command Watch Officer, besides having access to the same phone systems around the conference table, is responsible for updating of the displays and status board which show the progress of the current operation or exercise. Thus, the displayed contents of the Multiple Layer Status Board and its adjacent TV monitor must be legible from his position. He is assisted in meeting this requirement by the Assistant Flag Command Watch Officer and the

Watch Supervisor. The Watch Supervisor is stationed approx-imately five meters from the Multiple Layer Status Board.

The Console Operator (2119) and the CRT Typewriter Opera-tor (2120) have the greatest amount of direct involvement with electronic displays. The Console Operator selects information for display on the TV monitor above his console, as requested by the CATF and his staff. Similarly, the CRT Typewriter Operator answers queries from the CATF and staff by accessing the data base. He also updates the data base through his console.

Lastly, the Messenger/Status Board Keeper (2126) main-tains status boards showing meteorological data, voice calls, radio frequencies, initial contacts, unit locations, etc.

From the above general description, it is apparent that there are two general types of displays which are candidates for flat-panel technology: the status board and the CRT/TV monitor. These are discussed below.

2.3.1   Generic Display Requirements

No requirements specifications are known to exist for either the fixed Status Boards or the Multiple Layer Status Board. For this reason, the following section will describe illustrative generic requirements compatible with existing status board usage and displayed content, as well as with design guidelines summarized in Section 5. The same approach will then be taken with the CRT displays.

- 51 -

2.3.1.1 Status boards

The Multiple Layer Status Board and the fixed Status Boards are used to display various combinations of alphanumerics, map outlines, and symbol overlays to depict in plan view the operational situation. These boards are written on (updated) manually from the rear, and are transparent to permit viewing of written information from the front side. Thus, all information placed on the board must be <u>hand</u> <u>written</u> <u>and</u> <u>reversed</u> left-to-right. Some status boards are illuminated (or edge-lit) to permit viewing under the low ambient illumination levels of the command center.

The following generic requirements are typical of such status boards:

* Size -- 2 m wide by 1.75 m high

* Ambient illuminance -- 150 lm/m^2

* Content -- graphics, symbols, alphanumerics

* Information density -- variable, up to one symbol/3 cm

* Contrast ratio -- $\geq$ 9:1

* Line width -- 1 mr (3.4 arcmin)

* Character/symbol height --- 7 mr (24 arcmin)

* Viewing distance -- 5 m, maximum

* Luminance nonuniformity -- $\leq$ 30%

* Linearity -- $\leq$ 5%

* Character font -- 7 x 9 dot matrix (Lincoln-Mitre or Huddleston), or stroke (Lincoln-Mitre or NAMEL)

* Color -- four distinguishable preferred, monochrome acceptable

As was the case for the airborne display, requirements follow largely those derived in Section 5, those dictated by the LCC 19 Control Room layout, and staff functions. The 2 m by 1.75 m size, ambient illuminance, display content, viewing distance, and color are specified by current practice and layout. The maximum information density results from the minimum angular subtense of 24 arcmin for single alphanumeric characters or symbols at the maximum viewing distance of 5 m (Section 5.2.2). The line width (1 mr) requirement is the same as that for other moderate contrast displays, and permits line legibility well above threshold (Section 4.1) without unduly cluttering the display with thick symbol strokes. In addition, the strokewidth-to-height ratio of 3.4/24 = 1:7 is in the recommended range for light-on-dark displays. The contrast ratio ($\geq$ 9:1) insures good character legibility (Section 5.1), as does the character font selection (Section 5.5). Because the display content contains no video information, the luminance nonuni-

formity requirement can be somewhat relaxed (Section 5.2), as can the linearity requirement (Section 5.1).

While it is not possible to specify a minimum viewing angle requirement because most personnel are stationed largely in front of the status board, there are some personnel (e.g., 2126, Figure 16) who might benefit from a wide angle display. Thus, attention should be given to this parameter.

2.3.1.2 CRT displays

The existing 23-inch TV viewers (Figure 16) are used to display alphanumeric information in data format. No graphics are currently used, and none are apparently required for the data sources in use. Rather, these displays contain listings of combat units, calls, readiness states, locations, and the like. Therefore, the generic requirements for the CRT displays are generally as follows:

* Size -- 51 cm high by 56 cm wide, minimum

* Ambient illuminance -- 150 lm/m^2

* Content -- alphanumerics, symbols

* Information Density -- 12 lines of 16 characters each

* Contrast Ratio -- $\geq$ 9:1

* Character/Symbol Height -- 7 mr

* Viewing Distance -- 5 m, maximum

* Luminance Uniformity -- $\leq$ 30%

* Geometric Linearity -- $\leq$ 5%

* Character Font -- 7 x 9 dot matrix or raster equiva-
lent (Lincoln-Mitre or Huddleston, all upper case),
or stroke (Lincoln-Mitre or NAMEL)

* Color -- monochromatic, white or green preferred

* Refresh Rate -- adequate to avoid flicker

* Information Update Rate -- 1 s^{-1}, maximum

The driving requirements for this selection are the
information density, character/symbol size, and viewing dis-
tance. A 7 x 9 dot matrix character must have a "dedicated"
9 x 11 dot matrix cell to provide adequate spacing between
characters and between lines. Thus, the vertical dimension
of the display requires 12 lines x 11 dots/line, or 132
dots. Because each character must subtend 7 mr at 5 m, the
character must be 3.5 cm high. In turn, the "dedicated" 9 x
11 dot cell must be 4.3 cm high and the display 51 cm high,
if all characters on the display are to be readable at 5 m.

Similarly, each "dedicated" cell 9 dots wide requires 3.5
cm, for a display width of 3.5 cm/cell x 16 cells or 56 cm.

Thus, a square display 56 cm on a side, having dot center-line spacing of 0.39 cm will meet the requirements for viewing at 5 m.

However, the 0.39 cm dot spacing will cause the display to be both unattractive and difficult to read at close viewing distances because of the large dot structure. To avoid this difficulty, it is necessary to keep the spacing between dots (assuming a dot matrix display) or between active lines (for a raster display) to 0.5 mr, or 0.1 cm at a near viewing distance of 2 m (2110, Figure 16). Therefore, the dot width should be at least 0.29 cm and the edge-to-edge spacing between dots should be no more than 0.1 cm, for a dot-spacing/dot-width ratio of 0.34, thereby conforming with good legibility design criteria (Section 5.2.3).

Contrast ratio, linearity, and uniformity requirements are compatible with good alphanumeric legibility design (Section 5) and pose no particular technological problems. Similarly, the low data rate ($1\ s^{-1}$) is compatible with any technology.

2.3.2   Technology Selection

2.3.2.1   Status board

The difficulty with the present status board is that all information on it must be hand-written, reversed, on the back side. Errors are possible and legibility is compromised by hard writing. Further, the speed with which

information can be added or deleted is limited. Technological improvements therefore seem quite possible through either direct or projection display.

The most likely candidates for direct display are the EL or LED arrays. While EL arrays have only been made 1.6 m on a side thus far (Table 4), there seems to be no significant reason why a 2 m x 1.5 m panel could not be constructed. The EL power requirements can be met by ship's power, although heat dissipation must be carefully considered.

LED arrays have not been made this large either, and power requirements would probably preclude this selection.

Information density, contrast ratio, element sizes, uniformity, and linearity requirements present no problems to either the EL or LED technologies for this application. However, neither is capable of the four color presentation desired. With either of these technologies a monochromatic image is likely. Thus, a detailed analysis of color requirements would have to be conducted before either the EL or LED were selected.

A CRT projection system is another likely candidate. Such a system can present the full color spectrum, can project to the required size, and can meet contrast, luminance, linearity, uniformity, and resolution specifications. (Except for the relatively wide viewing angle, some of the "home" projection TV systems meet the requirements.)

Thus, a projection CRT provides color and all other performance specifications. It can be computer compatible, and would have the greatest flexibility for other information content, should the requirements change as data processing becomes more sophisticated. Its only possible constraint is the installation space required, but this appears to be available either above the crew or behind the status board (using rear projection).

Accordingly, the projection CRT seems to be the best candidate, and the EL array warrants consideration if color is not absolutely required.

2.3.2.2   CRT displays.

The present 23-inch (diagonal) CRT displays do not meet the size requirement of 51 x 56 cm, and are therefore of marginal legibility at 5 m. Addressing the size issue first, it is apparent that this can be met by CRT, EL, plasma, and perhaps EC technologies (Table 4).

All four of these technologies can meet the display content (Table 15), contrast ratio (Table 10), uniformity (Table 11), linearity (Table 11), refresh (Table 12), update (Table 13), and color (Table 6) requirements.

The plasma display might be discarded due to its near orange color. The CRT would be available with a white or green phosphor, the EL would be a yellowish-white, and the EC could be nearly any desired color.

There is really no logical basis for selecting among these three alternatives except cost. While all three can meet the design requirements, only the CRT costs on the order of $150 to $400, nonruggedized. The other displays would be well in excess of $1000, nonruggedized.

Thus, there appears to be no logical reason to replace the CRT technology with any other technology. However, the present 58.4-cm (23-in) CRTs should be replaced by a 75-cm (diagonal) CRT to obtain the desired 51 x 56 cm image size. This CRT should then have its line rate modified to about 145 lines/frame to obtain 132 active lines/frame. The display should be driven at a 60 Hz frame rate, noninterlaced. A video bandwidth of less than 5 MHz is adequate, well within current commercial capabilities. These modifications to the line rate and interlace, while not "off the shelf," are easily made by a competent video designer. The remaining modification would be to slightly defocus the CRT spot to an equivalent width of 0.29 cm, which is about 0.18 cm at half amplitude.

2.3.3 Summary

The procedure and the two application examples described above are realistic, though clearly not definitive. They are intended only to serve as illustrative simplified examples of how the systems/human factors engineer can make use of existing data to select or evaluate specific displays for

specific uses. Much more information is required to do that
job accurately and correctly. This information, summarized
in Sections 3, 4, and 5 following, should be used by the
systems/human factors engineer for any _specific_ application.

Section 3
TECHNOLOGY OVERVIEW

This section of the report summarizes the candidate flat panel technologies. Although the conventional cathode ray tube (CRT) and its several variants is neither solid state nor flat panel, it is also summarized at the end of this section as a baseline for comparison with the other technologies. Essentially the CRT remains the workhorse of the display applications field, has been adapted or modified to meet many requirements, is fairly reliable with an acceptably long life, and is produced quite inexpensively. In short, any successful technology must compete with or displace the CRT.

To provide a basis for comparison of the pertinent technologies, each of the display technologies is described relative to a set of 13 categories or parameters. These categories range from physical/engineering characteristics (size, configuration, power requirements, addressing requirements) through visual system pertinent variables (spectral emission, luminance, element size, element shape, contrast, uniformity, temporal characteristics) to cost, and include more subjective comments as to the utility of the technology for three categories of information presentation.

Finally, a future technology projection is offered for each category.

Lest there be some ambiguity of the meaning of any of these 13 categories or parameters, each will be defined as precisely as necessary for our present purposes.

3.1 PARAMETER DEFINITIONS

3.1.1 Physical Size and Configuration

This category describes the typical size and the range of physical sizes over which the display type can or may be fabricated. In some cases, the discussion refers to commercially available sizes, in other cases to potentially available sizes. In a couple of cases, limits to size are noted, as constrained by the inherent technology characteristics.

In addition, the basic physical configuration(s) of each technology is described so that the reader might appreciate the design limitations of the device on a parameter-by-parameter basis. No effort is made to present detailed quantitative design tradeoffs—such considerations fall into the bailiwick of the display designer. Rather, the present discussion is intended to give the reader an appreciation of the design tradeoff areas, so that available design data might be used better in device selection or system analysis.

3.1.2   Power and Voltage Requirements

While power and voltage requirements for a given device
or application are typically of greater concern to the
device designer, the user also must be aware of these
requirements. Often a given application may have some types
of power available (e.g., direct current) but not others.
Similarly, some applications have severe power limitations
(e.g., personally carried, portable display). Thus, total
power and voltage requirements remain the concern of the
human factors engineer or system engineer, as well as that
of the device designer.

3.1.3   Spectral Emission

The human visual system is not equally sensitive to all
wavelengths of visible light energy. Accordingly, wherever
possible the spectral emission is given in either radiant or
luminous energy per unit bandwidth. In keeping with current
scientific usage, bandwidth is expressed in nanometers (1 nm
$= 10^{-9}$ m).

The color spectrum may be described in several ways. For
example, visible light energy can be described in electro-
magnetic energy space as that portion of the electromagnetic
wavelength (or frequency) domain to which the eye is sensi-
tive, ranging approximately from 380 to 720 nm. Very narrow
wavelength bands produce "pure" colors, or any visually dom-
inant wavelength can be synthesized from other colors, in

accordance with the CIE Standard Observer chromaticity diagram.

For convenience, the x,y,z chromaticity coefficients, which define all spectral colors, are conventionally plotted in x,y coordinates, noting that $x + y + z = 1$. Subjective colors existing in various parts of the CIE space are labeled in Figure 17.
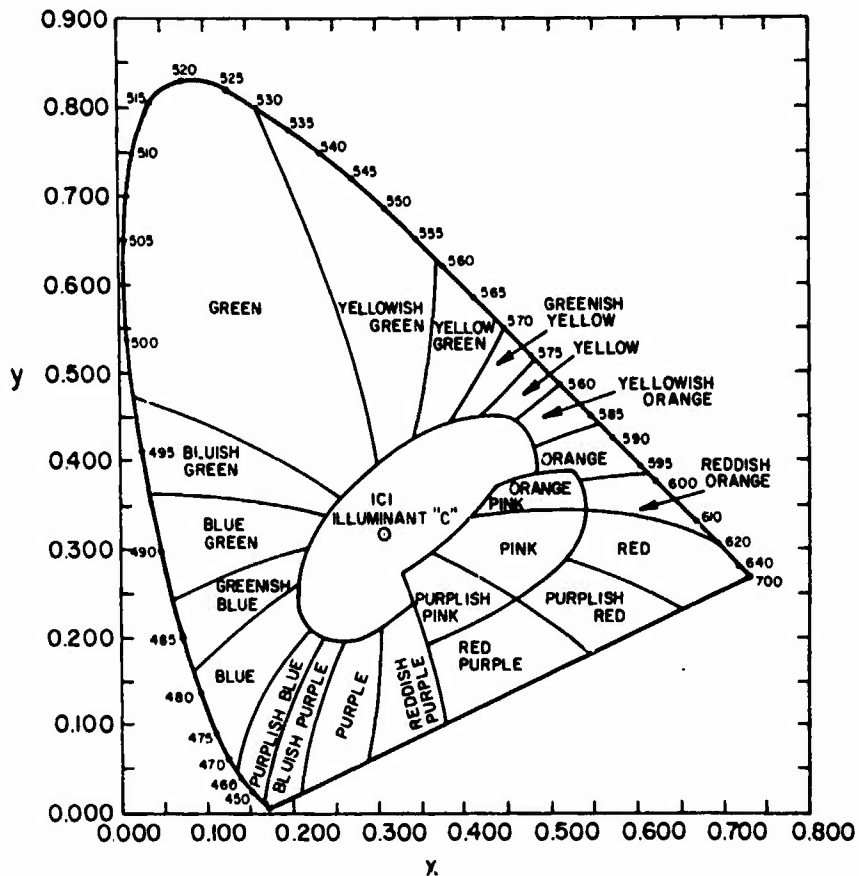
Figure 17: Subjective colors within the chromaticity diagram

It is often convenient to think of luminance as a dimension orthogonal to the x,y chromaticity diagram. For emis-

sive displays, luminance can be independent of the x,y coordinates of the display, subject only to the emissive properties of the display device. For a reflective display (e.g., liquid crystal), color is obtained by selective absorption. Thus, the maximum luminance (or maximum reflectance) occurs with white light, assuming a white light (x = y = z) ambient source. For selectively absorbing displays, greater absorption produces "purer" colors, at the expense of reduced luminance or reflectance. The maximum possible reflectance, as a function of x,y coordinates, is shown in Figure 18.
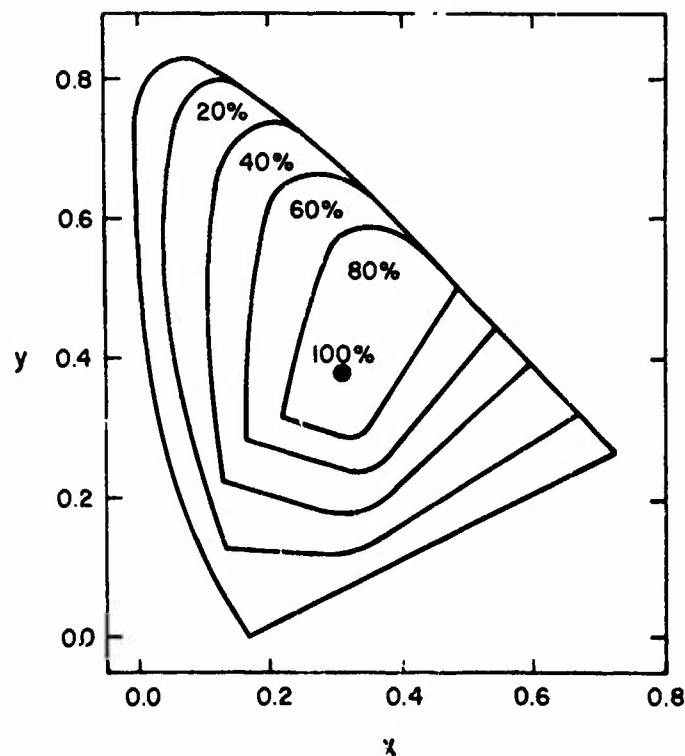
Figure 18: Contour curves of maximum luminous reflectance for materials illuminated by CIE source C (daylight)

These relationships will be referred to in later sections of this discussion.

3.1.4 Luminance

Because the visual system is not equally sensitive to all wavelengths of visible radiant energy, the radiant energy (in watts/nanometer, W/nm) must be weighted by the sensitivity of the eye to that wavelength. This sensitivity weighting function is termed the photopic luminosity function. The eye is most sensitive in the middle, or green section of the visible spectrum, and least sensitive at the extreme red (long-wave) and blue (short-wave) ends.

The weighting of radiant energy by the photopic luminosity function yields the physical measure of luminance, expressed in candelas/square meter (cd/m^2). Units often used, but not ISO units, include the foot-Lambert (ft-L), millilambert (mL), apostilb, and others. The cd/m^2 is commonly called the nit, and one foot-Lambert equals 3.426 cd/m^2.

It is important to note that the term "brightness" refers to the psychological perception of the stimulus or display, and is not a physical measure or property of a display surface. Thus, one cannot measure "brightness" in ft-L, cd/m^2, mL, or any other physical unit, in spite of this common erroneous usage in many otherwise accurate publications. "Brightness" is affected by spectral emission of both the

display and the surround, the visual adaptation state of the observer, and the luminance of both the display and the surround. Thus, it is a useful and meaningful measure in many contexts, but it must be determined through psychophysical methods, not by direct physical radiometric or photometric measurement.

For a thorough discussion of radiometric and photometric measurement, the reader is referred to Walsh (1965).

3.1.5 Luminous Efficiency

Not all devices produce the same amount of emitted radiant energy per unit of electrical power consumed. To permit comparison of devices in their efficiency to convert electrical energy into radiant energy, the concept of radiant efficiency, measured in radiant watts/electrical watts, is used. However, because all radiant energy is not equally visible to the observer, it is more meaningful to weight the emitted radiance by the photopic luminosity function, to yield the ratio of luminous energy to electrical power. This ratio, luminous efficiency, is expressed in lumens per watt, lm/W.

3.1.6 Element Size, Shape, Density

Flat panel displays are generally segmented in one of two forms. Some alphanumeric readouts are designed as seven-segment or starburst patterns (Figure 19), the selective addressing of which creates any given alphanumeric. The

- 67 -

other form is that of an element matrix, in which the elements are arranged in an X-Y array. By selective addressing of various elements in the X-Y array, one can create alphanumerics, symbols, line graphics, solid (shaded) areas, or even pictorial information such as a commercial TV image.
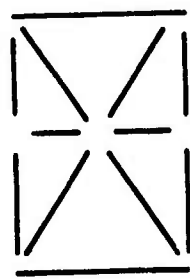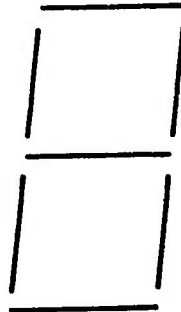
Figure 19: Seven-segment and starburst alphanumeric patterns

In recent years there has been rapid growth of the number of flat panel matrix designs. Simultaneously, the research required to relate human operator performance to the geometric properties of these designs has characteristically lag-

ged behind the hardware development activity. That research which has been completed to date, however, indicates that readability and legibility of X-Y matrix addressed displays are significantly affected by such geometric properties as individual element size, element shape, and spacing betwen elements (Snyder and Maddox, 1978). To date, no research has investigated the effect of the luminance energy distribution within the X-Y element, although it is reasonable to assume that this parameter may also affect legibility and readability.

Element size can be specified in appropriate display plane units such as diameter, length and width, or other dimensions. It may be also specified in angular subtense units at the eye. The former is more typical, although viewing distance must also be given to be useful for design evaluations.

Element shape can be specified in whatever terms are appropriate, such as round, square, rectangular, Gaussian, etc.

Spacing between elements is critical, and can be specified as edge-to-edge, or centerline. If element edges are poorly defined, center-to-center specification is more valid and meaningful.

3.1.7 Contrast and Dynamic Range

While display element luminance is important in display design, an equally important parameter is the contrast between any "on" display element and its "off" background. Unfortunately, the literature contains many definitions of "contrast," and many technical papers fail to specify which definition is intended. If the maximum or "on" luminance is symbolized as L_{max} and the background or "off" luminance is indicated by L_{min}, then the following definitions and relationships hold:

$$\text{Modulation (M)} = (L_{max} - L_{min})/(L_{max} + L_{min}), \tag{4}$$

$$\text{Contrast Ratio (CR)} = (L_{max}/L_{min}) = (M + 1)/(1 - M), \tag{5}$$

$$\text{Dynamic Range} = L_{max} - L_{min} = L_{max}(2M)/(M+1), \text{ and} \tag{6}$$

$$\text{Relative Contrast (C)} = (L_{max} - L_{min})/L_{min}$$
$$= (2M)/(1-M). \tag{7}$$

An often used, but typically misunderstood, concept in the literautre is "shades of gray." Usually, shades of gray refers to the number of just distinguishable increments in luminance that can be detected on a display. If so, the number of gray shades should be determined with a specified, appropriate psychophysical method. Unfortunately, the number of shades of gray attributed to a display is usually

estimated by inserting a gray step input and determining how many such increments an (often untrained) observer can detect. "Good" displays are said to produce 10 to 12 shades of gray.

An equally questionable procedure is also followed by assuming that a shade of gray increment is equal to a $(2)^{0.5}$ luminance increment. Since $(2)^{0.5} = 1.414$, this is equivalent to asserting that a just noticeable luminance increment is a 41.4% increase, which most visual psychophysicists would suggest to be quite erroneous. However, because the concept of modulation, M, is quite useful in display design and evaluation (as will be shown later) and because $(2)^{0.5}$ shades of gray are often reported, the following equation is noted:

$$(2)^{0.5} \text{ shades of gray, } N = 1 + \left[\frac{\log \dfrac{(1 + M)}{(1 - M)}}{\log (2)^{0.5}}\right], \qquad (8)$$

for L_{min} greater than zero.

3.1.8 Uniformity

Uniformity of a display can best be defined in its absence, or by nonuniformity. Following the general approach of Goede (1978), three important types of nonuniformity can be meaningfully distinguished: large area nonuniformity, small area nonuniformity, and edge discontinuities.

- 71 -

Large area nonuniformity is a gradual change in luminance (or color) from one area of the display to another, such as center-to-edge or edge-to-edge comparisons and gradients.

Small area nonuniformity pertains to element-to-element changes in luminance (or color) over small areas.

Edge discontinuity refers to changes in luminance or color over an extended boundary.

While the attempt by Goede (1978) to classify nonuniformities is heuristically helpful, it does not discuss suitable metrics of measurement. Still to be defined are "large area" and "small area," "changes in luminance," and the like. While one might easily suggest candidate metrics for evaluation (e.g., edge gradient, acutance, rms luminance), none has been tested or justifiably recommended. Sherr's (1979) excellent text on electronic displays includes no mention of uniformity in its representative specification parameter set (p. 38).

In this report, the subjective estimates of uniformity, physical measures (e.g., element failure rates), and applied specifications will be stated where they exist. However, the reader should note that there exists a real need for research to define these measures, to develop suitable criteria, and to recommend acceptable levels for representative observer tasks.

3.1.9 Temporal Characteristics

Some flat-panel displays have inherent memory, such that an element turned "on" will remain on until it is turned "off." Most technologies, however, have finite memory or persistence display elements, requiring periodic "refreshing" to avoid a perceived reduction in luminance over time or, more typically, a perceived flicker. To determine the required refresh rate to avoid perceived flicker, one must know the rise time and fall time of the luminance of the device.

Rise time is the time period required by the device to reach maximum luminance after the application of a square-wave "on" pulse or command. It is usually measured in microseconds (1 μs = 10^{-6} s) or milliseconds (1 ms = 10^{-3} s), as appropriate.

Fall time is the time, following cessation of the "on" pulse or command, for the luminance to reach 10% of its maximum value. It is also measured in μs or ms. The 10% value is used conventionally because decay or fall times are often exponential, making it difficult to ascertain the time to reach zero (or previous state) luminance.

3.1.10 Addressing/Driving Interfaces

A major consideration in any display system design is the electronic circuitry required to turn "on" and "off" any element or location on the display at the correct time and

with the proper luminance. This requirement is usually a determining factor in the success or failure of any display technology, for some technologies are more compatible with inexpensive, reliable addressing/driving concepts than are others. Very simply, "addressing" refers to the electronics that determine which element is turned on, whereas "driving" refers to provision of the right amount of power to the element to turn it on. For purposes of this discussion, addressing and driving requirements will be considered together, although it should be realized that as much, perhaps more, research and development activity is devoted to addressing/driving technology than is devoted to the display medium per se.

Segmented or starburst alphanumeric readouts typically present no serious problems in addressing or driving. Compatible power supplies exist, the number of segments to be addressed is relatively small, the logic of selection of the "on" segments is typically simple, and the total display refresh rate is not difficult to establish or meet.

Far greater problems exist with graphic displays, matrix-addressed character displays, and video displays. For such applications, the number of elements is usually large ($512 \times 512 = 2.6 \times 10^5$ is fairly typical), and addressing each element with a pair of leads is expensive, difficult, and often physically unattainable, at least in the size domain required for effective visual displays,

where there may be as many as 15-20 elements per square millimeter. Thus, the compatibility of any flat panel display technology with addressing/driving circuitry is critical to its potential utility. For this reason, somewhat detailed attention will be given to alternate addressing modes and concepts.

As a basis for comparison, the conventional raster scanning of the home television addresses over 3×10^6 picture elements per second, and high resolution television displays are available with the capability to address over 50×10^6 picture elements per second. No matrix addressing technique has achieved a comparable rate.

Consider a 512 x 512 element matrix. If each element is connected to a common ground, then there would need to be $(512)^2 = 2.6 \times 10^5$ wires, one per element, to address each element directly. This number of wires is not physically possible in the usual display size. However, it is possible to have one wire for each row and column, provided the row and column densities do not exceed about 4/mm. Thus, the 512 x 512 element matrix would require only 512 + 512 = 1024 wires. By this form of X,Y matrix addressing, a picture element is selected (addressed) by selecting a pair of orthogonal wires. As shown in Figure 20, when a suitable voltage is applied across this pair of wires, current flows through the picture element, and the element is turned "on" (Tannas, 1978).
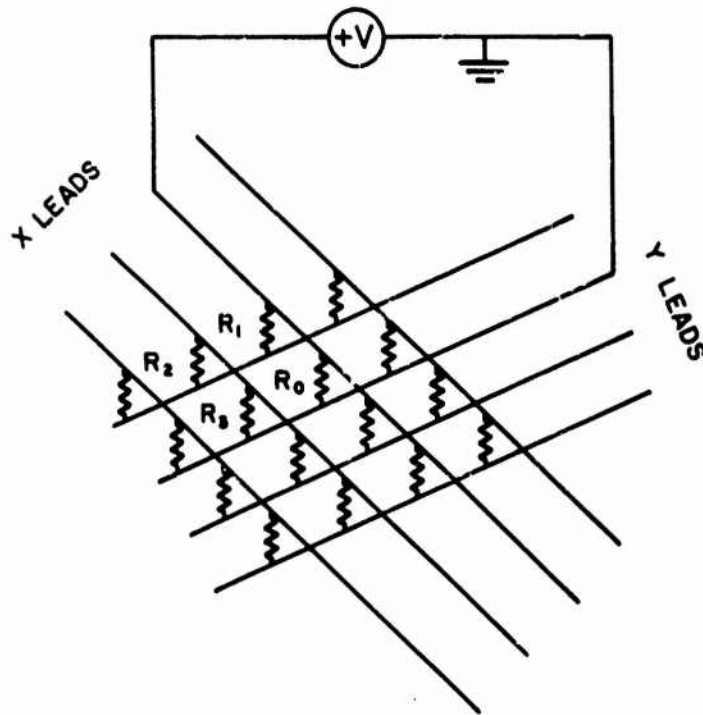
Figure 20: X-Y addressing scheme, from Tannas (1978)

For this typical two-wire addressing, each element has
one wire common to all other elements in the same row, and
the element has the other wire common to all other elements
in the same column. Row wires are usually on one side of
all picture elements, and column wires on the other side, as
in the two sides of a sandwich.

Unfortunately, selection of any given element by this
type of X,Y addressing also permits the unintentional selec-
tion of other elements. In Figure 20, the addressed element
is indicated by resistor R_o. However, nothing prevents cur-
rent flow from the selected X lead through element R_1, down
its Y lead, through R_2, across its X lead, and through R_3 to
the addressed Y (ground) lead. This path through R_1, R_2, R_3

is called a "sneak path." While the current through R_o, by Ohm's law, is V/R_o, the current through each of resistive elements R_1, R_2, R_3 is $V/(R_1 + R_2 + R_3)$, assuming the display device has a linear impedance. Figure 21 illustrates the effect of this sneak path on the nonaddressed cells, which are "on" but with less luminance, assuming the display material's luminance is linear with voltage, as are many flat panel devices. Under these circumstances, the sneak path addressed elements are at one-third the luminance of the directly addressed cells, and the maximum luminance modulation is:

$$M = (1 - 1/3)/(1 + 1/3) = 0.50. \tag{9}$$

This is termed the "half brightness level" and is typically unacceptable for most applications.

Fortunately, several flat panel technologies are not linear in luminance response. Examples are those showing hysteresis (plasma displays) or a threshold requirement (electroluminescent), as indicated in Figure 22. These types of devices are highly compatible with matrix addressing.

The above described problem is, of course, greatly complicated when two or more elements are selected on the same row or column. In this case, multiple sneak paths exist, and when many elements are simultaneously addressed, the numerous sneak paths essentially cause the entire image to be washed out.
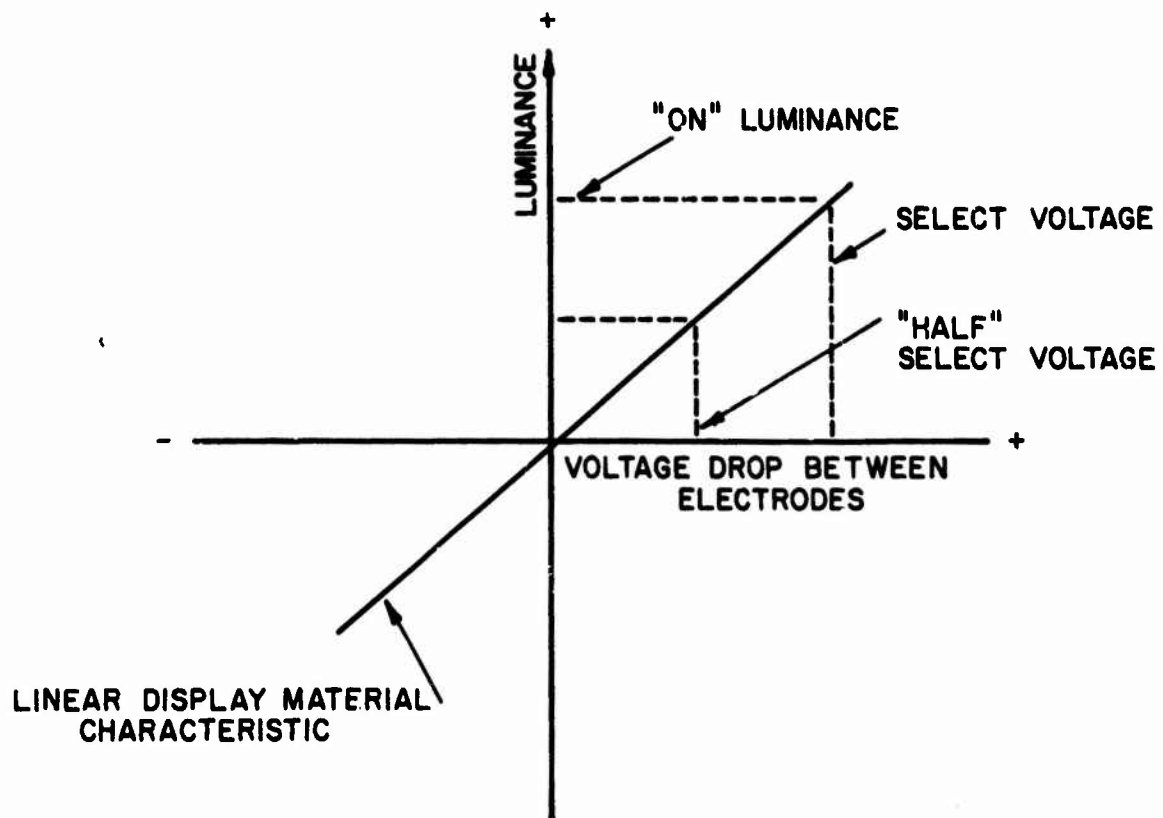
Figure 21: Voltage applied to sneak path cells, after Tannas (1978)

Two techniques are used to permit matrix addressing to work. First, each picture element must be nonlinear in response, so that the sneak path reduced current will not turn it "on" at all. Secondly, elements are addressed in a time-sharing or multiplexing sequence so that no two elements are addressed simultaneously. Thus, each element in the matrix is addressed only during a short period of time, and then the next element is addressed. Fortunately, as shall be seen later, some technologies respond favorably to this multiplexed, short duty cycle addressing.

In sum, several of the flat panel displays to be discussed later can be matrix addressed with success, and with acceptable results. The techniques work, but the cost can

HYSTERESIS
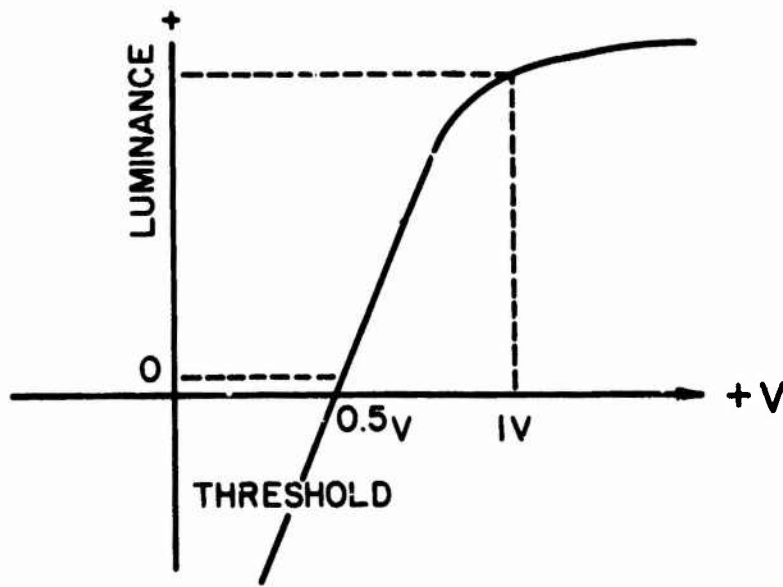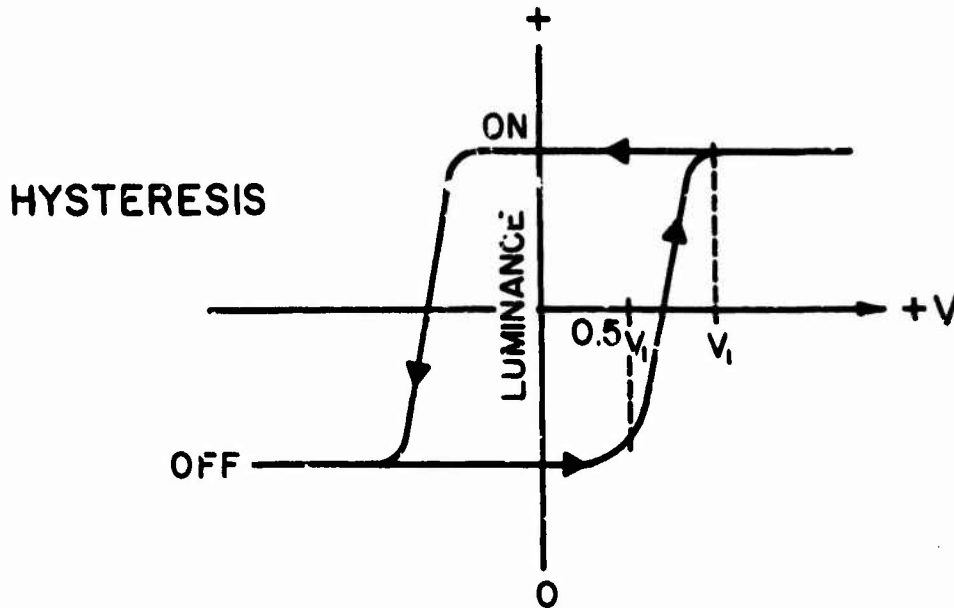
Figure 22: Nonlinear display thresholds, after Tannas
(1978)

be high. Tannas (1978) concluded that "The battle is won in
matrix addressing a flat panel array. But in all cases, so
far, the war is lost as the resulting complexity is too
costly to compete with the cathode ray tube" (p. III-2).

This is not totally true, as some flat panel displays have been costly and yet sold in thousands of copies. Others are improving and becoming less costly. The battle is still on.

3.1.11 Cost

The cost of flat panel displays is probably the most critical factor. Most technological problems have been or can be solved; however, the production costs far exceed the cathode ray tube for most graphics and video applications. To permit the reader some estimate of cost sources and comparisons, rough cost estimates are provided wherever possible.

3.1.12 Utility for Display-Type Applications

For convenience and directness of evaluation, four generic types of display usage or application are considered in this report. While these four types do not cover all possible applications of flat-panel displays, they do cover the range of difficulties and permit consideration of problems applicable to most military and civilian design requirements.

The least demanding type of display, from a technology standpoint, is that required for alphanumeric readout of data and messages. Information presented on this type of display is usually constrained by a given alphabet of letters, numerals, and symbols. Further, the display is con-

strained by a discrete set of locations in which any alphabet character can be located. Incandescent, mechanical, and vacuumfluorescent devices currently dominate this type. Characters displayed can be composed of segmented, stroke, raster scanned, or dot matrix elements.

Graphics (or vectorgraphics) requirements describe a more complex type of display. Graphics displays require the ability to position curved and straight lines essentially anywhere on the display, to move those lines individually and independently at often rapid rates, to shade various display areas, and to overlay alphanumeric information on the graphics information.

A most demanding requirement of a display is to present video information, as in a television or radar display. This category typically demands the greatest resolution, rate of motion, levels of distinguishable intensities, and (potentially) color, with no noticeable flicker or blur.

Last, each technology will be evaluated for its ability to be used in a large-screen, or group viewing situation. We will consider a "large screen" display to be that required for more than three simultaneous viewers, normally a display in excess of 0.5 meter on a side. Some large screen displays are directly viewed, while others are obtained by a projection technique. This distinction will be drawn for each technology type.

As will be seen, hard quantitative requirements do not exist in detail for these four types of displayed information for all applications. However, sufficient human operator performance data and design guidelines are available to permit a comparison of the various display technologies to meet generic needs to each of these display types.

3.1.13 Future Technology Projections

For each technology, and where available, information is given on the future directions of research and development. Areas of improvement critical to meeting various application requirements, performance criteria, or reducing costs are noted.

3.2 TECHNOLOGY SUMMARIES

3.2.1 Cathode-Ray Tube (CRT)

For a great majority of data and imaging applications, the CRT dominates the market. Part of its popularity is due to cost, and part is due to a long lasting familiarity. Even among knowledgeable system designers, however, the CRT is chosen for numerous applications because of its tremendous flexibility. CRTs are available in a variety of sizes and shapes, provide gray scale and color, can have reasonably good resolution, can provide a storage capability, and can be addressed with both raster and stroke patterns.

Information on CRT display capabilities is readily available from many sources and will not be summarized here for that reason. However, for comparison purposes, the CRT will be listed in the summary tables at the end of this section. For additional information, the interested reader should consult such sources as Sherr (1979), Herold (1974), and Luxenberg and Kuehn (1968).

3.2.2   Flat-Panel CRT

The conventional CRT has great flexibility in information display, but also a few substantial disadvantages. In many applications, a major disadvantage is its depth--as the displayed image size is increased, so is the length of the tube. Primarily for this reason, considerable effort has been placed on the development of a CRT having much less depth. This section summarizes the present status of that "flat" CRT technology.

3.2.2.1   Physical size and configuration

The flat panel CRT concept is best illustrated by the Northrop Corporation's development of the DIGISPLAY, illustrated in Figure 23 (Goede, 1973). The electron area source is a cathode which is less than 12 mm thick, and consists of a number of cathode elements requiring fairly low power. The modulation plate controls the electron beam current from the cathode, much as the control grid does in a conventional CRT.
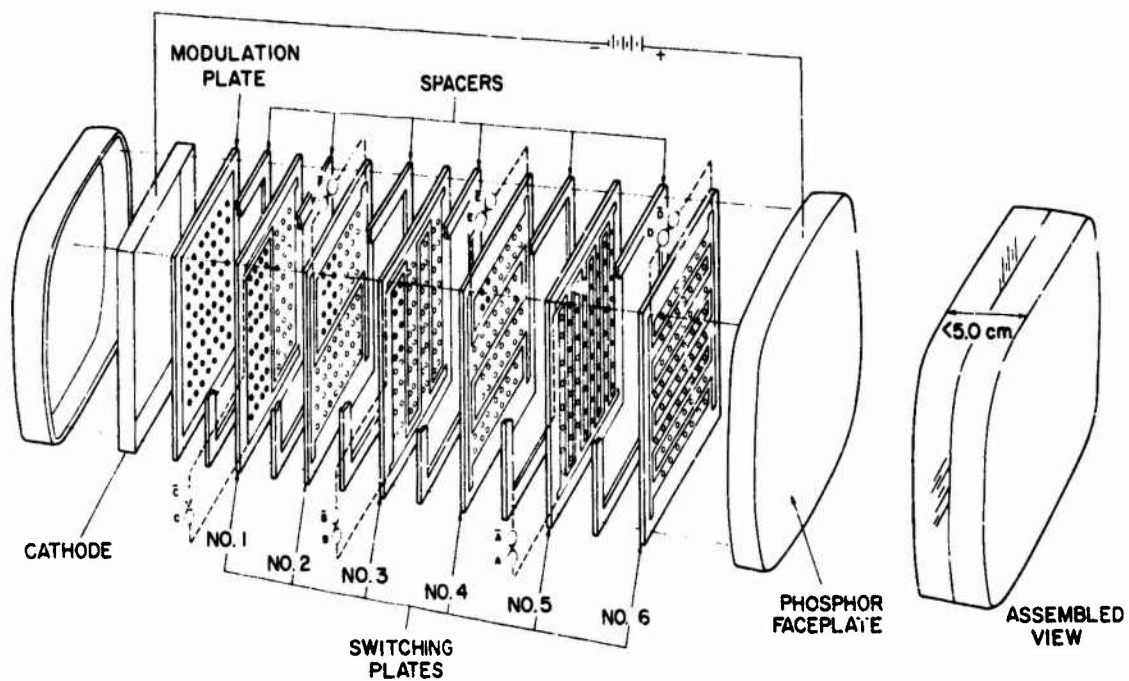
Figure 23: Schematic drawing of flat-panel CRT construction, from Goede (1973)

The modulation plate is followed by a series of switching plates, each of which has an array of channels ("holes") which pass electrons. These switching plates accomplish two functions: (1) they keep the electron flow in well defined channels or directions, and (2) they either pass or stop the flow of electrons in a given area, by voltage addressing of each plate.

Dielectric spacers between the switching plates provide structural strength. Lastly, the electron streams impact on a monochrome or multicolor phosphor to provide a modulated image much the same as in a conventional CRT.

The most efficient control of the switching plates is by having two leads to each plate. Then, in sequence, the electrons emerging from the cathode are either passed or blocked in a binary logic by each succeeding switching plate. The number of addressable elements on the display is 2^N for N switching plates. Thus, an 8 x 8 = 64 element DIGISPLAY would require N = 6 (2^6 = 64) switching plates, as illustrated in Figure 24 (Goede, 1973). Each successive plate, if energized, increases by two the number of elements that might be individually selected to be "on." Thus, to turn on only the upper left element of the display, there is only one combination of commands to all six plates, if each plate has two leads.

6-2 LEAD PLATES

3-4 LEAD PLATES
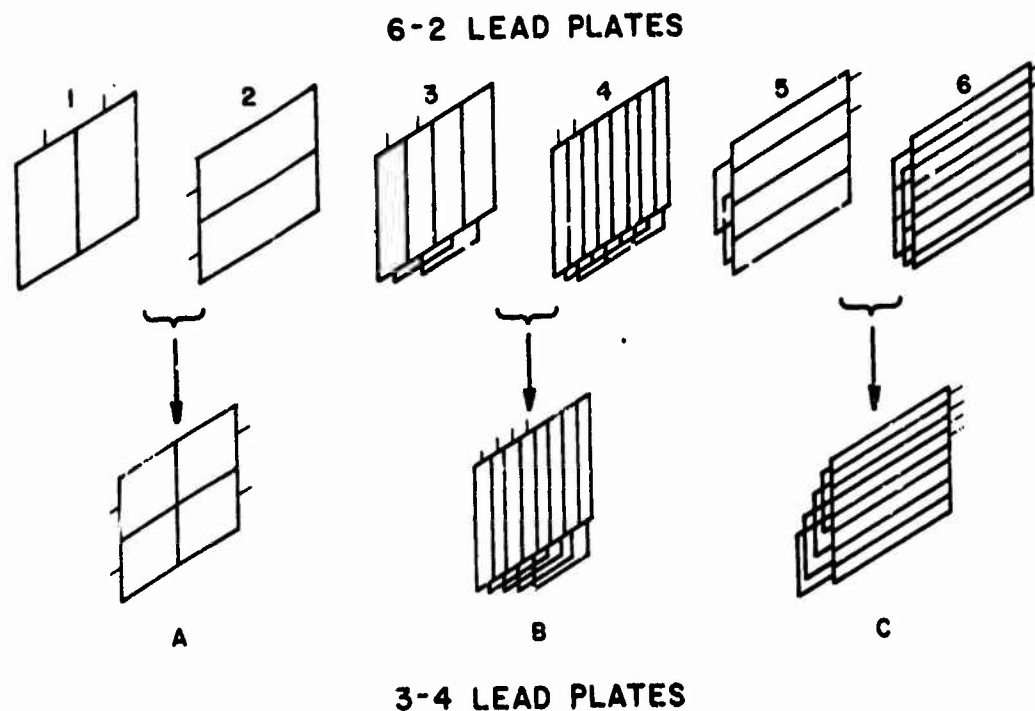
Figure 24: Alternative addressing techniques for flat-panel CRT, from Goede (1973)

Because the switching approach can be designed to "write" in more than one place at a time, the DIGISPLAY can be used in a multi-beam mode, which generally results in greater mean display luminance (Goede, 1973).

Due to the parallel plate design of this device, there is no particularly critical size limitation. To date, functioning DIGISPLAY devices have ranged in size from 7.37 by 11.68 cm to 16.26 cm square. Larger panels could be fabricated if the need arose. Thickness will generally be less than 5 cm.

3.2.2.2 Power and voltage requirements

Sherr (1979) indicates that power dissipation of the flat-panel CRT can be calculated by:

$$P_b = V_c I_c + V_i I_i + V_p I_p, \qquad (12)$$

in which

$$P_D = \text{beam power,}$$
$$V_c = \text{collector voltage,}$$
$$I_c = \text{collector current,}$$
$$V_i, I_i = \text{plate voltage and current,}$$
$$N = \text{plate number, and}$$
$$V_p, I_p = \text{phosphor voltage and current.}$$

This can be simplified to:

$$P_b = V_p I_p, \qquad (13)$$

because all other currents are negligible (Sherr, 1979).

Because the switching approach can be designed to "write" in more than one place at a time, the DIGISPLAY can be used in a multi-beam mode, which generally results in greater mean display luminance (Goede, 1973).

Due to the parallel plate design of this device, there is no particularly critical size limitation. To date, functioning DIGISPLAY devices have ranged in size from 7.37 by 11.68 cm to 16.26 cm square. Larger panels could be fabricated if the need arose. Thickness will generally be less than 5 cm.

3.2.2.2 Power and voltage requirements

Sherr (1979) indicates that power dissipation of the flat-panel CRT can be calculated by:

$$P_b = V_c I_c + V_i I_i + V_p I_p, \tag{12}$$

in which

P_b = beam power,

V_c = collector voltage,

I_c = collector current,

V_i, I_i = plate voltage and current,

N = plate number, and

V_p, I_p = phosphor voltage and current.

This can be simplified to:

$$P_b = V_p I_p, \tag{13}$$

because all other currents are negligible (Sherr, 1979).

In general, Sherr (1979) states that the cathode power is about 100 mW/cm^2, and the total power therefore ranges from 2 to 30 W. Goede (1973) described four different DIGISPLAY devices in which the beam power ranged from 0.65 to 8.7 W/cm^2, and total tube power from 1.6 to 27 W. The mean required display luminance levies the major burden on power consumption.

3.2.2.3 Spectral emission

The phosphor screen may be chosen to be any of the standard phosphors. Thus, a variety of spectral emissions is possible for monochrome displays.

Multi-color displays are obtainable with two techniques: (1) the beam penetration phosphor coating, and (2) a tri-color dot phosphor screen (Goede, 1973). The spectral range of the penetration phosphor approach will have the same limitations as those of the conventional CRT; i.e., color saturation will be limited by partial inadvertent activation of the nonaddressed phosphor.

The color triad approach of the shadow-mask CRT has some advantages when applied to the flat-panel CRT. In this design, the color dots are deposited in red, green, and blue triads, and each color is addressed through a separate hole or channel in the final switching plate. Thus, three-beam simultaneous addressing can be used to avoid high single-beam scanning rates and reduced display luminance. This

approach, while theoretically feasible, has yet to be adequately demonstrated, perhaps because it requires three times the number of leads needed for the monochrome display.

3.2.2.4 Luminance

The display luminance decreases as the number of elements scanned per unit time increases because the dwell time per element increases, much as is the case with a conventional CRT. A means to combat this reduction in luminance is to use multiple-beam scanning. The four devices described in Goede (1973) have from 13 to 40 beams, and result in spot luminances ranging from 103 to 822 cd/m^2. Since no data are available on spot luminance distribution or percent active area, it is difficult to estimate average display luminance levels, but it appears that 500 to 800 cd/m^2 are achievable. However, Goede (1978) estimated the flat-panel CRT luminance at 103 cd/m^2.

3.2.2.5 Luminous efficiency

The only estimate of luminous efficiency for flat-panel CRTs is 2.0 lm/W, offered for a 512 x 512 element TV display by Goede (1978). This is, of course, much less than that of a conventional CRT with a P4 phosphor, which typically emits about 44 lm/W, or a P22 (color) TV CRT emitting about 8 lm/W for a white image.

3.2.2.6 Element size, shape, density

Monochrome flat-panel CRTs are limited in element density to the accuracy with which channels in the switching plates can be aligned. To date, the highest density flat-panel CRT has 3.15 elements/mm (Goede, 1973) in a 512 x 512 element TV/graphic DIGISPLAY. It is believed that element shapes will be Gaussian in both directions. Dot overlap is not known, but will obviously depend on channel spacing in the switching plates.

3.2.2.7 Contrast and dynamic range

Goede (1973) claims contrast ratios for four different DIGISPLAY models to range from 25:1 to 100:1 minimum. Presumably these are the ratios of element-on to element-off luminances under zero ambient illuminance. The only intrinsic limitations to dynamic range are phosphor scattering and maximum beam current, which is limited by heat dissipation capabilities.

There are two means by which luminance levels are controlled in the flat-panel CRT. In amplitude modulation, the voltage applied to the modulation plate controls the beam current, as in a conventional CRT. If more than one beam is used, the modulation plate simultaneously modulates all beams at the same level.

The luminance can also be controlled digitally through pulse-width modulation. In this case, the "on" time for

each element is time modulated, thus controlling luminance by the on time at each element location.

Goede (1973) has demonstrated a DIGISPLAY with seven $(2)^{1/2}$ shades of gray, and claims that sixteen $(2)^{1/2}$ shades are possible, which would require a dynamic range of 256:1.

3.2.2.8 Uniformity

No data are available on any form of flat-panel CRT uniformity. However, some discussion is appropriate, based on the design concept.

First, the distance from the cathode to the phosphor should be fairly constant for the entire display surface. For this reason, there should be no reduction in luminance or increase in spot size from one display location to another. Of course, cathode uniformity must be accurately controlled.

Secondly, the geometric distortion should be negligible, and is limited only by the positional accuracies of the switching plates, which should be very good.

Third, element failures, row/column failures, and large area failures can all occur if there is a failure in the switching plate circuitry or the plates themselves. However, since no production units have been made as yet, no reliable estimates of failure types and failure rates are available.

Thus, the flat-panel CRT has many of the uniformity features which exist for solid-state panels of other types. Image and geometric stability should be very good, and luminous uniformity should be excellent.

3.2.2.9 Temporal characteristics

Since the rise and fall times of the luminance of addressed elements will result directly from the phosphor characteristics, it is expected that the temporal characteristics of the flat-panel CRT will be similar to those of the conventional CRT; that is, much depends on the phosphor selection.

However, it is possible to add a "storage" target to the flat-panel CRT just as one does in creating a conventional storage CRT. These principles are well understood (Sherr, 1979) and need no repeating here. In using the storage tube approach, the dwell time per element is increased by several orders of magnitude (Goede, 1973), but with considerable sacrifice of dynamic range and total loss of gray scale capability. Thus, the approach is acceptable for alphanumeric and graphic static information display, but not for dynamic video images.

3.2.2.10 Addressing/driving interfaces

As noted earlier, the driving voltages of the flat-panel CRT are in the tens of volts and present no particular prob-

lems. The addressing requirements vary with the plate/lead

scheme. The simplest scheme is to use two leads per plate,

which requires only binary addressing logic. However, as

the number of elements becomes large, so does the number of

plates. Thus, a tradeoff is made between plate number and

addressing complexity. For a 512 x 512 element graphic/TV

display, Goede (1973) used 96 drivers, which does not seem

unwieldy.

3.2.2.11   Cost

No production runs have been made for flat-panel CRTs.

In 1973, Goede (p. 7) estimated that

> all processes involved are similar to processes
> used in other low-cost devices. Current cost pro-
> jections indicate that the DIGISPLAY can be cost
> competitive with the CRT for many applications.

Somewhat thereafter, the Northrop Corporation sold all

rights and technology to Texas Instruments Corporation who,

according to Sherr (1979, p. 150), has

> embarked on a extensive program of research and
> development in order to achieve a fully practical
> device.

It is this author's understanding that the TI R&D effort has

been terminated because production cost estimates suggest

the flat-panel CRT will not be competitive with other tech-

nologies.

3.2.2.12 Utility for display-type applications

The flat-panel CRT has been shown to meet most requirements for alphanumeric displays, graphics, and television. Goede (1973) has presented characteristics of models which meet these needs, ranging from a low-power radar display to a 512-character (5 x 7 font) alphanumeric display to a 512 x 512 element TV/graphic display. Data rates, dynamic range, and element densities appear generally compatible with these needs. The flat-panel CRT is not a likely candidate for very large displays, however, because of cost considerations.

3.2.2.13 Future technology projections

At the present time, additional development of the flat-panel CRT technology is hampered by a predicted high cost in production quantities. While the perceived need for a flat-panel video/graphic capability is generally accepted by most people in the display engineering community, it is also generally thought that other technologies will prove to be more economical.

If any future development takes place, areas of emphasis are likely to be full-color development, cost reduction, and fabrication of larger size displays (e.g., 1024 x 1024 elements at 3 to 4 elements/mm). At this time, however, the flat-panel CRT appears to hold very little interest in the display community.

3.2.3 Light-Emitting Diode (LED)

The light-emitting diode is a basic example of the concept of electroluminescence, and is probably the most widely used and successful form developed to date. The LED has been used successfully in calculators, wristwatches, instrumentation readouts, and discrete miniature lamps. Its popularity is based upon the combination of good luminance, low cost, low power, high reliability, and good compatibility with integrated circuit technology. More recently, larger arrays of LEDs have been developed for message readout and graphics displays, although such applications and devices must be considered to be in the development stages.

The basic form of the LED is a two-terminal semiconductor device that produces visible energy by voltage application to a forward-biased P-N junction. (A good basic review of the physics principles involved is given by Sherr, 1979). Figure 25 illustrates the basic concept. The most typical device has a substrate of either gallium phosphide (GaP) or gallium arsenide (GaAs), with GaP leading to greater luminous efficiency because it can be made optically transparent to the LED emission. As the forward voltage is increased, the LED behaves like a classic diode, with a sharp knee and then linear current flow roughly proportional to the forward voltage in excess of approximately 1.5 V. Figure 26 illustrates this relationship, which is of fundamental importance since the luminous output is approximately proportional to the forward current over the linear range.
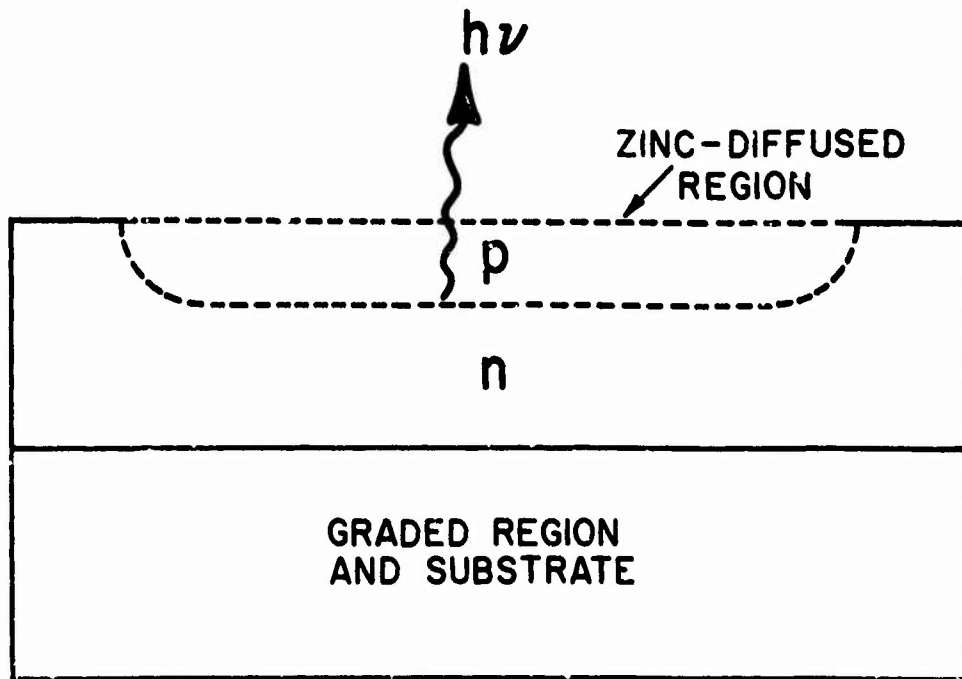
Figure 25:   Cross-section of basic LED, after Goodman (1975)
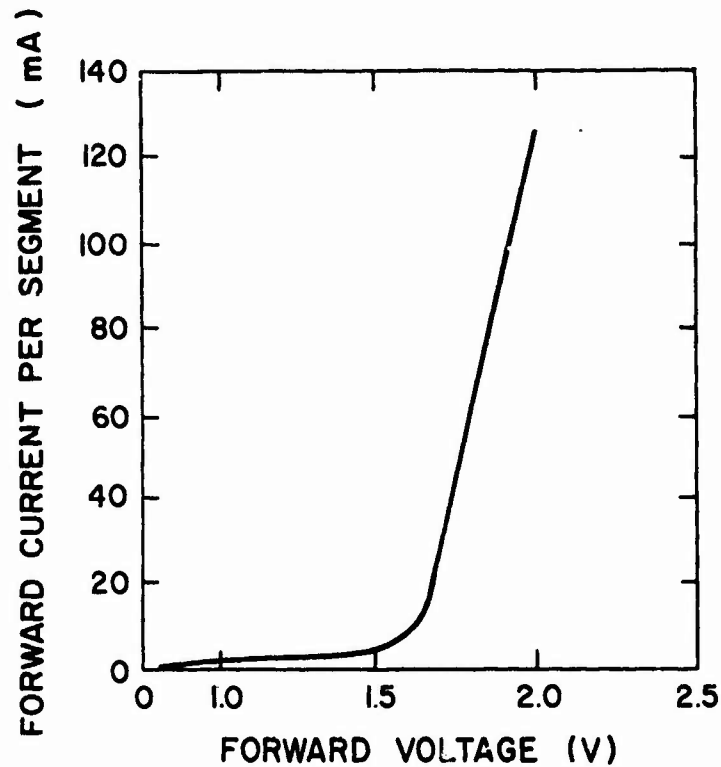
Figure 26:   Current-voltage characteristics in forward-bias direction for an LED, after Goodman (1975)

3.2.3.1 Physical size and configuration

LEDs are available in a variety of sizes and configurations. A common form is the discrete status (ON-OFF) indicator using a sealed and light-filtered output. Small discrete elements of this type often have an integral lens to disperse the output energy over a wide viewing angle.

Pocket calculators and wrist watches have created a heavy demand for LED segmented numeral readouts, in which each of the seven segments in the numeral is a separate P-N junction, controlled by external logic and switching circuitry. Arrays of discrete LED dots have also been used in pocket calculator displays to provide a more flexible readout form than is available from the geometrically restrictive seven-segment character.

Finally, larger two-dimensional LED arrays are available, in which the LED elements are part of a two-dimensional monolithic pattern, addresed in matrix form by external logic circuitry.

A cross-section of a gallium arsenide phosphide (GaAsP) array discrete element is shown in Figure 27. In this version, the substrate is GaAs upon which a GaAsP layer has been epitaxially deposited. The P-region has been diffused (doped) with zinc. The process used is similar to that employed in making silicon integrated circuits. In one example of this approach, the elements are 0.38 mm square, spaced on 0.51 mm centers to give approximately 2 ele-

- 97 -

ments/mm. Thus, a typical 512 x 512 element panel would occupy approximately 260 mm x 260 mm.

LIGHT
OUTPUT

GOLD WIRE

EPOXY

CONDUCTOR
CABLE

ZINC DOPED
P-TYPE

GALLIUM ARSENIDE PHOSPHIDE

GALLIUM ARSENIDE

LED JUNCTION

GOLD
METALLIZATION

COPPER HEAT SINK
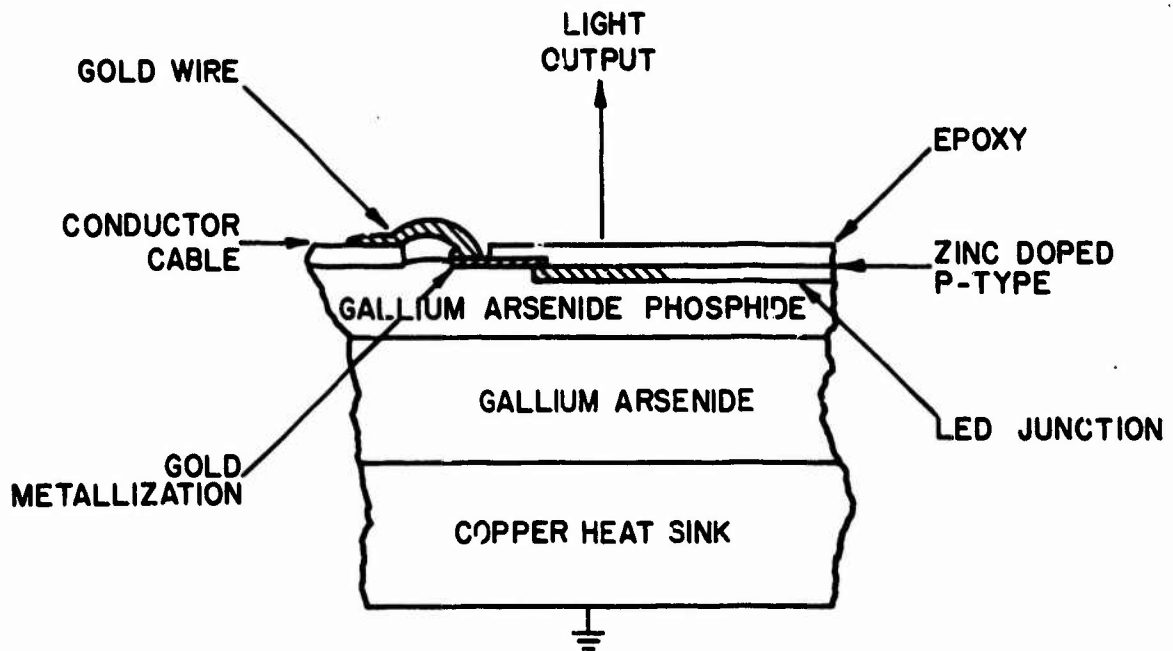
Figure 27: Cross-section of GaAsP array discrete element, from Scanlan and Carel (1976)

Smaller elements and spacings are also possible. Texas Instruments has developed, for example, a linear (one-dimensional) array with an element size of about 0.05 mm on 0.1 mm centers, for approximately 10 elements/mm (or 250 elements/inch). Even smaller elements and spacings appear feasible, should the need arise.

3.2.3.2 Power and voltage requirements

A significant advantage of the LED, in either single element or array form, is its low power consumption and voltage

compatibility. Requiring between 1.5 and 5 VDC, the LED is highly compatible with solid-state circuitry, especially TTL (transistor-transistor logic) and LSI (large-scale integration) circuitry.

Typical power requirements are 1 to 5 mW/element for moderate luminance levels. Because output luminance is directly proportional to current through the P-N junction (and therefore to power dissipated), greater power requirements are attached to increasing luminance. For example, in high ambient environments, power requirements can be as high as 10 to 30 mA per element. At 1 mA/element and a switching voltage of 2 V, a 512 x 512 element array, fully illuminated, would require over 130 amperes of current, an impossible amount in many applications. Under high ambient conditions, requiring 10 mA/element, the same array would draw 1300 amperes or 2.6 kW. Even if such power were available, structural limits on heat absorption and heat sinking requirements would become prohibitive.

3.2.3.3 Spectral emission

LEDs are available in red, orange, yellow, and green. The standard red GaAsP on a GaAs substrate emits light at about 655 nm, a fairly deep red. The use of nitrogen-doped GaAsP on GaP substrates shifts the dominant wavelength toward the orange, or about 635 nm. Elimination of the nitrogen doping shifts the dominant wavelength towards the red end and reduces the luminous efficiency.

Orange, yellow, and green LEDs are possible with GaAsP on
a GaP substrate, although GaP on a GaP substrate will pro-
duce the highest intensity green. Control of the dominant
wavelength, ranging from orange to green, for the GaAsP/GaP
compound is obtained by varying the amount of GaP. Large
amounts of GaP produce green (565 nm) emission, while rela-
tively small amounts shift the emission toward red (at 650
nm). Spectral coverage is continuous throughout this range.

Another way to achieve chromatic variation is by dual
doping such that the emission wavelength varies with current
density. Litton Data Systems (Scanlan and Carel, 1976, p.
155) has developed a $(152 \text{ mm})^2$ panel of 132 x 132 dual-doped
elements that range from red to orange to green as the cur-
rent density is increased from 0.7 to 3.3 mA/element.

A summary of the spectral emission and luminance charac-
teristics of a variety of LEDs is given in Table 2.

3.2.3.4 Luminance

Inherent in the LED is a linear increase in luminance
with increasing current, assuming uniform doping. Scanlan
and Carel (1976) reported this nearly linear relationship
for two GaAsP numeric indicators, as illustrated in Figure
28. This figure also indicates an important characteristic
of LEDs, namely that the output luminance can be signifi-
cantly increased beyond the DC-excited level if the diode is
pulsed repetitively, at the same average power input.

TABLE 2

Luminance and Spectral Characteristics of Representative
LEDs (adapted from Bhargava, 1975)

LED	Color	Peak Wavelength (nm)	Luminous Efficiency (l/W)	Luminance (l/Acm^{-2})	Commer. Avail.
GaP:Zn,O	red	699	3.0	350	yes
GaP:N	green	570	4.2	470	yes
GaP:NN	yellow	590	0.45	---	yes
GaAs$_{0.6}$P$_{0.4}$	red	649	0.38	---	yes
GaAs$_{0.35}$P$_{0.65}$:N	orange	632	0.95	---	yes
GaAs$_{0.15}$P$_{0.85}$:N	yellow	589	0.90	---	yes
Ga$_{0.7}$Al$_{0.3}$As	red	675	0.48	140	no
In$_{0.42}$Ga$_{0.58}$	amber	617	0.28	470	no
SiC	yellow	590	0.01	10	no
GaN	blue	440	---	---	no
GaN	green	515	---	4000	no
GaAs:Si with YF$_3$YbEr	green	550	0.60	---	no
GaAs:Si with YF$_3$:Yb:Tn	blue	470	0.006	---	no
InSe	yellow	590	---	---	no

Luminous output can also be increased by adding a silver
or reflective backing to the LED, or by adding optics to the
single element to increase directional emittance. The upper
LED luminance limits are not totally defined, and will
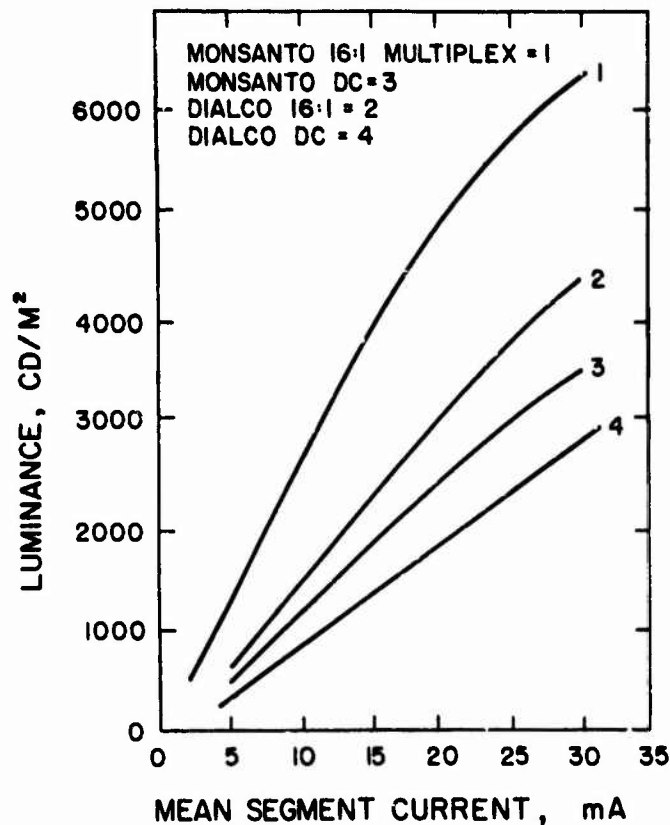
Figure 28: Luminance versus current for four LEDs, from Scanlan and Carel (1976)

depend on future research and development. At the present time it is possible to achieve over 68,500 cd/m^2 luminance from a single GaAsP LED with an instantaneous drive current of 1 ampere.

3.2.3.5 Luminous efficiency

Because the sensitivity of the visual system varies by nearly 100:1 from the green LED wavelength (560 nm) to the red LED wavelength (650 nm), we must consider the luminous output/power input ratio, or luminous efficiency, of the several available LED types. Figure ?9 illustrates the

luminous efficiencies of GaAsP/GaAs, GaAsP, GaP, and GaAsP:N/GaP (Allan, 1975). As illustrated, the nitrogen-doped GaAsP:N/GaP compound is most efficient, witn its maximum efficiency in the red end of the spectrum, even though the eye is less sensitive to the red wavelengths. Thus, the red LED emits over 65 times as much radiometric energy at 630 nm than it does at 570 nm, more than compensating for the 20:1 loss in visual sensitivity between these two wavelengths. Table 2 gives the luminous efficiency for a variety of LEDs.
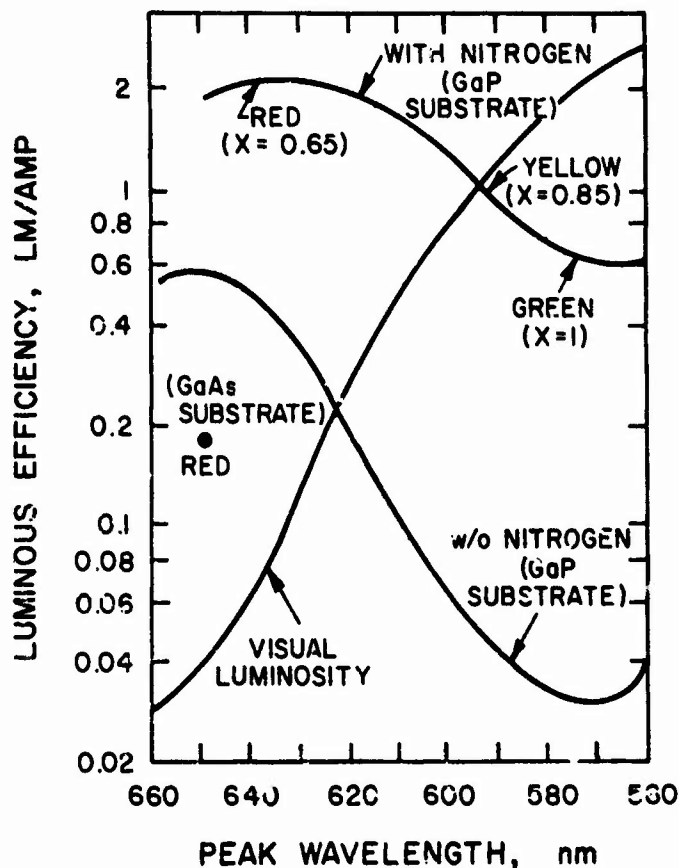
Figure 29: Luminous efficiency of several LEDs, from Scanlan and Carel (1976)

3.2.3.6   Element size, shape, density

The optical and geometric characteristics of the LED depend considerably on the materials used and the fabrication techniques employed. For these reasons, and because there are many types of LEDs currently in the experimental and developmental stages, it is not possible to define parametric limits for the size, shape, and density of elements in LED arrays. However, some existing examples can be described and extrapolations from these examples can be offered.

For discrete element (e.g., OFF-ON) indicators, LEDs are commercially available to a 0.5 mm diameter or less, and are usually fitted with an integral lens to achieve optical dispersion. The geometric challenge of LED technology is not in small discrete indicators, however, but in fabrication of larger one- and two-dimensional arrays of high density to be acceptable for message readouts, graphic displays, and possibly video presentation.

Current technology seems to be adequate for most purposes in achieving small LED element sizes. The Litton display built for the Air Force's Multimode Matrix Display program has 2.5 elements/mm, while fairly standard GaAsP arrays use elements of 0.4 mm square on 0.5 mm centers, to obtain 2 elements/mm. The greatest density array known, an experimental linear array being investigated by Texas Instruments, has 0.10 mm center-to-center spacing, or 10 elements/mm (Scanlan and Carel, 1976).

- 104 -

The shape of the individual elements is rarely, if ever, reported in the literature. Scanlan and Carel (1976) published a slit photometer tracing of the luminance distribution of the TI array, reproduced in Figure 30. Although no distance dimensions or luminance values are given, it appears that the element width is approximately equal to the edge-to-edge element spacing, for an active area of about 25%.
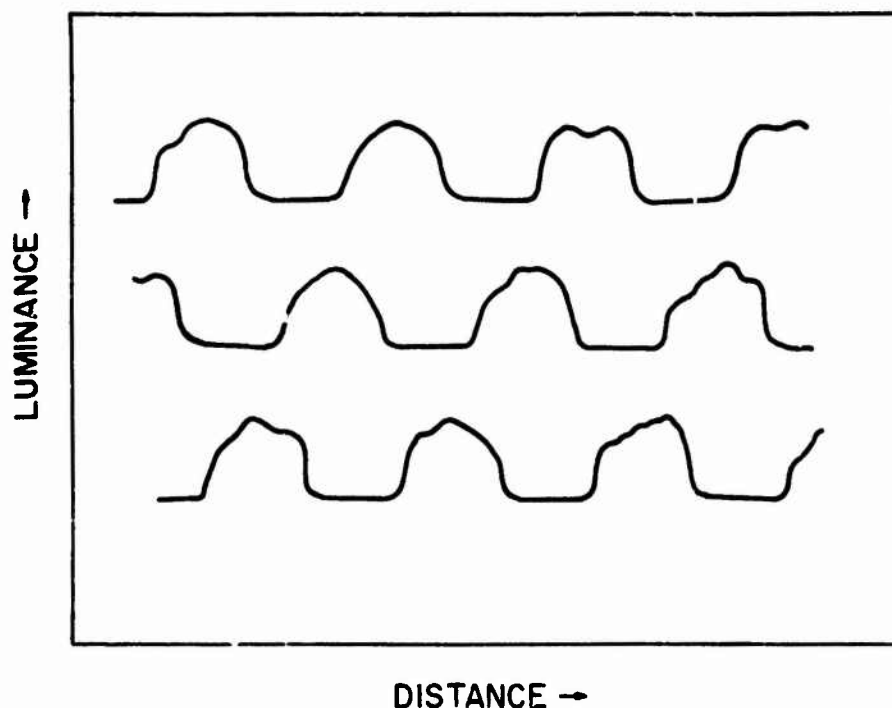
Figure 30: Luminance distribution of TI linear LED array, from Scanlan and Carel (1976)

The Litton GaP prototype array, having 2.5 elements/mm, has been photometrically measured (Kaelin, Nakahara, Piatt, Price, and Riggs, 1975). Its effective element width is invariant with luminance (varied by driving current), and is

- 105 -

approximately 0.25 mm at half amplitude, or 0.36 mm full width. Using the latter number as element width, the edge-to-edge spacing is calculated to be approximately 0.04 mm, for an active area measure of 81%.

It is reasonable to assume that the 10 element/mm density of the TI array is adequate for most purposes. Even at a relatively close viewing distance (e.g., 0.25 m), each element subtends only 0.69 arcminute. With the roughly equal elements and spaces between elements, the spatial frequency of the elements is about 87 cycles/degree, which is beyond reasonable visual requirements.

3.2.3.7   Contrast and dynamic range

While maximum luminance data, typically as a function of current, are often presented for LED arrays, the derived measure of contrast is often not accurately obtainable. That is, the photometric data rarely include the luminance of the (inactive) area between the "on" elements, and only sometimes include the luminance of the "off" areas of the display, i.e., areas where all elements are turned off.

Litton (Scanlan and Carel, 1976) claims a contrast ratio (or dynamic range) in excess of 25:1 for their GaP array, at a peak element luminance of 685 cd/m^2. Data from Kaelin et al. (1975) show a peak luminance, at 40 mA, of 2×10^4 cd/m^2 and a minimum, at the edge of the element, of less than 3.4×10^2, for a contrast ratio or dynamic range in

excess of 50:1. However, their data are for a single "on" element, and may or may not be representative of obtainable contast at all spatial frequencies. This again highlights the lack of measurement standards in the flat-panel technology area.

3.2.3.8 Uniformity

Discrete LEDs are known to vary in luminance, and must be individually tested and grouped when such nonuniformity appears critical. No data on sample variance by LED type are known to exist, however.

LED arrays probably present the most critical uniformity problem. For example, the Litton Systems display, with the Optotek Limited LED array, had only 10 inoperative LEDs out of 49,152 (0.02%). Thus, small area nonuniformity due to element outage is not likely to be a major problem in LEDs. Similarly, as previously shown in Figure 30, diode-to-diode uniformity in peak luminance appears quite good, although no quantitative data have been published to date.

The major problem in LED arrays is that of obtaining chip-to-chip uniformity. When arrays of 8 x 8 LEDs/chip (Scanlan and Carel, 1976) or monolithic arrays of 30 x 36 LEDs (Frescura and Luechinger, 1975) are abutted, edge discontinuities become frequent and prominent. While no measurement of such typical variation has been published for various arrays, Burnette and Melnick (1980) have measured

luminance nonuniformities of 102 cd/m^2 $\pm$ 28% for the Litton arrays. This degree of large-area nonuniformity is quite visible and considered unacceptable for production quality.

3.2.3.9   Temporal characteristics

As previously noted, although LEDs can be DC driven, high frequency refresh produces much greater luminance and luminous efficiency. A key question then revolves around the required refresh rate. Piley (1977) has described this problem, and has shown experimentally that a refresh rate in the range of 400 to 1000 Hz is needed to avoid flicker or image "breakup." A rate of 500 Hz is typically acceptable, but a higher rate is required when there is relative motion vibration between the display and the observer. This need for a high refresh rate is largely due to the very sharp rise and fall times of LEDs, typically about 10 ns (Sherr, 1979).

3.2.3.10   Addressing/driving interfaces

The short rise and fall times of the LED cause it to be characterized as having essentially no "memory." As a result, it must be scanned repetitively to (1) maintain high luminance, and (2) avoid flicker. Refreshing a 512 x 512 element display at 400 Hz in an element-at-a-time scanning pattern would require an address/refresh rate of slightly over 1 x 10^8, or 100 MHz, which is impractical if not impos-

sible. Line-at-a-time addressing and refresh, of this same display, would reduce the bandwidth by approximately 512, or to 2×10^5 MHz, which is achievable but very expensive.

However, the luminance of any element is inversely proportional to its refresh duty cycle. Thus, an element in the array which would have a peak luminance, under ideal conditions, of 17,000 cd/m^2, would have a maximum luminance on the order of 17-34 cd/m^2 if refreshed at a 1/500 duty cycle by line-at-a-time addressing/refresh (Van Raalte, 1976). Compared to other technologies, large-array matrix-addressed LED displays appear uneconomical and impractical.

3.2.3.11   Cost

At the present time, small indicator LEDs are available for a few cents each, in all four main colors.

In large arrays, the cost is estimated at 2.5 cents/element, or about $6500 for a 512 x 512 element array, without addressing circuitry (Scanlan and Carel, 1976). Compared to other alternatives, this cost appears unacceptably high. Of course, large volume production costs might be much less, but it is impossible to estimate them at this time.

3.2.3.12   Utility for display-type applications

Without question, LEDs are excellent candidates for single and multiple alphanumeric readouts. They are very inexpensive, draw little power, come in several colors, have a

- 109 -

very high luminance (if desired), and can be made into vir-
tually any size or shape.

When single LEDs are put into a matrix display for a
graphics or television-type image, however, the cost and the
power requirements lessen the attractiveness of LEDs. For
example, while small alphanumerics that draw 50 to 100 mW
per element can be tolerated, a matrix display of 10,000
elements will require 100 W of power, even if only 10% of
the elements are on at a time (Sherr, 1979, p. 297). While
very small monolithic arrays have been fabricated which draw
little power, they also have unacceptably low luminance lev-
els for most applications (less than 1 cd/m^2). In sum,
large LED arrays are impractical unless there is a large
source of power available, and a means by which to remove
the heat generated.

For the same reasons, large screen LED arrays appear to
be impractical at the present time.

3.2.3.13   Future technology projections

Discrete diode indicators and small alphanumeric fixed-
format readouts are readily available at low cost, exist in
several colors, have high reliability, and are integrated
circuit compatible.

Larger area displays, suitable for graphics and video
displays, are available or are in development with up to 512
x 512 diodes/panel and densities ranging from 2.5/mm (Mon-

santo, Litton) to 5/mm (Bowman) to 10/mm (TI), the last being a linear array. The display technology is available for high-density LED arrays, but the development (and perhaps production) costs are very high. Coupled with the significant heat dissipation problems of many-element arrays, it appears doubtful that large array LED panels will replace other technologies for graphics and video display requirements. It is likely, however, that future research and development will be devoted to increasing luminous efficiency, obtaining flexible color capability, improving fabrication techniques, and reducing production costs. Even with significant progress along some or all these lines, LED arrays seem far behind other flat-panel technologies as the choice for many display applications.

3.2.4   Electroluminescence (EL)

Electroluminescence, or more properly field-excited electroluminescence, has been a familiar phenomenon for many years. The earlier EL displays typically consisted of imbedded phosphor particles in a dielectric medium located between two parallel conductors or electrodes which were about 50 μ apart. Zinc sulfide (ZnS) is a phosphor material often used in this form. Other phosphors have included cadmium sulfide (CdS), zinc selenide (ZnSe), and cadmium selenide (CdSe), with activators such as copper (Cu), silver (Ag), manganese (Mn), and the lanthanides (Sherr, 1979).

- 111 -

Details of the electron physics are described in a variety of publications (e.g., Ivey, 1963; Piper and Williams, 1957).

In recent years, research in both the United States and Japan has led to the very promising film-layer EL. In the film EL, the phosphor powder is replaced by a film which has been deposited on a substrate, typically glass. The film is polycrystalline and is composed of the same material (ZnS) with Cu or Mn activators as in the phosphor powder.

Both phosphor powder and film layer EL displays can be driven by AC or DC current, thus providing four possible generic types of EL displays. Since these four types have different properties and uses, they will be discussed separately, as necessary, in some of the following sections.

3.2.4.1   Physical size and configuration

The AC powder (or AC layer) type of EL device consists of a powder phosphor which has been mixed with a binding material, typically an organic liquid, and deposited on a transparent sheet, usually glass. The glass substrate will typically have been previously coated with a conductive layer, such as SnO_2 or In_2O_3. A metallic electrode, such as aluminum, is then placed over the phosphor powder. Because this is a capacitive device, an AC voltage applied between the transparent (e.g., SnO_2) electrode and the metallic (e.g, aluminum) electrode will cause luminescence of the phosphor

- 112 -

powder. The size of the elements so turned "on" will vary directly with the electrode sizes, which can be as large or as small as necessary.

The DC powder (or DC layer) has become popular only in the last 10 to 12 years. The DC powder is typically ZnS:Mn, usually with copper as a coactivator deposited only on the surface of the ZnS phosphor granules (Schlam, 1973). It can be fabricated to the same variety of sizes as the AC powder.

The DC film EL is physically similar to the AC powder in that it is a polycrystalline film that has been deposited on a conductively-coated glass substrate. The phosphor is usually ZnS:Mn,Cu and has the greenish-yellow manganese emission. Because these films all require copper to achieve the EL process, their operating life is generally short (Schlam, 1973). DC film EL panels have not been fabricated commercially in many sizes or shapes because of this low operating life, although research continues on this type of device.

The ZnS:Mn,Cu film can also be driven by an AC voltage, in which case there is little difference between the AC and DC film characteristics. Figure 31 illustrates an AC thin film EL panel (Inoguchi, Takeda, Kakihara, Nakata, and Yoshida, 1974). A different approach originally developed by Sigmatron (Schlam, 1973), however, incorporates a contrast enhancement technique that apparently improves the AC film, which Sigmatron terms the light-emitting film (LEF). This configuration uses a black layer which absorbs ambient

light upon the back metallic electrode. Viewing the metallic electrode through the transparent glass substrate, dielectric film electrode, and transparent EL film, one sees only the black light absorbing layer, which has a reflectivity of only 0.25%. The AC film, as with other EL types, can be fabricated in a variety of sizes. This black absorbing layer approach was further developed by Aerojet ElectroSystems into 10 x 10 cm AC film panels and tentatively into 1.63 x 1.63 m panels (Fugate, 1977).

METAL ELECTRODE (Al)

INSULATING LAYER (~2000Å)

ACTIVE EL LAYER (ZnS: Mn, ~5000Å)

INSULATING LAYER (~2000Å)

TRANSPARENT ELECTRODE (SnO$_2$)

GLASS SUBSTRATE
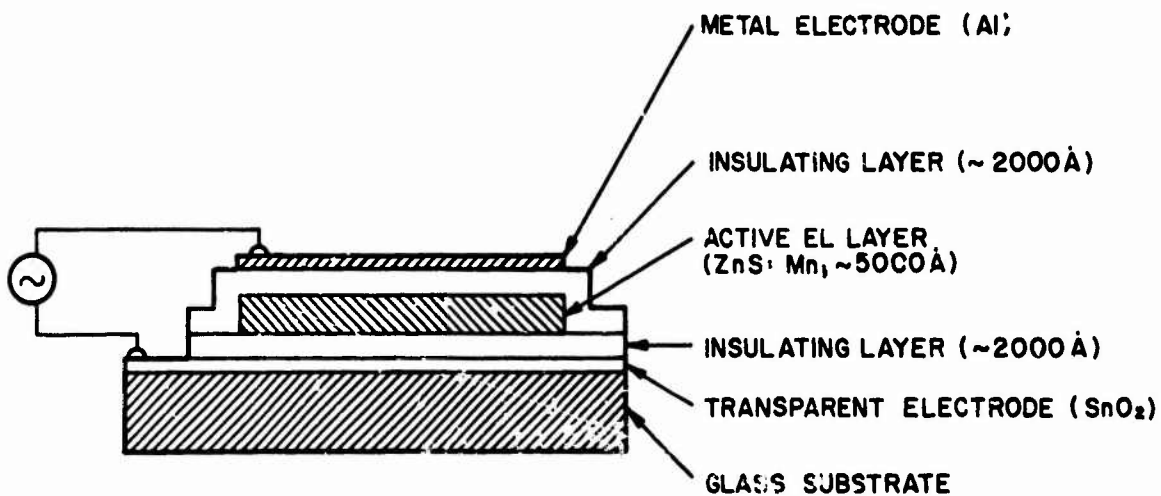
Figure 31: Cross-section of an AC thin film EL panel, from Inoguchi et al. (1974)

3.2.4.2 Power and voltage requirements

EL displays are generally driven by voltages in the 30-650 V region, with display luminance directly affected by the driving power. Thus, it is not meaningful to talk about a specific driving voltage for EL displays. However, the four EL display types can be compared in terms of luminous efficiency, which is done subsequently.

In all cases, operating life decreases as driving voltage increases. Thus, designs which require lower driving voltages are more desirable.

3.2.4.3 Spectral emission

As with other flat-panel technologies, the emission spectrum of EL varies with the specific chemical material. Copper-activated zinc sulfide (ZnS:Cu) is probably the most commonly used EL powder. Depending on the activator concentration, the dominant hue of this compound can range from green to blue. Registered phosphors of ZnS:Cu are the P2 and P31, although the following, which also contain silver, cadmium, or selenium, are quite closely related: P4, P6, P7, P11, P14, P17, P20, P22, P28, P32, P35, P36, P37, and P40 (Schlam, 1973). As the concentrations of silver and copper vary, as activators, with ZnS, it is possible to obtain any emission hue from red to blue. The ZnS:Cu green emission is about 100 nm wide, centered at approximately 525 nm. The emission spectrum of ZnS:Mn peaks at about 585 nm,

and is somewhat narrower than that of ZnS:Cu, as indicated in Figure 32 (Schlam, 1973), which compares both with the CIE photopic luminosity function of the eye. Note that a greater portion of the ZnS:Mn emission lies within the luminosity function than is the case with ZnS:Cu. The ZnS:Mn appearance is distinctly a greenish yellow, perhaps tending toward a brownish yellow.
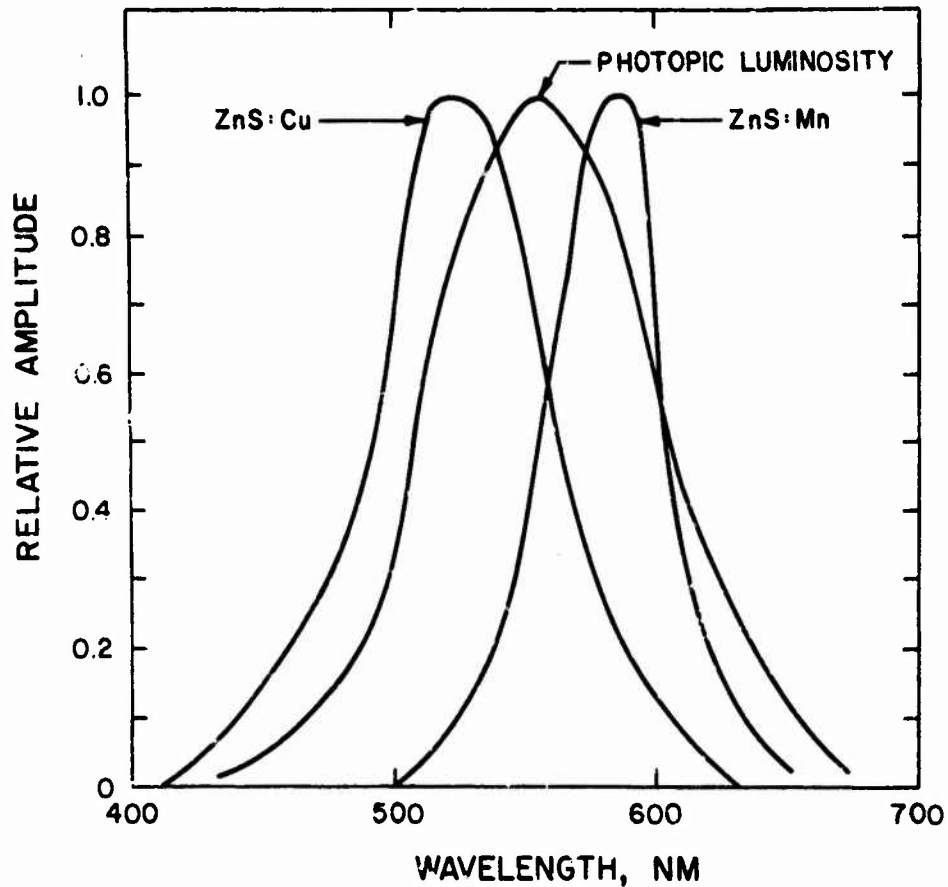
Figure 32:   Emission spectra of ZnS:Cu and ZnS:Mn El phosphors, from Schlam (1973)

3.2.4.4 Luminance

Figure 33 indicates the steady-state luminance of a typical ZnS powder when driven at various audio frequencies. As Kazan (1976) points out, the increase in luminance is superlinear with voltage, increasing approximately as the third power. The figure also shows the increase in luminance with the AC frequency, from 400 to 2000 Hz. Some materials reach maximum luminance at 10,000 Hz.
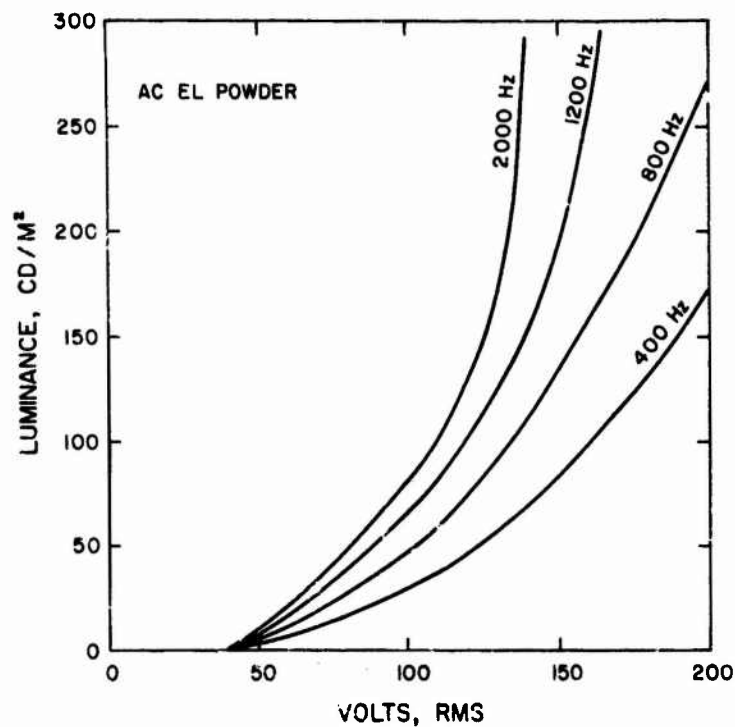
Figure 33: Effect of driving voltage and frequency on EL luminance

EL powder phosphors, when excited by a DC voltage, can also produce relatively high luminance, for example 340 cd/m^2 at 100 VDC (Vecht, Werring, Ellis, and Smith, 1973). Of particular practical interest is the fact that high lumi-

- 117 -

nance (e.g., 170 cd/m^2) can be obtained at low duty cycles, such as 0.5% (at 120 VDC, 3 µs pulse). With this type of phosphor and DC excitation, a very steep rise in luminance with applied voltage is obtained. Whereas the luminance increases with the third power of voltage for AC driven pow-ders, the increase with DC powders is proportional to the sixth (or greater) power of the voltage (Kawarada and Ohshima, 1973).

Luminance output with the thin-film EL is greater than that obtained with powders (Kazan, 1976). For voltages above 140 Vrms, the ZnS:Mn thin film produces an order of magnitude greater luminance, reaching 1.3×10^3 cd/m^2 at 250 Vrms, at 5 kHz frequency. Of importance also is the steep slope of the luminance/voltage relationship prior to satura-tion, for relatively small voltage differences (110 to 160 Vrms) can be used to control very large luminance differ-ences of 0.1 to 10^3 cd/m^2. Using this technology, it is possible to achieve about 10,000 cd/m^2 luminance at 20 kHz and high voltage, although the lifetime of the device is substantially shortened at this voltage/frequency combina-tion.

With line-at-a-time addressing, Matsushita Electric (Mito, Suzuki, Kanatani, and Ise, 1974) has demonstrated off-the-air TV imaging at a highlight luminance of 200 cd/m^2, with a contrast ratio of 50:1. As has been noted, longer device life is obtained by decreasing the drive volt-

age, which is feasible when high contrast displays are possible.

3.2.4.5 Luminous efficiency

According to Kazan (1976), the luminous efficiency of AC powder phosphors is about 5 to 10 lumens/watt over a considerable range of frequencies for the ZnS greenish-blue emission. Brody (1979) has reported 19 lumens/watt at non-peak luminances, and Schlam (1973) indicates that 18 or 19 lm/W is obtainable with the addition of Se to the phosphor powder.

Although DC phosphors achieve high luminances, their luminous efficiency is lower than that of AC phosphors, typically being on the order to 0.5 lumen/watt (Vecht et al., 1973), although 1.5 lm/W has been observed (Schlam, 1973).

To date, AC driven thin-film EL displays have had low luminous efficiency, on the order of 0.3 lumen/watt (Kazan, 1976). However, Fugate (1977) claims 1-2 lm/W is obtainable and 10 lm/W theoretically possible. No data are known to exist on the luminous efficiency of DC thin-film EL materials, perhaps because of their short operating life.

3.2.4.6 Element size, shape, density

DC driven EL powder displays can be fabricated in a virtually endless combination of sizes and shapes. The element size can be made as small as desirable, limited only by the

electrode photofabrication processes. Element sizes as small as 75 μ x 75 μ are quite achievable (Vecht et al., 1973). Similarly, single areas as great as 800 cm² are also commercially available. (Large panels, however, are subject to heat sinking requirements which. if not adequately handled, can cause burn-in nonuniformities.)

In matrix addressed AC thin film displays, densities of 4.9/mm can easily be achieved, and 19.7/mm should be attainable (Fugate, 1977).

3.2.4.7   Contrast and dynamic range

Appropriate selection of driving voltage and EL materials can achieve up to 10^5 cd/m² luminance. When coupled with an absorbing background, or a good ambient absorbing filter, a discrete element alphanumeric EL display can provide as much contrast as needed for most applications. The Sharp 480 character alphanumeric display panel, which is addressed line-at-a-time and consists of a ZnS:Mn thin film layer, has a 25:1 contrast ratio, with a maximum luminance in excess of 50 cd/m². Thus, the "off" cells emit approximately 2 cd/m² due to the "half-brightness" effect or other scattering. (The difference between this 25:1 contrast ratio and the 100:1 of the TFT EL points up the importance of the elemental electronic control.)

Dynamic range and contrast become limited, however, for large matrix-addressed EL displays due to (1) refresh duty

cycle, and (2) the "half-brightness" effect. Each of these is important in understanding EL display design tradeoffs.

Because the fall time of the EL element is on the order of 1.5 ms, refresh is necessary, as in the case of the LED. Pulse length, voltage, and duty cycle tradeoffs will be discussed below. However, it is immediately clear that a large matrix, such as 512 x 512, will lead to limited luminance with an element-at-a-time addressing (< 0.0004%) duty cycle or a line-at-a-time addressing (< 0.19%) duty cycle. At such low duty cycles, high voltages are needed to produce reasonable luminance. Thus, a built-in element electronic circuit is desirable and necessary (1) to obtain individual element control without crosstalk or the half-brightness effect, and (2) to provide element-by-element memory, effectively increasing the duty cycle.

To eliminate these problems, considerable research at the Westinghouse Research and Development Center and at Aerojet ElectroSystems Company has been dedicated to the development and feasibility demonstration of a thin film transistor active circuit design which is integral with the EL element matrix, each EL element being directly associated with an addressing circuit.

Fugate (1977) has pointed out that suspended phosphor layers (e.g., powder EL) are diffuse scatterers of light (i.e., nontransparent), whereas the thin film EL is transparent. Thus, the marriage of thin film EL with TFT driv-

ing/memory electronics is a natural one. With this approach, a 50 cd/m^2 EL character is quite legible under 20,000 lm/m^2 illuminance, and a 70 cd/m^2 display is legible in a brightly illuminated aircraft cockpit (Fugate, 1977).

As an example, the Snarp TFEL TV (Figure 34) has 240 vertical by 320 horizontal picture elements, is noninterlaced, and is refreshed at 60 Hz. It presently has 2.7 elements/mm, and 3 elements/mm is considered possible. The maximum luminance is about 90 cd/m^2, but no contrast measurements are known.
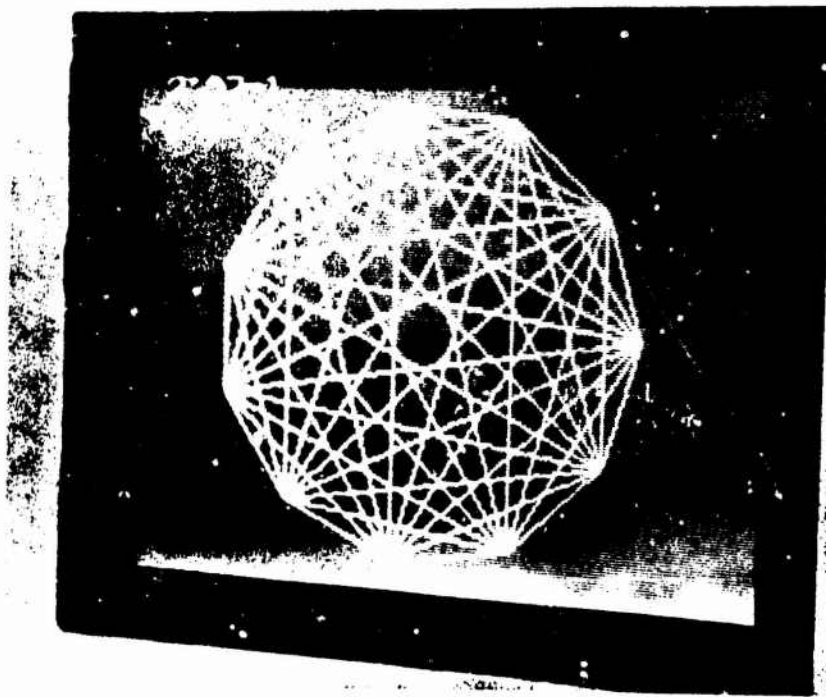
Figure 34:   Photograph of Sharp TFEL TV display, courtesy of Sharp Corporation

Both Matsushita and Sharp in Japan have produced EL matrix addressed panels, which are also compatible with com-

mercial TV data rates, as is the Westinghouse TFT EL display. Matsushita's TFEL TV panel has 230 x 230 picture elements, is noninterlaced, presents 60 fields /second, and has a maximum luminance of about 35 cd/m^2.

The Sharp 12-row alphanumeric display has a maximum luminance of 140 cd/m^2, is made of TFEL, and is addressed line-at-at-time (Figure 35).
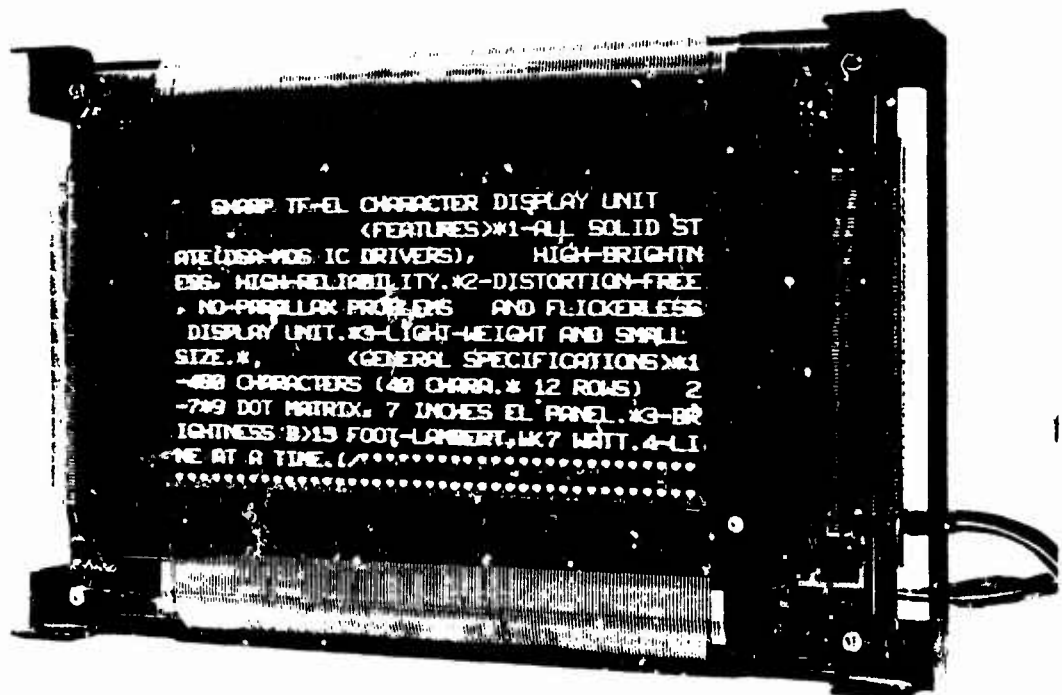
Figure 35: Photograph of Sharp TFEL alphanumeric display, courtesy of Sharp Corporation

3.2.4.8   Uniformity

No quantitative data have been published on the uniformity of EL displays, powder or film.   Assuming fairly even

phosphor deposition, one would expect good uniformity when the display is operated at high luminance, for under the high luminance condition the EL element is in saturation.

However, many applications (e.g., cockpit displays at night) require uniform dimming, which is achieved by either analog (voltage) or digital (pulse width or duty cycle) modulation. If the element-by-element uniformity is low, then analog modulation is unacceptable because of the steep luminance/voltage slope (Figure 33). Digital modulation at high (e.g., TV) data rates can be very difficult to obtain.

With EL powder displays, the uniformity will be limited by the phosphor density and the activator uniformity. For EL film, uniformity will be limited by the EL layer thickness, which is difficult to control over a large area (Goede, 1978). Quantitative values for both large and small area nonuniformity due to these fabrication variables apparently do not exist, although researchers and viewers readily see such nonuniformity.

Perhaps more critical than large area uniformity is the element-out or line-out nonuniformity which can be caused by defective driving/addressing circuitry. Although prototype panels frequently exhibit such errors, no severity or frequency data are known to exist. Edge nonuniformities caused by driving electronic errors are considered to be particularly likely with the TFT active circuit approach (Goede, 1978).

- 124 -

3.2.4.9   Temporal characteristics

The rise and fall times of EL phosphors are short enough
for many dynamic display applications. For a DC driven ZnS
powder at 50 V (Vecht et al., 1973), the rise time is
approximately 1 ms and the fall time is about 1.5 ms, both
measured to 90% of peak.

The average luminance of a DC powder EL panel will depend
upon its pulse duration and duty cycle, as well as upon the
driving voltage. Figure 36 illustrates the manner by which
luminance increases monotonically with driving voltage and
pulse length for voltages up to about 70 VDC, and varies
nonmonotonically with pulse length beyond 70 VDC (Vecht et
al., 1973). As voltage is increased, the optimal pulse
length decreases.

Because the available duty cycle will depend upon the
number of lines or elements to be addressed in an X-Y
addressed display, the voltage and pulse length must vary
with the information requirements and size (number of ele-
ments) on the display. The maximum duty cycle of an ele-
ment-at-a-time addressed display is the reciprocal of the
number of elements; for a line-at-a-time addressed display,
the maximum duty cycle is the reciprocal of the number of
lines.   Clearly, dense displays without element memory
require high voltage driving to compensate for short duty
cycles and short pulse widths.
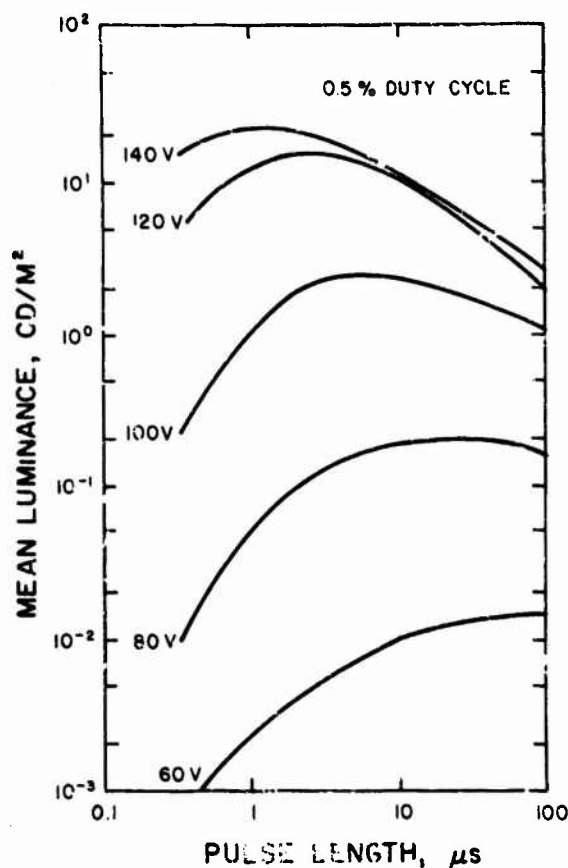
Figure 36: Effect of pulse length and duty cycle on EL luminance, from Vecht et al. (1973)

3.2.4.10 Addressing/driving interfaces

Directly addressed EL alphanumeric segment readouts present no particular interface problems. Although they require voltages in the 50-600 V region, low current power supplies are adequate and driving logic can be interfaced with IC electronics.

Matrix panels have long been considered suitable for X-Y or line-at-a-time addressing. However, crosstalk (e.g., half-brightness) and low duty cycle limitations are substantial, typically leading to high driving voltages. The advent of TFT drivers for each element, especially when cou-

- 126 -

pled with a black layer, EL film display, appears to solve
many of the addressing/memory/driving voltage problems. The
resulting display is highly compatible with integrated cir-
cuitry and available power supplies.

3.2.4.11   Cost

The cost of EL displays is not dependent on the cost of
the phosphor material, for ZnS:Cu can be obtained for about
$20/lb. (Schlam, 1973). Rather, the cost is more likely to
be driven by the display fabrication process complexity,
yield rates, and, in the case of TFT-driven displays, the
associated addressing/memory circuitry.

High voltage line drivers for EL panels cost approxi-
mately $.25/driver, with a driver required for each line and
some fraction of the number of elements/line (Tannas, 1978).
For a 512 y 512 display, line drivers alone would cost on
the order of $150. TFTs cost about $8/cm^2. "Pinout" prob-
lems, typical of all flat panels, cost on the order of $.01
per connection times the number of rows plus columns (Tan-
nas, 1978). Line drivers must ultimately be individually
connected to one or two large circuit boards.

Taken together, the electronics cost of a 15 x 15 cm, 512
x 512 element TFT driven EL panel is estimated to be on the
order of $3500-$5000, in production quantities. To this
must be added the production cost of the display panel
itself, which simply cannot be estimated at this stage of
development of the technology.

3.2.4.12 Utility for display type applications

The EL display, especially the black layer, thin film EL is potentially compatible with requirements for alphanumeric readout, graphics, and TV displays, although TFT addressing electronics are probably needed for environments having any significant ambient illuminance in which a graphics or TV display is required. This assumes, cf course, that small and large area nonuniformity is not a problem, and that element and line addressing errors are virtually nonexistent

3.2.4.13 Future technology projections

EL panels have several distinct advantages. They can produce reasonable gray scale and dynamic range; have a wide acceptance angle for viewing; can be fabricated in sizes ranging from a few centimeters to greater than a meter; and are potentially capable of very high element density (Kazan, 1976).

They also have some distinct disadvantages, including the cost of a large number of high voltage drivers, low luminous efficiency, and color limitations.

Applications of thin-film EL technology, especially when coupled to TFT electronics, are several and are growing in number. Flexible alphanumeric readout and graphics capability are already here. Monochrome TV is acceptable, although the subjective yellow color, decidedly nonachromatic, is disliked by many. Color TV, at luminance levels of 150

cd/m^2 or above, does not appear imminent. Until more pure red and blue emitting EL phosphors are well developed, color EL TV appears to be far away in the development cycle. Thus, immediate emphasis on EL research and development will likely be in the areas of color capability, achieving greater luminous efficiency, and reducing driving/addressing costs. Cost reduction, in particular, may determine the ultimate utility of matrix addressed EL displays.

3.2.5 Plasma Displays

In a 1975 review article, Weston wrote that

> the gas discharge plasma panel offers the only available alternative to the highly developed cathode ray tube. Furthermore, the technical success of such panels has encouraged research on . . . half-tones and colour with the ultimate goal of the large screen picture-on-the-wall television, which is the television engineers' dream for the future.

While that dream still remains a dream, the gas discharge, or plasma, display has been produced in large volume, in many configurations, by several manufacturers, and for different applications. Of all the flat panel technologies, more attention and more success have accompanied the plasma display panel than any other.

There are two types of plasma display, one AC driven and one DC driven. Both have had good success in the market, and have been fabricated in alphanumeric readouts as well as in matrix-addressed panels for graphics and alphanumerics.

Plasma displays necessarily have one transparent (front) electrode, through which the display is viewed. The rear electrode can be black, reflective, or clear. In the last case, it is possible to rear project an image on the display, thereby using the plasma display as overlay information on the projected image.

3.2.5.1 Physical size and configuration

Gas discharge or plasma displays have existed for a long time, typified by the well-known Nixie^{TM} numeric indicator and the common commercial neon sign. The basic mechanism is a gas-filled volume, across which an electrical field can be controlled. The electrical potential can cause the movement of an electron from one energy to a lower energy level, simultaneously separating the electrons from the atoms. When a sufficiently large number of atoms have lost at least one electron, the gas is said to be in its ionized state. Light emission accompanies this process. When the concentration of electrons and ions is less than 1%, the gas is called a "plasma," hence the term "plasma display."

The general display physical configuration is illustrated in Figure 37. On the left, the electrodes are located inside the glass plates, in direct contact with the gas-filled center cavities. This format is typical of DC-driven plasma displays. The AC-driven display, however, must have the electrodes separated from the gas, as indicated on the

right of Figure 37. Alternatively, a dielectric layer can
be placed between the electrodes and the gas. This dielec-
tric layer is needed, in the AC panel, to provide the capa-
citance needed to store a "sustaining" voltage, as will be
described below. A familiar AC configuration using dielec-
tric glass layers is the Digivue[TM], illustrated in Figure
38.
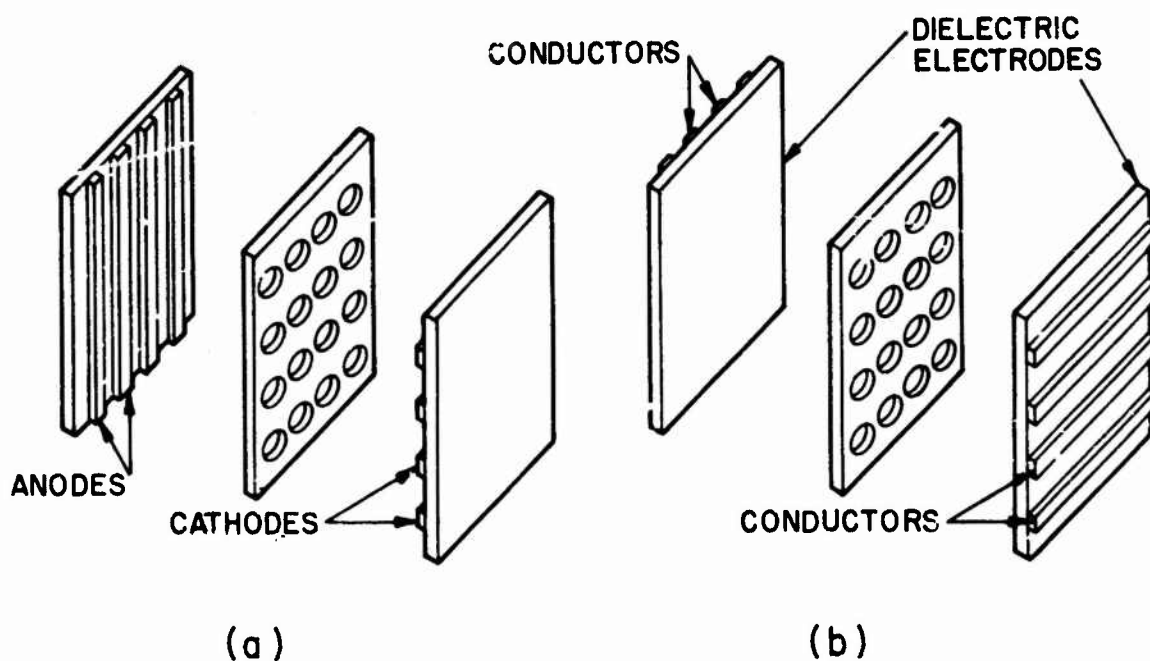
Figure 37: Generalized plasma panel configurations in DC
panels (left) and AC panels (right)

AC and DC driven plasma displays are available commer-
cially in a variety of configurations. Small alphanumeric
displays are typified by the Burroughs Panaplex[TM] display
and the Beckman planar display. Standard single-digit or
single alphanumeric displays are from 5 to 18 mm in vertical
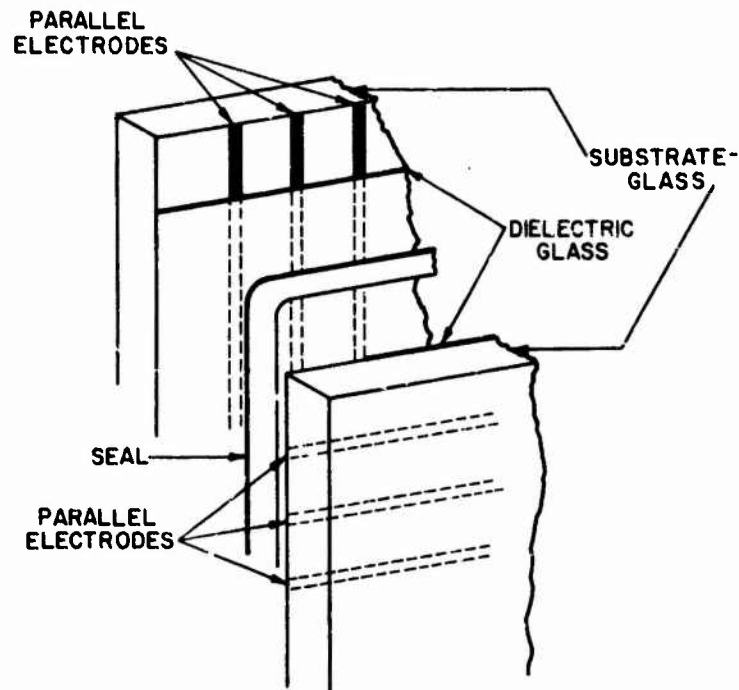
Figure 38:   Owens-Illinois DIGIVUE configuration

height, although up to 200 mm is available on special order
(Sherr, 1979).   Large matrix-addressed panels range from the
single line Self-Scan[TM] to 43 cm square panels having 1024 x
1024 addressable elements.   The largest known panels are 54
cm square.   The plasma display is 12 to 13 mm thick.

3.2.5.2   Power and voltage requirements

Figure 39 illustrates the generalized voltage/current
relationship for a DC plasma display. Briefly, as the volt-
age is increased, the current flow increases in an acceler-
ating fashion until the ignition threshold voltage, V_s, is
reached. This current increase is due to the ionization in
the gas and the secondary emission of electrons from the
cathode.   Up to this point, the current is very low, less

than 1 μA cm^{-2}, and no visible glow is present. When the ignition threshold is reached, however, there is a large current increase accompanied by a voltage drop across the electrodes to a minimum, V_e, at which a glow discharge is observed. Lowering the voltage below V_e, the extinction voltage, extinguishes the glow discharge. Thus, if the voltage is held between V_e and V_s, the display can be turned on by applying a pulse which will temporarily raise the potential above V_s. Similarly, the emission can be turned off by a temporary pulse which drops the potential below V_e. In essence, the display has "memory" or a storage capability as long as the sustaining voltage is between V_e and V_s. The values of V_e and V_s depend upon the geometry, gas pressure, and the cathode material.
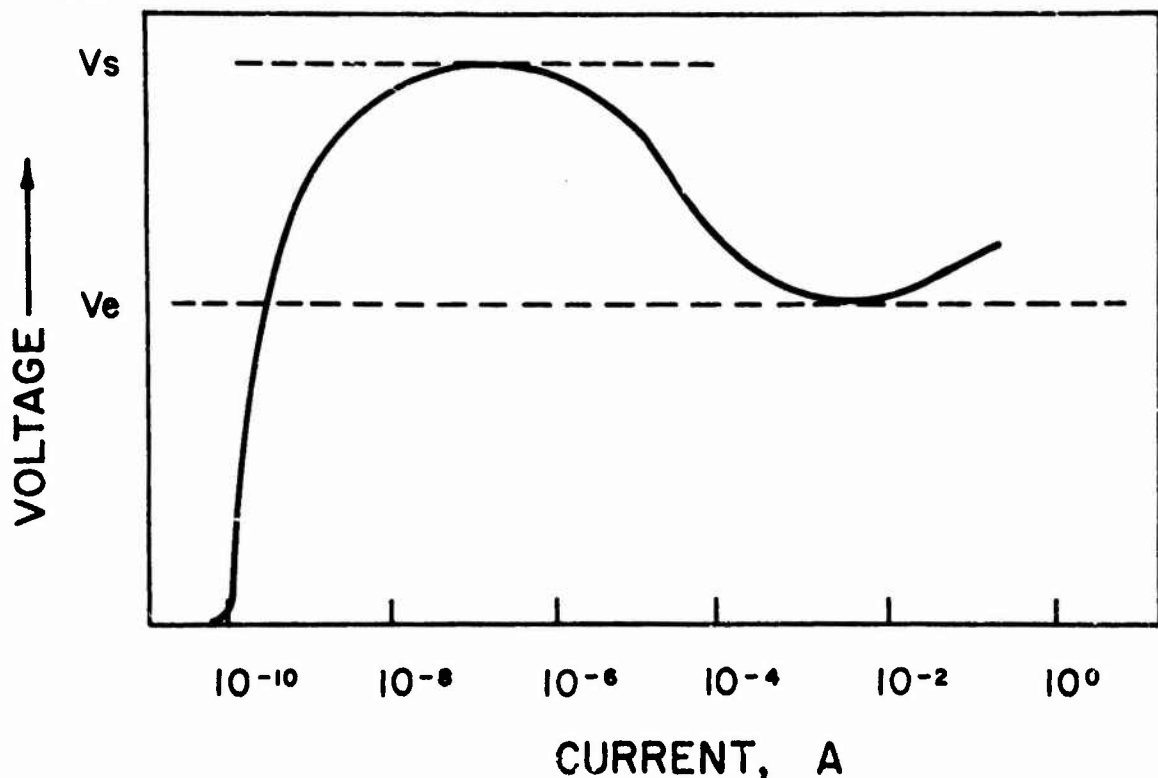
Figure 39: Voltage/current characteristics of a gas discharge display, from Weston (1975)

- 133 -

The AC plasma cell, as noted previously, often has the electrodes placed on the outside of the glass plates. This geometry was originally proposed to protect the electrodes from sputtering, or the erosion of the cathode from ion bombardment during discharge. However, the discharge causes charge particles to flow to the cell walls and accumulate there, eventually building up a charge that quenches the discharge. Therfore, it is necessary to reverse the potential polarity, by using an AC drive. A fundamental advantage of the AC plasma display, as indicated in Figure 40, is that a discharge occurs on each half cycle of the driving voltage. By selecting a voltage below the discharge voltage, but in excess of the extinction voltage, the discharge can be maintained as in the case of the DC cell. As Sherr (1979, p. 219) points out,

> the lower voltage can maintain the discharge because the wall charge voltage adds to the sustaining voltage on each half cycle, and the firing voltage, which is the same for a comparable DC cell, . . . is exceeded by the sum of the sustaining and wall-charge voltages, thus causing the cell to fire and the wall charge to reverse.

This sequence is shown in Figure 40, where V_S is the sustaining or driving voltage, V_t is the firing voltage, V_w is the wall voltage, and V_c is the voltage across the cell.

General current/voltage requirements are as follows. The DC alphanumerics require an anode supply voltage of about 200 V, and each character draws approximately 1 to 2.5 mA, for a power requirement of 0.2 to 0.5 V per character.
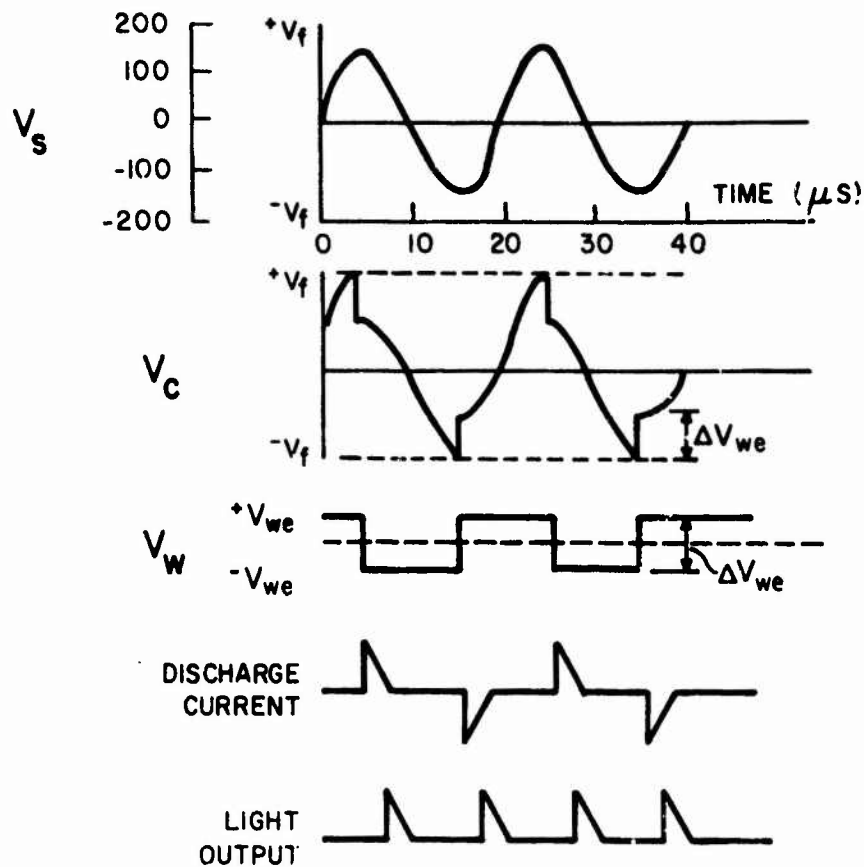
Figure 40: Voltage, current, and light output waveforms for a plasma cell

Larger DC matrix displays of 10,000 to 15,000 elements will draw between 25 and 100 W, averaging between 0.002 and 0.01 W per element.

AC driven plasma panels usually use a sustaining voltage of about 140 V and a firing voltage of 200 V, drawing a current of about 5 μA per element, or 200 W for a 512 x 512 element panel (Scanlan and Carel, 1976).

3.2.5.3 Spectral emission

The most common emission of plasma displays is the orange emission, at 585.2 nm, from a neon gas. Other gases, and their discharge colors, which have been used in gas discharge displays incude argon (blue), cadmium (red), helium (yellow), mercury (purple), and sodium (yellow).

Recently, there has been increasing attention to techniques by which multicolor, e.g., television, displays might be obtained with plasma panels. The most promising technique appears to have the ultraviolet emission of selected neon gas cells excite phosphor dots much as the scanning electron beam in a shadow mask CRT selectively excites red, green, and blue phosphor dots. Ohishi, Kojima, Ikeda, Toyonaga, Murakami, Kioke, and Tajima (1975) described a device (Figure 41) in which red, green, and blue phosphor dots were turned on by the neon emission below 200 nm. The spectroradiometric output of a variety of red, green, and blue phosphors was measured, resulting in CIE chromaticities for the best combinations which were not greatly different from the NTSC TV primaries. In a similar set of measurements, using a DC driven panel (Figure 42), Fukushima, Murayame, Kaji, and Mikoshiba (1975) obtained good saturation in red, green, and blue phosphors excited by the 129.6 and 147.0 nm emission lines of the xenon gas. Their CIE chromaticity results also indicate good correspondence with the NTSC primaries.
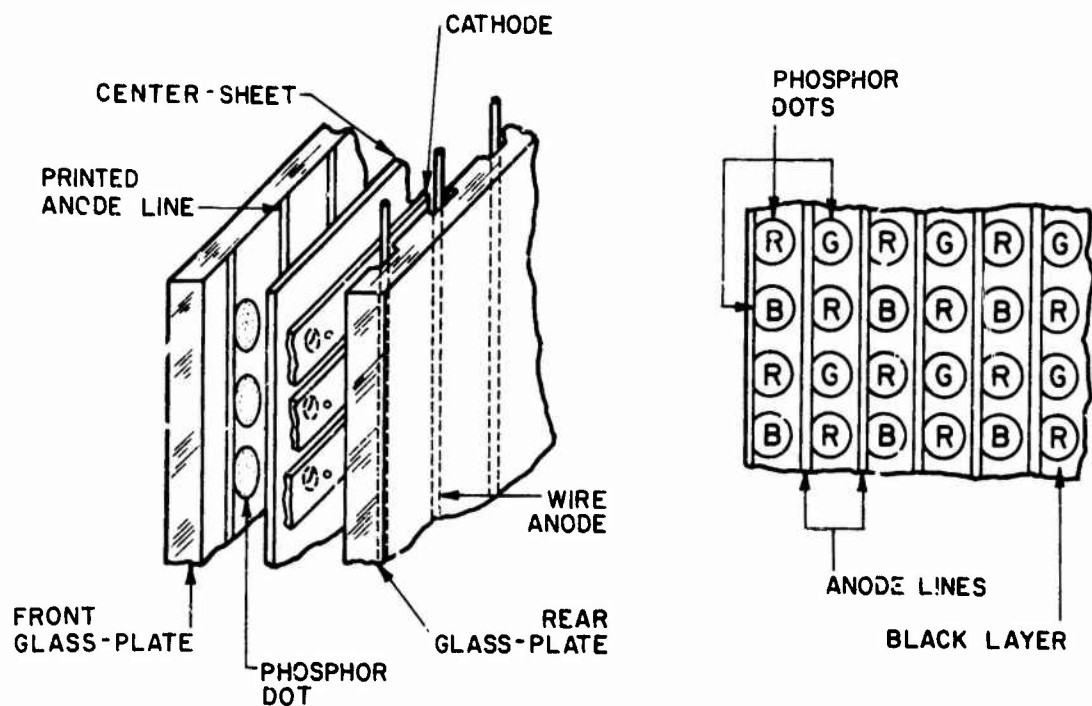
Figure 41: Construction of tricolor plasma display, from Ohishi et al. (1975)

Thus, it appears that several single gas emission colors are achievable, with different luminous efficiencies. Moreover, by using the ultraviolet gas emission to selectively excite red, green, and blue phosphors, one can obtain a fair coverage of the full color spectrum.

3.2.5.4 Luminance

DC-driven alphanumeric displays are capable of about 150 cd/m^2 (Panaplex^{TM}) to over 600 cd/m^2 (Beckman planar gas-discharge), although 30 to 50 cd/m^2 is more typical for long life. DC-driven matrix displays (e.g., Self-Scan^{TM}) nominally provide about 85 cd/m^2, while the AC-driven DIGIVUE^{TM} provides about 150 cd/m^2 average luminance and 250 cd/m^2 at the peak intensity within the element.

- 137 -
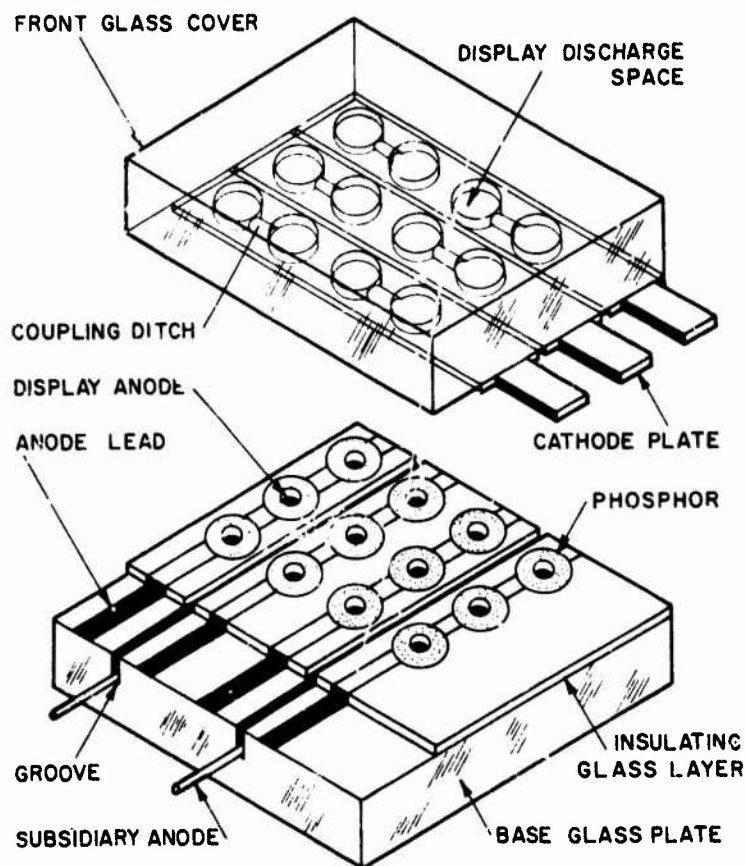
Figure 42: Construction of tricolor plasma panel, from
Fukushima et al. (1975)

Gas-discharge excited color displays developed thus far
are fairly low in luminance, one of the main limitations to
wide acceptance. The experimental panel of Ohishi et al.
(1975) only developed 19.5 cd/m^2 for white (all three phos-
phors driven), and 5.14, 8.57, and 1.37 cd/m^2 for red,
green, and blue, respectively. Even at these low lumi-
nances, panel life is short due to cathode sputtering.

3.2.5.5 Luminous efficiency

The most common plasma panel, the DIGIVUE[TM], has a lumi-
nous efficiency of about 0.3 lm/W (Crisp, Hinson, and Sie-
gel, 1974). Other orange panel displays have luminous effi-
ciencies ranging from 0.05 to 0.8 lm/W, while single cell
luminous efficiencies have been reported at 1.2 and 3.4 lm/W
(Chodil, 1976).

3.2.5.6 Element size, shape, density

As in other technologies, element size and density are
constrained by photodeposition and photolithography techni-
ques used in electrode and cell fabrication. The DC Self-
Scan panel has 1 mm element centerline spacing, with each
element of the rectangular element version approximately
0.75 x 0.43 mm. The Self-Scan is also available with round
elements. The AC DIGIVUE, on the other hand, usually has
2.36 elements/mm, or 0.42 mm center-to-center spacing. The
element size of the DIGIVUE is not easily defined, however,
because its intensity distribution is irregular, as shown in
Figure 43. As an approximation, however, its width between
the intensity peaks is on the order of 0.2 mm. At normal
viewing distances, the individual elements are indistin-
guishable, as their edges blend together.

The nontransparent front electrode in the AC panel cre-
ates a dark line through the element, as shown in Figure 43,
and also reduces the element's luminance. To increase the
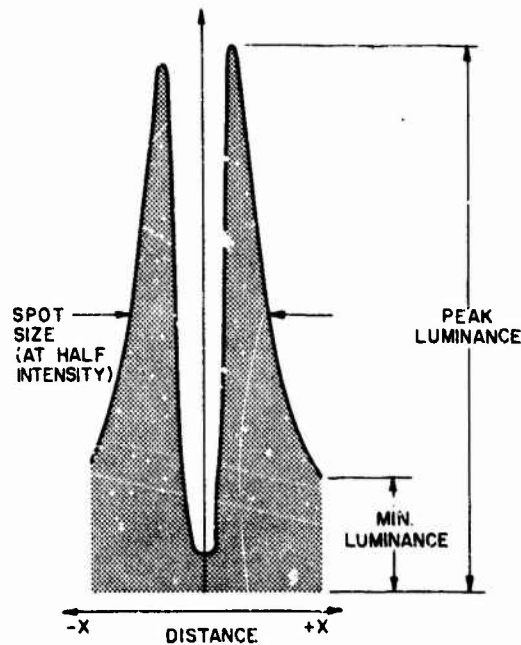
Figure 43: DIGIVUE plasma element luminance distribution

element luminance, some researchers have designed panels with double electrodes. A scan through a double electrode panel gives the luminance profile illustrated in Figure 44. Unpublished microphotometric measurements in our laboratory verify the increased luminance of this configuration.

AC-driven panels are now available with up to 1024 x 1024 elements at densities of 2.36, 2.87, and 3.27 elements/mm. Large displays of 54 cm on a side, with element densities of 3.94/mm have also been made by Photonics Technology. Magnavox is currently fabricating a $(50 \text{ cm})^2$ panel also.
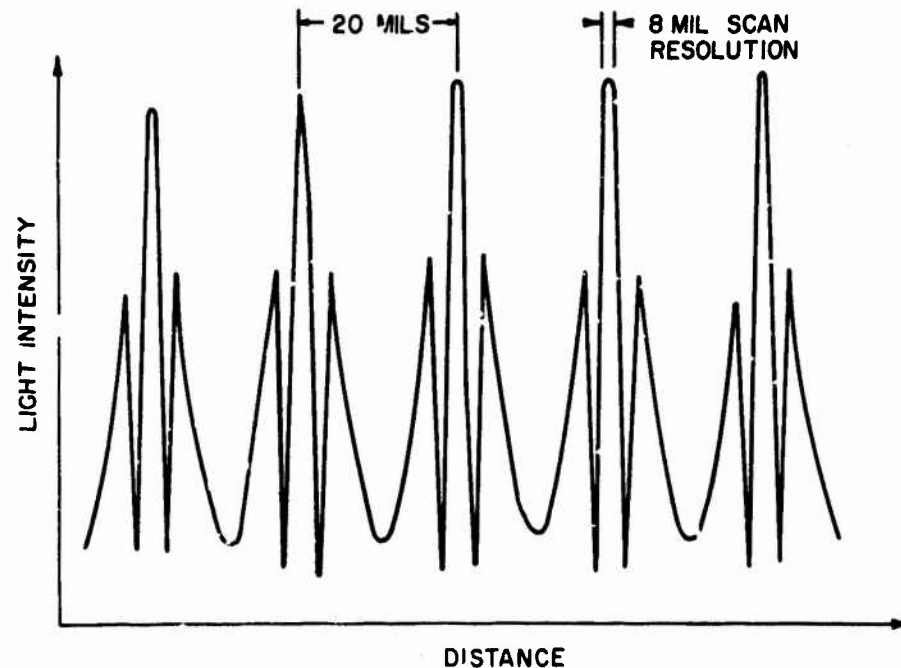
Figure 44:   Double-electrode light intensity distribution in
plasma panel, from Scanlan and Carel (1976)

3.2.5.7   Contrast and dynamic range

DC-driven plasma displays can be controlled in luminance
by either current variation or duty-factor modulation.
Duty-factor modulation requires constant current but varies
the pulse width of each "on" command, noting that the
refresh rate must remain above the flicker frequency, typi-
cally at least 50 Hz.

AC-driven displays remain "on" until turned "off." Since
they are essentially bistable, two techniques have been
developed to control gray levels: spatial modulation and
temporal modulation.

In spatial modulation several cells (e.g., four or eight)
are used to form one resolution element. Combinations of
"on" and "off" cells within this resolution element are

selected which, with the spatial integration of the visual system, form up to 16 gray levels (2^4) for four cells. This technique has been successfully demonstrated by Umeda, Murase, Ishizaki, and Jurabeshi (1973).

Temporal modulation requires selective duty cycles of each element, proportional to the average required luminance. Care must be taken to avoid flickering refresh rates, and to modulate the entire display above the flicker threshold. Mitsubishi has successfully demonstrated this technique on a 128 x 128 cell display at a refresh rate of 30 Hz (Sasaki and Takagi, 1973). Anderson and Fowler (1974) similarly achieved a full-resolution interlaced TV display with 64 temporally modulated intensity levels.

In general, contrast ratios from 40:1 to 8:1 have been reported for DC panels (Chodil, 1976), while a ratio of 20:1 is typical for AC plasma displays (Scanlan and Carel, 1976; Sherr, 1979).

2.2.5.8   Uniformity

All cells in both AC and DC plasma displays share a common gas supply and the same power supply. Thus, small area and large area uniformity are dependent upon physical construction uniformity. For existing panels, the manufacturing technology appears to be sufficiently mature to provide good small and large area uniformity (Goede, 1978), although no empirical data are known to exist. Since most panels are

of unit construction, edge discontinuities due to butting appear to be no problem.

3.2.5.9 Temporal characteristics

The DC alphanumeric plasma display requires at least 125 µs to fully ionize a cell, which sets an upper limit on the scanning rate (Sherr, 1979). Alternately, it sets a limit on the maximum number of characters that can be sequentially addressed at a nonflickering refresh rate. Sherr (1979, p. 279) indicates that this number of digits is given by:

$$N_d = (T_s)/(T_i + T_o) \qquad (14)$$

in which

T_d = scanning time,

T_s = scanning time,

T_i = ionization time, and

T_o = off time.

Using 125 µs for T_i, 16.7 ms for T_s (reciprocal of 60 Hz), and a minimum of 25 µs for T_o, then:

$$N_o = (16.7 \times 10^{-3})/((125 \times 10^{-6}) + (25 \times 10^{-6})) \qquad (15)$$

$$= 111 \text{ characters.}$$

However, in practice it is necessary to maintain each character "on" for 1 ms to avoid flicker from very low duty cycles, so about 16 digits is a realistic maximum (Sherr, 1979).

Large DC panels, such as the Self-Scan^{TM}, avoid this refresh rate limitation by using a "priming" approach. Very

simply, this type of panel uses "priming" holes in the cathodes (or row electrodes) and a priming discharge running in vertical (column) slots on the back of the cathode. Once a discharge is begun, it is easier to start another discharge nearby. This priming channel permits this "easier" discharge, scanning from top to bottom of the panel. Priming thus reduces the number of required drivers by a factor of 70 or so. It also improves gray scale rendition by reducing noise at low luminance levels (Chodil, 1976). Sherr (1979) discusses tradeoffs among duty cycle, number of cathodes, and blanking times for this type of display, concluding that adequate refresh rates are achievable (60 Hz) with a loss in average luminance. However, since the DC panel is capable of 30,000 cd/m^2 maximum luminance, even a 0.74% duty cycle yields an acceptable luminance of over 200 cd/m^2.

Representative temporal characteristics for AC plasma displays include a 100 ms rise time, 2 μs decay time, and a 20 μs "on" time per element, assuming the display is scanned at 50,000 elements per second. Bulk erase times of 11.5 μs/line have been demonstrated (Scanlan and Carel, 1976).

3.2.5.10 Addressing/driving interfaces

As described in several paragraphs above, plasma display present unique and complex opportunities for creative driv-

ing circuitry. Such techniques are well beyond our purposes here. However, it should be noted that such techniques have been successfully demonstrated and implemented by many organizations, with the result that at least 13 different gas-discharge flat panel TV displays have been announced, as indicated in Table 3. Most of these are based on the Self-Scan™ (DC) or DIGIVUE™ (AC) panel technology.

An advantage of plasma technology is compatibility with integrated circuit characteristics. Although the sustaining voltages are in the 100-300 V region, the "on" and "off" pulse commands are IC compatible. Thus, interfacing with computers and other devices is straightforward.

3.2.5.11   Cost

Current prices of 512 x 512 element displays, with all associated driving and addressing electronics, vary from about $4500 for serial input capability to over $9000 for high-speed, parallel processing devices with some internal memory. In part, the cost is high because the number of manufacturers was small. More recently, Owens-Illinois, the developer of the DIGIVUE™, has sold the rights and manu-facturing technology to several companies. Competition and volume sales should help reduce prices considerably. One large manufacturer estimates that the display and its driv-ing circuitry can be sold for less than $200 in very large quantities.

TABLE 3

Comparisons of various plasma panels for TV display, from Chodil (1976)

Company	Panel type, region of discharge utilized	Color	Reported luminance, cd/m²; duty factor	Calculated luminance, cd/m², at 1/500 or 0.2% duty factor	Luminous efficiency, lumens/watt	Contrast ratio	Video storage; Intensity Modulation	Panel Size (cm); Rows × Columns
Philips	DC; negative glow visible emission	orange	205 1.1%	38	0.2 lm/W	10:1	Analog; continuous duty factor	5 × 13 40 × 100
Zenith	DC; negative glow visible emission	orange	27 0.2%	27	0.1 lm/W	40:1	Analog; continuous current	16 × 6 212 × 80
Bell Labs	DC; negative glow visible emission	orange	86 0.33%	51	0.1 lm/W	30:1	Analog; continuous current	21 × 7 222 × 77
NHK Broadcasting Co., Japan	DC; negative glow visible emission	orenga	51 NR	NR	NR	25:1	NR	NR
Bell Labs	DC; negative glow visible emission & phosphor excited by low energy electrons	white	38 1%	7.5	0.09 lm/W	NR	Analog; continuous	17 × 1 110 × 7
Philips	DC; positive column UV excites phosphors	red blue green	NR 2.5%	NR	0.8 lm/W	NR	NR	0.1 diam. 16 × 16
Toshiba	DC; constricted glow visible emission	orange	205 0-035%	2 et 4 × 10⁻⁴%	0.4 lm/W	20:1	Digital; 16 discrete current levels	7 × 10 48 × 65
Mitsubishi	AC; negative glow visible emission	orange	NR	NA	NR*	NR	Digital; 12 discrete duty factor levels	NR 128 × 128
Sony	DC; negative glow visible emission	orange	86 0.25%	72	0.5 lm/W	40:1	Digital; 32 discrete levels useing duty factor & current modulation	11 × 13 212 × 282
Hitachi	DC; negative glow UV phosphor excitation	red blue green	17 white 0.38%	9 white	0.05 lm/W	8:1	Digital; 64 discrete levels useing duty factor & current modulation	12 × 16 120 × 160
NHK	DC; negative glow UV phosphor excitation	red blue green	17 white 0.8%	4 white	0.07 lm/W	30:1	Analog; continuous duty factor & current modulation	13 × 16 127 × 160
GT&E	AC; negative glow visible emission	orenga	86; nearly 100%	NA	NR*	NR	Digital; 64 discrete duty factor levels	22 × 22 512 × 512
GT&E	DC; positive column UV excites phosphors	red blue green	65 green 0.15%	86 green	1.2 lm/W		Data are for single cells, no panels reported	
Zenith	DC; positive column UV excites phosphors	red blue green	1233 green spot luminance 0.18%	1370 green spot luminance	3.4 lm/W		Data are for single cells, no panels reported	
Bell Labs	AC; negative glow visible emission	orange	NR; 100%	NA	NR*	NR	Digital; ordered dither	512 × 512

NA = not applicable

NR = not reported

*AC "Digivue" panels typically have a luminous efficiency of several tenths of a lumen/watt.

- 146 -

3.2.5.12   Utility for display-type applications

Plasma displays have been successfully demonstrated and sold for alphanumeric readouts in single rows, multiple rows, and larger matrix panels. Alphanumeric indicators are available in both dot matrix and seven-segment forms, and lifetimes of 30,000 hours are typical.

Matrix displays ranging from 48 x 65 elements to 512 x 512 elements have been available for graphics and TV applications for several years. More recently, 1024 x 1024 element displays have become commercially available.

Displays larger than 54 cm on a side have not been attempted to date. It seems unlikely that large-screen displays will become feasible with the plasma discharge technology. Cost and fabrication tolerances would probably cause this technology to be noncompetitive for that application.

However, the optical compatability of transparent plasma panels with rear-projected images has been well demonstrated, a feature which should not be overlooked in many applications.

3.2.5.13   Future technology projections

Future R&D efforts in plasma technology will emphasize greater luminance ranges through multiple electrodes (AC), better color capability with greater luminance, and better color resolution. At the present time, the color is too dim

and too coarse, as only 1.2 color dots/mm are possible, compared, for example, with 3/mm for a high-resolution shadow mask CRT. Attention will also be given to cost reduction through innovative addressing/driving schemes.

3.2.6 Liquid Crystal Displays

The liquid crystal display (LCD) technology is the most popular and most developed of the flat-panel passive display types. Rather than emitting light energy, as do active displays, the passive display controls or modifies the passage of externally generated light. In the LCD case, this control or modification is achieved by light scattering, modulation of optical density, or color change (Goodman, 1975).

The liquid crystal material flows like a liquid, but has molecular ordering characteristics of a solid. Thus, the LCD is typically composed of transparent (e.g., glass) plates which sandwich and contain the liquid crystal substance, and which also serve as transparent electrodes (Figure 45). Voltages applied to these electrodes cause realignment of the crystal molecules, thereby changing the optical characteristics and light propagation through the liquid crystal substance.

There are several generic types of liquid crystal molecular organization, each of which has different electrical requirements and optical properties.
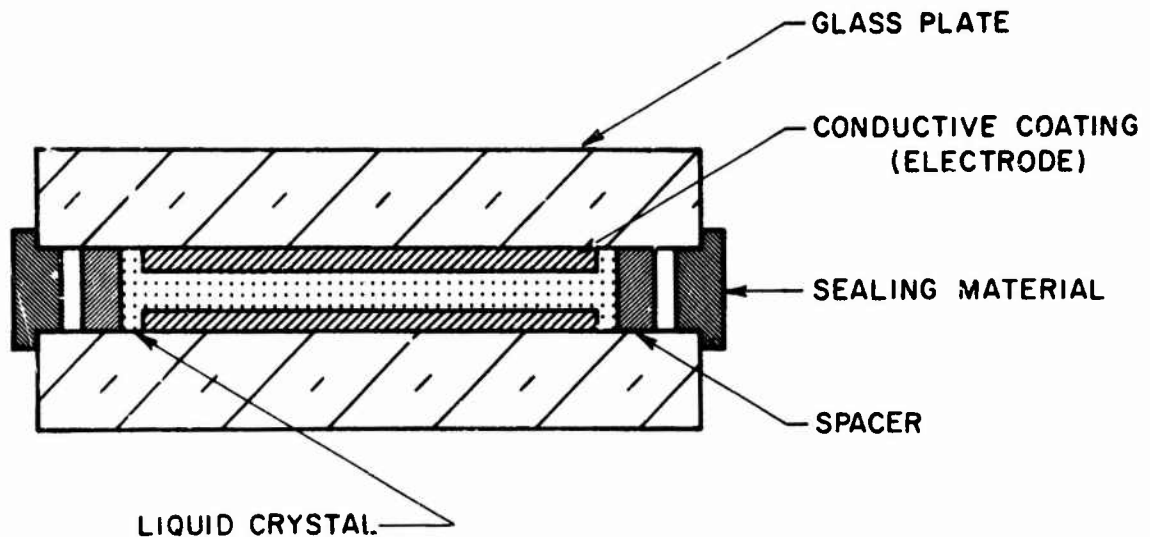
Figure 45:   Schematic cross-section of an LCD

Figure 46 schematically illustrates the crystal ordering
in the three generic liquid states--smectic, nematic, and
cholesteric.   The rod-shaped molecules are generally ori-
ented parallel to one another in the mesomorphoric (resting)
state.   The smectic has a layered structure, with all mole-
cules in each layer generally ordered in the same direction.
The nematic has all molecules ordered parallel, with no
apparent layering.   The cholesteric type has parallel orien-
tation of all molecules within a given layer, and a helical
rotation of molecular orientation from one layer to the
next.   Most displays use the nematic structure, although
other examples exist in particular applications.

SMECTIC   **NEMATIC**   **CHOLESTERIC**
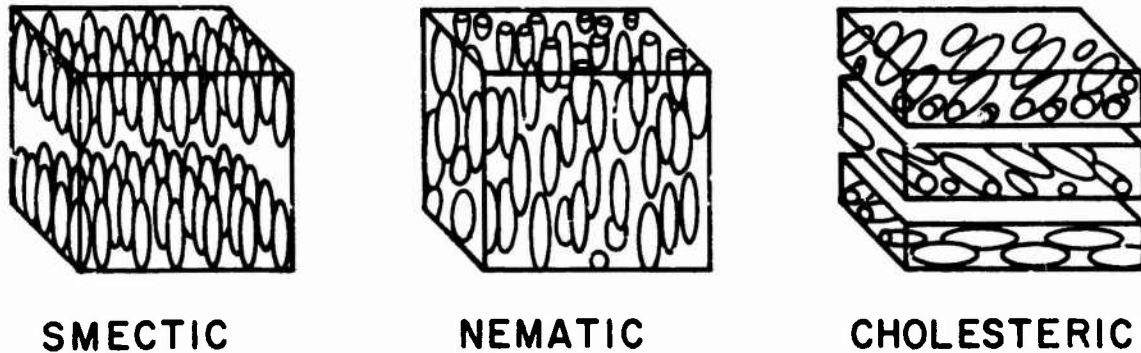
Figure 46: Schematic representation of three LCD types, from Sherr (1979)

3.2.6.1 Physical size and configuration

Research has been conducted on a wide variety of LCDs in many laboratories, both in the United States and abroad, most notably in Japan. As a result, it is not easy to categorize and generalize all materials, optical techniques, configurations, and interfacing approaches. For purposes of this discussion, however, it seems adequate to limit our description to the three most used LCD types--twisted nematic, distortion of aligned phases (DAP), and dynamic scattering.

The twisted nematic LCD concept is illustrated in Figure 47. The crystal molecules in the liquid are uniformly aligned in one direction (parallel to the surface), but have a twist angle of 90 deg between the top and bottom surfaces.

- 150 -

Thus, with no applied voltage, there is a gradual rotation through 90 deg of the molecule axes from one surface to the other. When an above-threshold voltage is applied across the electrodes, the fluid "untwists" and the molecular axes align perpendicular to the surfaces. The threshold voltages for various twisted nematic materials range from 1 to 5 volts.
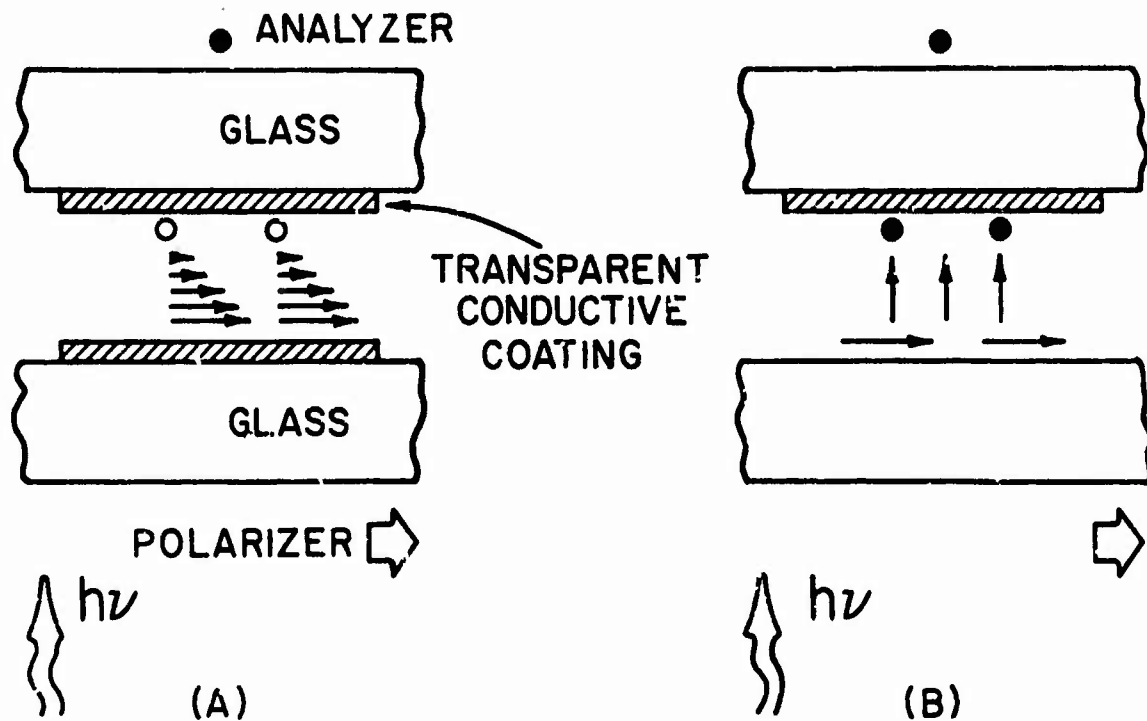
Figure 47: Side view of twisted nematic LCD, with zero applied voltage (A) and suprathreshold voltage (B), from Goodman (1976)

Light incident on one surface of the twisted nematic LCD is generally passed through by a polarizer (Figure 47).

- 151 -

Without an applied voltage, the polarized light is rotated about 90 deg as it passes through the fluid. If the second polarizer (the "analyzer" in Figure 47) is oriented 90 deg to the first polarizer, then maximum light transmission occurs. As the applied voltage exceeds threshold, the nematic molecules untwist and the transmission axis of the light becomes increasingly nonparallel to the analyzer, resulting in reduced transmission. Alternatively, one can mount the polarizer and analyzer axes parallel, so that no transmission occurs in the relaxed state, and transmission increases as the applied voltage exceeds the threshold value. Thus, depending on the analyzer and polarizer orientations, application of a suprathreshold voltage can produce either a black-on-white or a white-on-black display. In either case, the geometry of the sealed LC material determines the size and shape of the displayed element.

The DAP type LCD is not as frequently used as the twisted nematic, but it has some advantages and potential properties of interest. In the DAP LCD, the LC molecules, under an applied voltage, cause the incident light to leave the display at a different angle and polarization than that at which the light entered. By using a polarizer/analyzer combination, as in the twisted nematic, the pattern of transmitted light can be controlled by the applied voltage. Moreover, and of potentially greater interest, the DAP effect can achieve a color change in the transmitted light as a function of the applied voltage (Sherr, 1979).

Finally, the dynamic scattering type of LCD is a result of the fluid turbulence that occurs as a voltage is applied to nematic liquid cry...ls that have a high electrical resistivity (between 10^8 and 10^{10} ohm-cm). The process is illustrated in Figure 48. With a subthreshold or no applied voltage, the molecules are uniformly aligned parallel to the surfaces. At the threshold voltage, a periodically structured vortical flow pattern occurs. Above the threshold voltage, the fluid flow becomes unstable and aperiodic, permitting light transmission (Goodman, 1976). This transmitted light is strong, nonstable, diffuse, and scattered; thus, the name dynamic scattering.
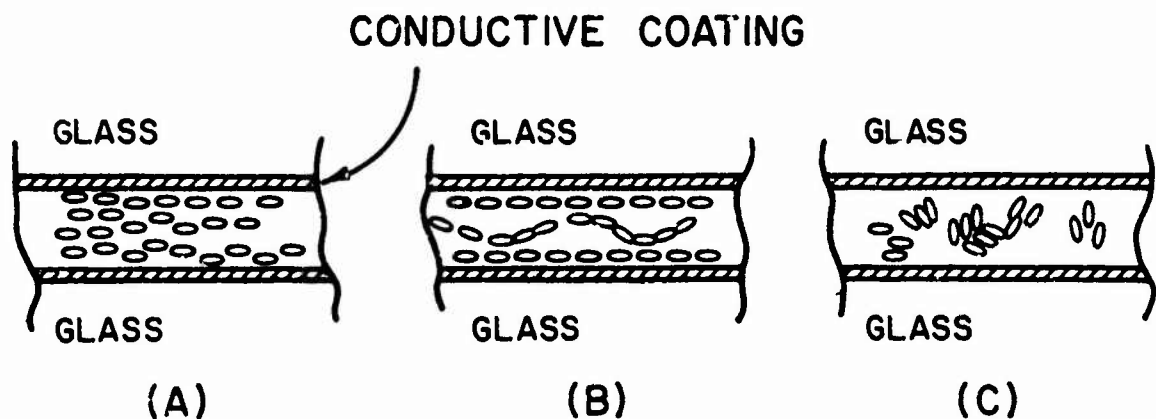
Figure 48: Steps in formation of dynamic scattering LCD with uniform parallel orientation. Initially the molecules are parallel (A) with zero voltage, changing into a sinusoidally arranged orientation with increasing voltage (B), and finally into the scattering mode with suprathreshold voltage (C). After Goodman (1976).

- 153 -

The size of an LCD is of course dependent upon the required fabrication techniques. The dynamic scattering mode is the easiest to use for large displays, which have been made 30 cm on a side (Sherr, 1979), since the alignment problem is less critical for dynamic scattering than for twisted nematics.

At the other extreme, the minimum pixel size is limited only by electrode and glass fabrication technology. LCDs with pixels as small as 0.25 mm have been achieved repeatedly on small 25 mm on a side panels.

Within these limits, there appears to be no serious limitation to the size and shape of LCD elements.

3.2.6.2 Power and voltage requirements

One of the most desirable attributes of the LCD is its very low power requirement. While it is possible to operate LC cells with DC voltages, the failure rate is much greater than with AC excitation. The usual driving frequency is 30 Hz, although frequencies as high as 1 kHz have been used (Sherr, 1979). At any voltage or frequency, the current requirement is usually very small, on the order of microamperes, so that a small alphanumeric display might operate for a year or two on a battery, drawing only microwatts of power.

Because the voltage requirements (e.g., 1-5 V) are compatible with low power MOS (metal oxide semiconductor) tech-

nology, sophisticated instruments having LC displays can operate long periods on batteries. Further, the voltage and current requirements are easily handled by integrated circuits (ICs), thereby achieving a packaging convenience.

Examples of current requirements are 10 µA for the Hamlin twisted nematic LCD having 7 segments, and 30 µA for the Hamlin 16-segment LCD, each having all segments on (Sherr, 1979). Thus, current and voltage requirements are typically very low and create no problems, even in battery-operated applications.

For purposes of generalization, dynamic scattering LCDs typically require 0.1 to 1.0 mW/cm^2 at a drive voltage of 15 V and at 30 Hz. Twisted nematic LCDs require about 5 V on the average, at 30 Hz, and dissipate approximately 1 µW/cm^2.

3.2.6.3 Spectral emission

As indicated above, the orientation of the optic axis of nematic liquid crystals is easily controlled by an applied voltage. In addition, the birefringence of nematic liquid crystals is quite large. (Birefringence can be considered as the separation of light into separate wavelengths, similar to that of white light by a prism. The resulting separate wavelengths are polarized differently.)

Wavelength variability results from the application of a voltage, V, in excess of the threshold voltage, V_o. For a 10 µm thick nematic LCD with crossed polarizers, the locus

of chromaticity coordinates for voltage ratios between 1.0 and 3.0 is illustrated in Figure 49. Clearly, all these colors are of low to medium saturation, except for the blue at $V/V_O = 1.2$.
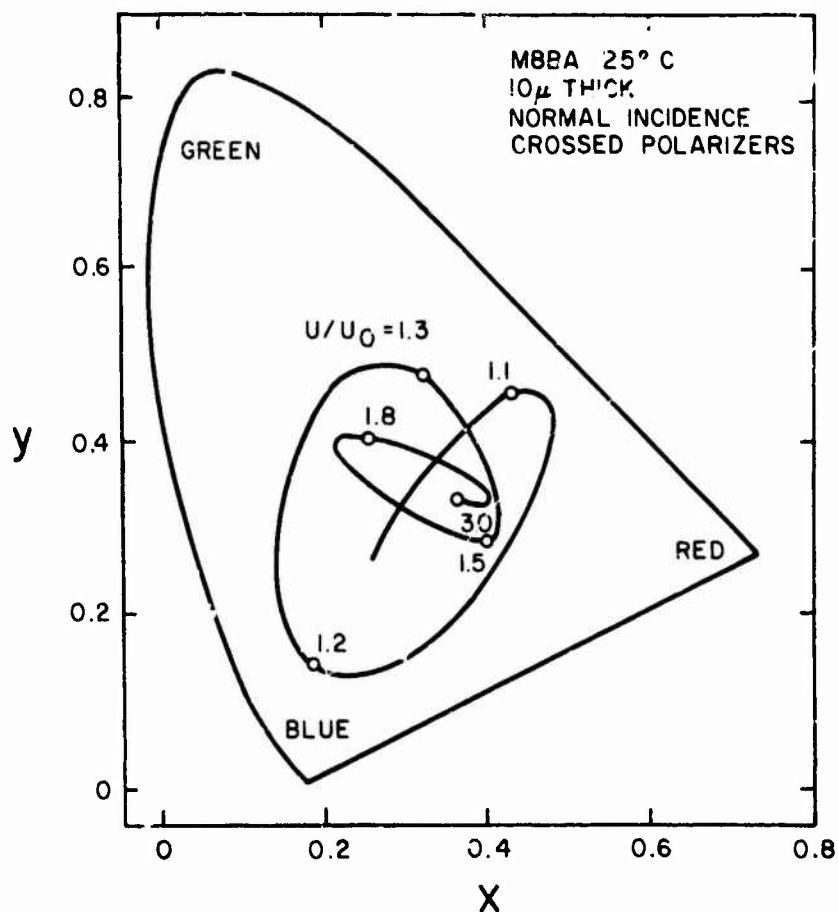
Figure 49: Calculated effect of voltage on transmission wavelength for tunable retardation effect, from Scheffer (1976)

Unfortunately, the dominant wavelength is quite sensitive to temperature changes, as the temperature directly affects the elastic constants, dielectric constants, and refractive indices, all characteristics of the nematic phase (Scheffer, 1976). Finally, the dominant wavelength is also affected by

the viewing angle. For example, at a constant angle of incidence of 30 deg from the normal, the dominant wavelength ranges from blue to green to red (Scheffer, 1976).

In the above example, color is controlled by the nematic LC cell which also was the birefringent layer. Sensitivity to angular orientation, voltage, thickness, and LC material type is very high for this type of cell. Another color-variable approach is possible, in which passive birefringent layers are separate from the LC electrooptic cell. In this type of arrangement, the display can be switched only between two discrete colors. However, the advantages are that there are no temperature or thickness effects on the resultant color (Scheffer, 1976, p. 59). Thus, one can get two discrete hues, one with the field voltage on and the other with the field off. Pairs of these complementary hues, when added together, produce white light. If this passive birefringence approach is applied to a reflective LCD rather than to a transmissive LCD, somewhat more saturated colors are obtainable (Scheffer, 1976, p. 64).

3.2.6.4 Luminance

The luminance of a light modulating transmissive or reflective display is a function of the ambient illuminance and the modulating ability of the display. Therefore, it is not particularly meaningful to consider light modulating LCDs in terms of their luminance characteristics. However,

- 157 -

one can discuss them in terms of the intensity of the emerging light compared to the light passing, for example, through the two parallel polarizers (polarizer and analyzer) in a nematic LCD.

This relationship is given by Goodman (1973) as:

$$I = I_p \sin^2(2\phi) \sin^2(\delta/2), \tag{16}$$

where

I = the intensity of the light emerging from the LCD,

I_p = the light transmitted through the two parallel polarizers (roughly the input illuminance, attenuated slightly by the optical density of the polarizers),

ϕ = the angle between the input light optical vector and the projection of the director (IC cell axis) on the plane parallel to the cell walls,

δ = $2 \pi d \Delta n/\lambda$,

d = the cell thickness,

Δn = the voltage-induced birefringence, and

λ = wavelength of the ambient illuminance.

Under these conditions, the light transmitted, I, is maximum at ϕ = 45 deg, and this parameter can be somewhat controlled in preparation of the cell wall substrates (Goodman, 1973). While many design variables can obviously affect the total transmission through the LCD, representative measured data indicate that the maximum transmission is 50%, even with ideal polarizers which absorb half of the incident light.

The dynamic scattering LCD mode, on the other hand, requires no polarizers, and therefore has greater transmission of the ambient illuminance. The price paid for this higher transmission (80-90%) is greater power consumption and lower contrast ratio.

3.2.6.5 Luminous efficiency

The concept of luminous efficiency for light modulating displays is, for reasons similar to those of the luminance discussion above, not particularly meaningful. Again, however, it should be noted that while the transmission of the display is no greater than about 50%, the power cost of the input illuminance can be essentially zero, as in daylight.

3.2.6.6 Element size, shape, density

The element size and shape in an LCD are limited largely by fabrication techniques, typically photomasking. In an individual element containing an active electronic driver, as in a matrix array, the element size may also be somewhat limited by the electronic miniaturization possible.

At the present time, fabrication has been demonstrated with element densities of at least 3.93/mm, with a spacing between elements of approximately 0.012 mm. Thus, the element size can be made as small as 0.254 mm. with an inter-element space as small as 0.012 mm, for a percent active area of over 91%. Of course, large element sizes are easier

to fabricate, and elements can be made as large as 5 to 10 mm on a side if necessary.

While extremely high density element fabrication has not yet been required, Wysocki, Becker, Dir, Madrid, Adams, Haas, Leder, Mechlowitz. and Saeva (1972) estimate that an element density of 10 to 20 per millimeter should be easily achieved.

The element shape is controlled also by the photofabrication process, and can be any shape desired, including square at the 4/mm density.

3.2.6.7 Contrast and dynamic range

The contrast of an LCD is limited by the level of the ambient illuminance and the transmission of the cell in its ON state versus its OFF state, typically defined as the ON/OFF ratio. Thus, one usually finds empirical contrast data plotted in the form of contrast ratio (ON/OFF) against viewing angle or driving voltage.

At the present time, contrast ratios of 20:1 to 30:1 are easily obtained, while contrast ratios of nearly 50:1 (modulation = 96%) are possible with careful selection of different front and back polarizers.

The contrast ratio also depends on the azimuth and elevation viewing angles of the display, as illustrated in a recent Hughes matrix display developed for head-up display applications.

Within the limits of the available contrast ratio, gray levels can be controlled by the driving voltage. However, the luminance/voltage gradient is quite steep and voltages must therefore be carefully controlled to avoid inappropriate gray levels. Sixteen addressable gray levels have been successfully demonstrated by Kaneko, Kawakami, and Hanmura (1978).

3.2.6.8 Uniformity

LCD uniformity is largely dependent on cell thickness, molecular alignment, and voltage control. Molecular alignment and cell thickness tolerances are not nearly as critical for dynamic scattering LCDs as they are for DAP and twisted nematic cells. For this reason, larger LCDs can be made more easily with dynamic scattering operation. For twisted nematics, cell thickness should be held within $\pm$ 0.25 μm to avoid perception of nonuniform colors or gray levels.

Larger arrays of LCDs are generally made by butting together four or more matrix arrays. Thus, a 350 x 350 element array is created by a 2 x 2 matrix of 175 x 175 element arrays, usually resulting in an LCD of about 90 mm square. Great care and accuracy are required to eliminate edge butting discontinuities in this approach. (Since the X,Y leads of all four 175^2 element arrays can be brought to the edges, no interior lead problems exist. Such interior lead problems exist with more than 2 x 2 section arrays.)

- 161 -

A different type of nonuniformity arises with off-axis viewing of LCDs. In general, the ON/OFF contrast varies greatly with the angle of view relative to the normal and relative to the incidence angle of the ambient illuminance. Figure 50 illustrates this problem for a transmissive cross-polarized twisted nematic LCD. Transmission in the OFF state is maximal normal to the surface (0 deg). As the voltage is increased, the maxima and minima (OFF and ON) transmisivities change. Thus, with higher driving voltages greater contrast is obtained further off axis, but the absolute contrast is reduced. In general, different LCDs have different usable viewing angles: dynamic scattering ± 90 deg; twisted nematic, ± 45 deg; DAP, ± 45 deg (Sherr, 1979).

3.2.6.9 Temporal characteristics

The major limitation to widespread use of LCD displays is the relatively slow response time to either an "on" or an "off" command. Sherr (1979) gives response times of 100 to 500 ms for dynamic scattering LCDs, 100 to 300 ms for twisted nematics, and 50 to 200 ms for DAP LCDs. Goodman (1973), in offering the theoretical basis for LCD rise and fall times, indicates that rise times of 10 ms are possible with twisted nematics and 200 to 300 ms for dynamic scattering LCDs. Goodman (1973) also states that decay (fall) times of 400 ms and 100 ms are typical for the twisted nematics and dynamic scattering LCDs, respectively. Since

- 162 -
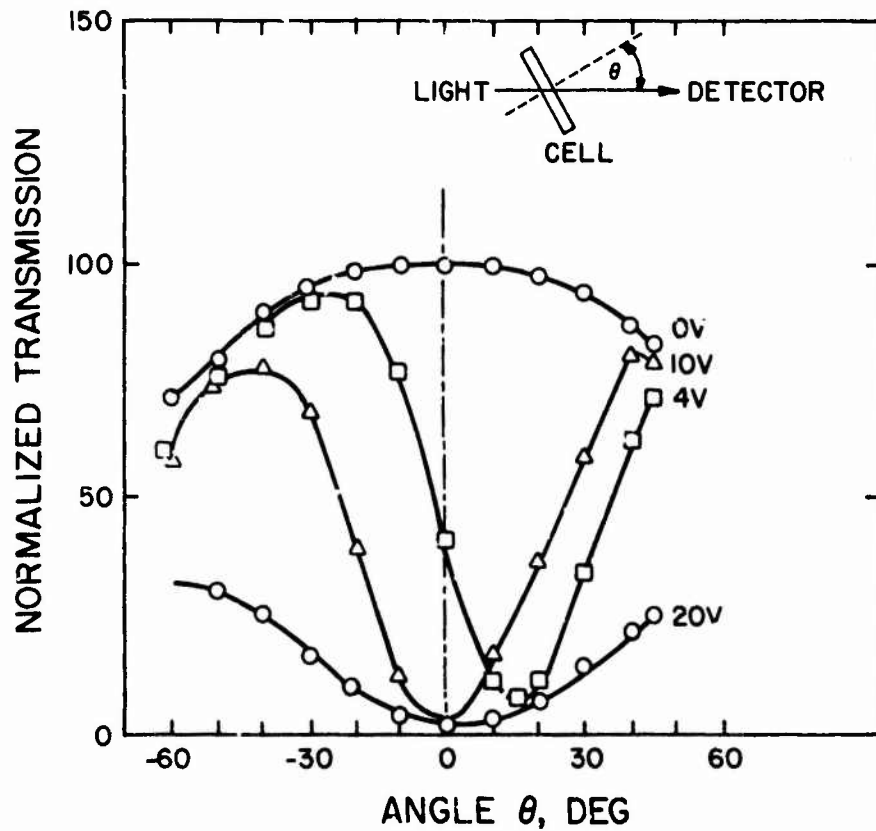
Figure 50: Effects of applied voltage and viewing angle on light transmission through a twisted nematic LCD, from Goodman (1976)

the response times are generally due to fluid viscosity, which in turn is affected by temperature, LCDs can become quite sluggish at low temperatures, particularly below 0 deg C.

While switching rates of 60 Hz have been used for LC television (LCTV) in keeping with the interlaced broadcast signal, moving objects in LCTV appear blurred and have a "trail." Some LCTV advocates believe that rise times on the order of 100 ms and fall times of 50 ms or less produce "acceptable" smear in video images. In part, the acceptability of such an image will depend on the observer's task.

- 163 -

3.2.6.10 Addressing/driving interfaces

Alphanumeric displays are easily driven by low voltage MOS devices, as indicated above. Matrix addressing of LCDs, however, is much more difficult because LCDs lack the well-defined thresholds, high contrast ratios, and rapid response desired of matrix-addressed displays.

As discussed above, the threshold voltage varies with temperature and, in twisted nematics, the contrast ratio varies with viewing angle. Sherr (1979, pp. 257-269) has an excellent discussion on alternate LCD matrix addressing techniques, as does Goodman (1973). Goodman concluded that a feasible approach might be a matrix of MOS FETs (metal oxide semiconductor, field effect transistor) to drive each LCD element, or a less expensive array of thin film transistors (TFTs). More recently (Ernstoff, 1978) Hughes has successfully developed a 200 x 200 element MOS driven LCD with line-at-a-time addressing and a 30 Hz, flicker-free refresh rate. Serial-to-parallel conversion is used to make the addressing scheme compatible with broadcast television signals.

Many of the interface addressing problems of LCDs have been solved, although a large amount of research is still dedicated to the remaining issues in this area. It appears that the attractiveness of the LCD itself has caused the needed activity to obtain compatible addressing/driving approaches.

3.2.6.11   Cost

The LCD has become quite inexpensive because of volume
production for use in calculators, wristwatches, and other
small units. Only when matrix addressing is required, as in
LCTV, does the cost become substantial. In such cases,
addressing and driving costs for LCDs are similar to those
for other nonemitters, such as electrophoretics and electro-
chromics. Tannas (1978) estimated such costs to run between
$10 and $50 per square inch for the TFTs or FETs needed for
addressing. In addition, power supply, line driver, pinout,
and serial/parallel conversion costs may be added, depending
on the application.

Estimated production costs for LCTV panels are about $15,
and for a complete broadcast-compatible LCTV system $150 in
large volumes (Kaneko, 1978) using a double matrix address-
ing approach (Kaneko et al., 1978). On the other hand,
production costs for an LCTV system using large-scale inte-
gration with FETs or MOS devices may run as high as several
thousand dollars (Kaneko, 1978).

3.2.6.12   Utility for display-type applications

The LCD is, for many applications, an ideal choice for
single or multiple character alphanumeric readout. The
character can be made any size, the contrast is typically
adequate, costs are very low, and voltage/power requirements
are compatible with battery sources.

Graphic and video displays, on the other hand, are just becoming feasible with LCDs. Matrix addressing remains expensive and complex. Moreover, the contrast sensitivity to voltage and incidence angle variation is a severe limitation to perceptual uniformity. As a result, applications requiring matrix addressing but only ON/OFF driving (no gray scale) can be considered suitable for LCDs, but gray scale requirements, especially under varying viewing conditions, remain difficult to meet.

Compounding the difficulty for television displays using LC is the slow response time, leading to significant blur of moving images.

One of the more interesting and potentially useful applications of LC technology is in the aircraft head-up display (HUD). Present designs using CRTs are limited in resolution, contrast, and positional linearity and stability. The substitution of diffraction optics (rather than refraction optics) has greatly improved HUD image quality. Using an LCD as the image source, replacing the CRT, in a diffraction optics HUD should result in improved contrast and positional accuracy. Development efforts along these lines currently exist at Hughes Aircraft (Ernstoff, in press), and were discussed in Section 2 of this report.

3.2.6.13   Future technology projections

As the demand or potential market increases for matrix addressed LCD panels, the developments required for reasonable production costs should progress. A 500 x 500 element array is possible and should be available in prototype form during 1980. Larger sizes are similary feasible, given the need for such. However, because of addressing complexities, they shall probably remain very expensive for the foreseeable future.

Areas requiring additional research and development appear to be the following. First, rise and fall times must be shortened considerably for acceptable TV or fast graphics applications. New materials with < 10 ms response times appear possible. Reductions in cell thickness also hold promise here.

Secondly, techniques are needed to reduce the effects of viewing angle and incidence angle on contrast. This problem must be solved to permit multiple viewer applications.

Thirdly, while LC light valves have been used, and are available, the contrast level is poor, color saturation is low, resolution is unimpressive, and overall luminance is marginal in all but a totally darkened room. Additional development of the LC light valve mechanization is needed before this concept will have diverse utility. Therefore, large screen displays using LC materials may ultimately prove popular, but current models are very limited in applications.

In spite of these limitations, LCDs are growing in popularity and frequency of usage. The wide variety of research being conducted on LCD applications suggests that the technology's popularity will continue to increase.

3.2.7 Electrochromic Displays

The electrochromic (EC) display is another type of light modulating display device. It has no light generating or emitting proporties; rather, it modulates the ambient (or reflected) illuminance. As compared to the liquid crystal technology, however, the EC device is generally transparent and absorbs only a selected portion of the visible spectrum upon application of an electric field. The apparent color of the "on" portion of the display is determined by the electrochromic material. While one type of EC material exhibits a color in the "on" state and is transparent in the "off" state, other types exhibit one color in the "on" and a different color in the "off" state.

3.2.7.1 Physical size and configuration

The general form of an EC display cell is shown in Figure 51. The front of the cell is a transparent glass plate upon which has been deposited a thin transparent electrode and the electrochromic film. The EC film is usually about 1 μm thick and consists of an EC material such as tungsten triox-ide (WO_3) (Sherr, 1979) or other metal oxides (Zeller,

1976). The film is usually backed with an electrolyte containing a mixture of water, glycerol, sulfuric acid, and a white reflecting powder. Next follows a layer of insulating material, and last a back electrode which is usually highly reflective.
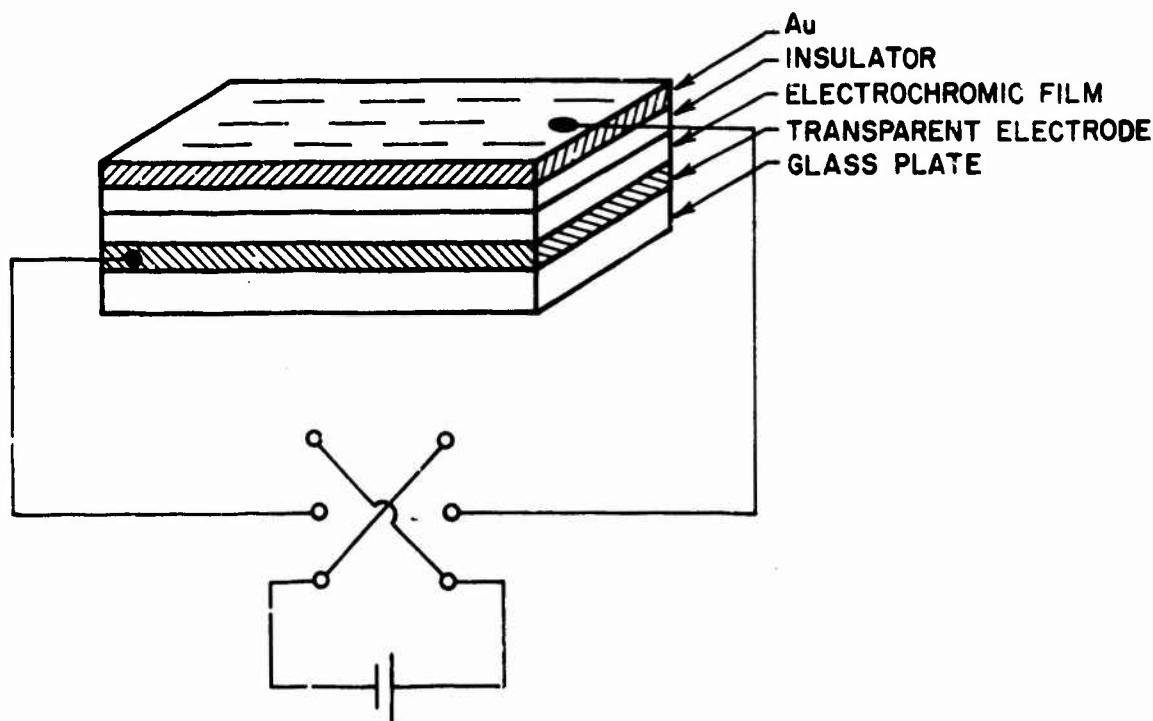
Figure 51: General EC cross-section, from Sherr (1979)

This type of EC cell is normally white because of the white reflective powder in the electrolyte. However, when a negative voltage is applied to the front (transparent) electrode, the WO_3 turns blue and results in a blue-on-white display. By shaping the locations of the EC film, one can of course control element or character shapes. An EC numeric display is illustrated in Figure 52.

Figure 52:   Photograph of Sharp numeric EC, courtesy of
              Sharp Corporation

It is also possible to create transmissive EC displays by
eliminating the white powder in the electrolyte and replac-
ing the back, reflecting electrode in Figure 51 with a
transparent electrode.

EC displays can be formed in virtually any desired size
and shape, which are controlled by the accuracy of EC film
deposition and electrode placement.

3.2.7.2   Power and voltage requirements

The attractiveness of the EC display is due in large part
to its low voltage and power requirements. Change from the
white to the color state may be obtained with about 1.5 VDC
and about 5 mA of current. To return from the color state

to the white (or transparent) state, a process termed "bleaching," the polarity must be reversed and a current of about 15 mA is needed. However, the current flow is required for only a short period of time, as shown in Figure 53. Thus, the total power used, because of the low duty cycle, is well below that required by LED displays, and is comparable to that used by dynamic scattering LCs.
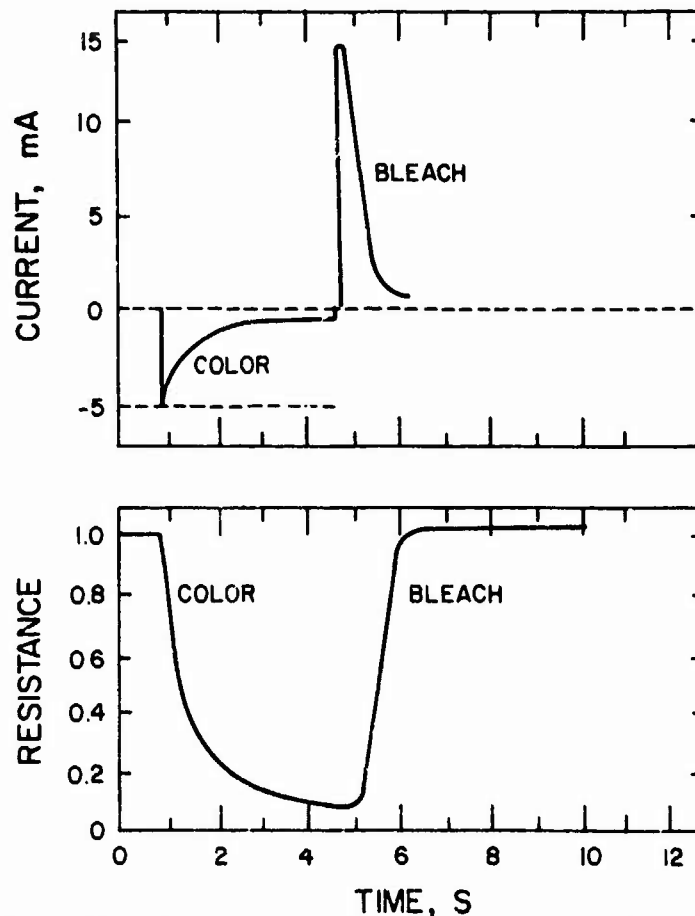
Figure 53: Time/current characteristics of WO_3 ECD, from Sherr

The current flow in the EC cell is a function of the area of the display and the voltage, as well as of the EC mate-

- 171 -

rial (Sherr, 1979).   The driving voltage is direct current,
which simplifies the drive circuitry and makes the device
.ore compatible with battery power, such as in wristwatches.

In general, an EC device with 10 mm^2 active area and a
switching rate of 1 s^{-1}, as a wristwatch, will have a power
consumption of about 100 µW,   compared to < 1 µW for a
twisted nematic LC display of the same size and switching
rate (Zeller, 1976).

3.2.7.3   Spectral emission

Tungsten oxide EC cells, when turned "on," absorb light
in the longer wavelengths, typically above 600 nm.   Thus,
the "on" state is characterized by a broadband blue color.
Most of the WO_3 absorption lies outside the visible spec-
trum, but some variation in color is possible by careful
selection of the film materials.   As Chang (1976) has indi-
cated, element selection can result in characteristic wave-
lengths ranging over the entire visible spectrum, although
at varying efficiencies.

3.2.7.4   Luminance

As with other light modulating display technologies, it
is not meaningful to talk about the luminance of the EC
device.   However, the optical density of the "on" cell, and
therefore the luminance contrast, has been studied for the
EC cell and has been found to relate to the flow of charge
across the cell.

The optical display density is propotional to the current and time the current is on, and is inversely proportional to the area. Figure 53 illustrates a representative reflectance for a WO_3 EC cell with an active area of 0.33 cm^2. Note that minimum reflectance is about 0.1, for an effective luminance modulation of $(1 - 0.1)/(1 + 0.1) = 82\%$.

3.2.7.5 Luminous efficiency

The luminous efficiency of the EC display is not a meaningful concept because it is a light-modulating device.

3.2.7.6 Element size, shape, density

As with the LC technology, EC elements can theoretically be formed in virtually any size, shape, or density, limited only by accuracies of film deposition and electrode placement. As a practical matter, however, little attention has been paid to this area of research because the relatively slow response of the EC material precludes its use for dynamic displays which change more frequently than once or twice per second. Thus, it is incompatible with most graphics and television-type applications.

3.2.7.7 Contrast and dynamic range

As with the LC display, the dynamic range of the EC cell is limited only by the ambient illuminance level. Within this range, luminance modulation of about 90% is readily attainable.

- 173 -

It must also be realized that the EC cell provides a chrominance contrast in addition to its luminance contrast. That is, the cell not only changes in optical density, but also changes in dominant wavelength from the "off" state to the "on" state. Unfortunately, a suitable "effective constrast" measure combining both luminance contrast and chrominance contrast does not currently exist, although many developers of EC displays attest to their increased "readability" over LC displays due to the chrominance contrast contribution. This argument appears quite valid and a means of measurement of such a combined contrast is truly needed.

3.2.7.8   Uniformity

While one might expect that the density of the EC cell, and therefore its optical uniformity, is highly related to the thickness of the EC film layer, the author knows of no measured data on this possible effect. It might be expected that EC film thickness tolerances are less critical than those of LCs because the EC does not rely on molecular orientation; however, research will be needed to support this notion.

3.2.7.9   Temporal characteristics

The EC display has inherent memory. Once it is turned "on," the color will remain on for days or until it is bleached by application of the reversed voltage. Typically,

- 174 -

the rise time is on the order of seconds, and the bleach time about 1 s (Figure 53), although Chang (1976) reports that rise and fall times as low as 100 ms are possible.

The temporal characteristics in terms of response time of the EC display are not impressive. The temporal characteristics that are desirable, however, relate to the facts that (1) the current duty cycle to change state is very small, and (2) the EC cell has inherent memory, i.e., it does not require an applied voltage to remain either "on" or "off."

3.2.7.10   Addressing/driving interface

According to Sherr (1979, p. 494), there are no commercially available matrix-addressed EC displays. While the driving/addressing problems would be relatively simple, requiring only a bidirectional DC driver, the response time of the EC cell precludes its consideration in matrix displays at the present time. Even if response times were significantly reduced, the typical nonsharp threshold of the EC cell would make matrix addressing difficult.

3.2.7.11   Cost

The cost of the EC display should be quite low. No particularly stringent tolerances are required, the driving electronics are inexpensive and simple, and the material costs are low. Further, the fabrication techniques are not particularly involved or complex. While no cost data are

available, it is estimated that the fabrication cost, in comparable production quantities, of the EC display should be less than that of the LC display.

3.2.7.12   Utility for display-type application

EC displays have found limited acceptance in alphanumeric display applications, mostly in battery-driven situations such as wristwatches and calculators. Because of their slow response rate and power drain upon switching, wristwatch applications are limited largely to displays of hours and minutes, and not seconds. If a watch were to display hours, minutes, and seconds, the life of the battery driving the EC display would be about 4 months, compared to about 2 years for the LC display.

Because of its poorly defined voltage threshold and slow response time, the EC display seems unsuitable for graphics or television-type matrix displays. The author also knows of no attempt to fabricate large-screen EC displays, either real image or projected image, although there appear to be no technological reasons to preclude such a development.

3.2.7.13   Future technology projections

The subjectively pleasing color contrast of the EC display will probably sustain some limited amount of research for commercial applications. However, until the threshold and response time problems are solved, no significant prod-

uct development is very likely. Future research will probably emphasize these two problem areas.

3.2.8 Electrophoretic Displays

The electrophoretic induced display (EPID) is also a light modulating, rather than a light emitting, display. It results from the process of electrophoresis, the movement of charged particles suspended in a liquid by the application of an electric field. The pigmented particles are selected to be a different color or optical density than the suspending liquid, so that the migration to the front surface of ʌe display cell permits the oberver to "see" the particles, whereas migration to the rear surface of the display causes the observer to see only the suspending liquid. Selection of colors or optical densities of the pigmented particles versus the suspending liquid determines the contrast or chromaticity of the EPID. The EPID is one of the newer display technologies, traceable only to the recent work of Ota, Ohnishi, and Yoshiyama (1973).

3.2.8.1 Physical size and configuration

The EPID concept is shown schematically in Figure 54. Like many other solid-state displays, it is essentially a transparent sandwich, with the front and rear plates coated with conducting electrodes. The cavity created by spacers between the two transparent electrodes is filled with a

fluid composed of a small pigmented particle suspension in a
dense liquid.
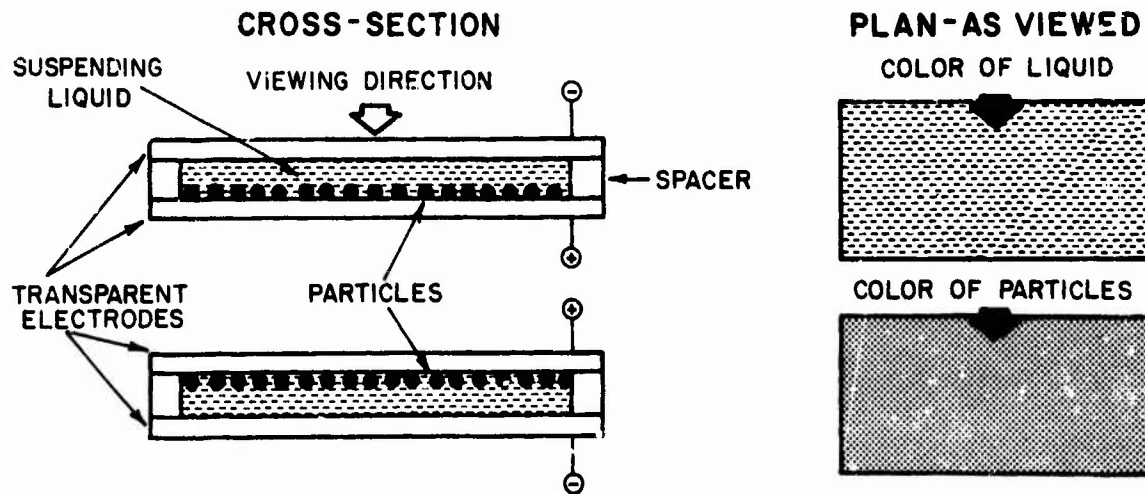
Figure 54: General EPID cross-section, from Lewis (1976)

Application of an electric field across the electrodes
causes the particles to migrate toward one or the other
electrode. The rate of migration of the particles depends
on several factors, among them the particle size, the cell
thickness, and the field voltage.

Because of the recent discovery of the EPID, little par-
ametric development in configuration has taken place. Five
millimeter digits have been successfully fabricated, as has
an experimental plain panel 300 x 150 mm (Ota, Sato, Tanaka,
Yamagami, and Takeda, 1975). The fabrication process is not

particularly difficult, however, and any reasonably sized panel should be possible. Individual element sizes will not be limited by particle size, since the particles need only be slightly larger than light wavelengths (e.g., 750 nm) to adequately change optical density.

3.2.8.2   Power and voltage requirements

Operating voltages explored thus far have ranged from $\pm$ 2.5 VDC to $\pm$ 100 VDC. It appears that cell lifetime is longer for greater voltages. Further, the combination of dye concentration, suspension thickness, and voltage determines the contrast level, such that there is an optimal voltage for each cell configuration (Sherr, 1979). A representative optimal voltage would be about 40 VDC for a pigment (particle) concentration of 30 mg/ml and a dye concentration of 4 mg/ml. A representative cell driven at 1 VDC/1 μm thickness would require about 0.3 μA/cm^2 current for a contrast with a cell thickness of about 100 μm. That is, this cell would draw (100 μm)(1 V/μm)(0.3 μA/cm^2) = 30 μW/cm^2 power. For lower voltages, < 5 μW/cm^2 seems feasible (Dalisa, 1977).

3.2.8.3   Spectral emission

EPIDs can be fabricated in a variety of color combinations, ranging from black/white to blue/white, black/yellow, red/yellow. Other combinations are possible, depending only

- 179 -

on the particle and dye choices. To date, no spectral meas-
urement data have been published.

3.2.8.4 Luminance

As with other light modulating display technologies, dis-
play luminance is a function of the ambient illuminance and
the light attenuation of the display. Thus, the contrast
measure is more meaningful and will be discussed subse-
quently. Subjectively, however, Lewis (1976, p. 234) has
noted that EPIDs are "easily legible at illumination levels
of 10 lux and improve . . . up to 10^5 lux."

3.2.8.5 Luminous efficiency

Luminous efficiency is also an inappropriate measure for
the EPID.

3.2.8.6 Size, shape, density

The size, shape, and density of EPID elements will be
limited by the accuracy of electrode placement and cell con-
struction. There seems to be no problem in fabricating
large cells up to 150 x 300 mm (Ota et al., 1975) nor as
dense as 5 elements/mm (Dalisa, 1977).

3.2.8.7 Contrast and dynamic range

The dynamic range of an EPID cell depends, of course, on
the ambient illuminance and the optical transmission through

the display. The optical transmission can be measured in the contrast ratio metric, which depends on dye concentration, pigment (particle) concentration, pigment and dye optical densities and light scattering properties, cell thickness, and applied voltage.

There is an optimal voltage for combinations of dye and pigment concentration. Similarly, the contrast ratio varies with EPID cell thickness and applied voltage. Contrast ratios up to 7 or 7.5 (modulation = 0.76) appear reasonable with the current, infant state of this technology. Lewis (1976) indicates contrast ratios in excess of 30:1 (modulation = 0.94) are possible with titania-based displays.

3.2.8.8   Uniformity

No data have been reported on the uniformity of EPID cells, again largely due to the newness of the technology. However, uniformity should be no significant problem as long as reasonably constant cell thickness and uniform suspension/pigment viscosities are maintained.

3.2.8.9   Temporal characteristics

Preliminary data indicate that rise and fall times for the EPID, at optimal thickness and voltage, can be on the order to 10 to 20 ms, although 100 ms may be more likely in practical devices (Sherr, 1979).

Lewis (1976) also points out the distinct advantage that the EPID may retain an image up to several months after the voltage is removed, i.e., it has intrinsic memory as long as the viscosities and densities of the dye and particle portions are comparable.

3.2.8.10   Addressing/driving interface

One of the niceties of the EPID is that it has a fairly stable threshold voltage, such that there is adequate discrimination for use in matrix displays based on the "half-select" principle (Lewis, 1976). This is usually accomplished by holding the front electrode at 0 V, and driving the individual elements at the threshold potential.   In larger arrays, or if greater than $\pm$ 15 VDC is desired to improve speed and contrast, then switching transistors can be placed at each matrix element. This approach, of course, becomes more costly.

3.2.8.11   Cost

Sherr (1979, p. 188) points out that the EPID technology is too new to estimate production costs. However, the simplicity of construction and the relatively low material cost suggest that EPIDs, in production, should cost on the same order as EC displays.

3.2.8.12 Utility

Laboratory prototypes have already been used to demon-
strate the feasibility of EPIDs for on-off displays, seven-
segment alphanumerics, and matrix displays. While these are
all feasible, the application of EPID matrix displays to
dynamic content, such as television, is probably precluded
by the response speed (e.g., 50-100 ms) of the device. Sim-
ilarly, applications involving large screen displays are
possible if matrix addressing and cost are not limitations.

In general, EPID displays are esthetically pleasing, due
in part to the color combinations available to the designer.
They should operate at all required temperatures, although
studies to date have been limited to -10 deg to +70 deg C.
Most importantly, they have a very wide useful viewing
angle, estimated to approach $\pm$ 90 deg (Lewis, 1976, p. 233),
apparently due to a nearly Lambertian transmission distrib-
ution (Goodman, 1975) Thus, there is potential application
for multi-viewer large-screen EPID displays.

3.2.8.13 Future technology projections

The EPID is too new to project its utility very far into
the future. Clearly, research is necessary on matrix
addressing at low cost, on materials compatibility for long
memory, and on reduced power consumption coupled with high
contrast and long memory. This type of research should con-
tinue because of the potential advantages and intrinsically
desirable characteristics of this display technology.

- 183 -

3.3 TECHNOLOGY COMPARISONS

The previous pages have attempted to summarize the seven flat-panel technologies. It is clearly of interest, at this point, to draw some comparisons among these technologies, and to provide summary data which might be used to select "the" appropriate technology for any given set of application requirements.

General comparisons are summarized in Tables 4 through 16, corresponding to the 13 categories on which the technologies were previously compared. Comparable data for the CRT are also given in these tables.

TABLE 4

Comparison of Physical Size and Configuration
Characteristics of the Display Technologies

Display Type	Display Size	Display Depth	Construction	Yield (in prod.)
CRT	0.75 m, diag.	$> 1.2 \times$ display diag.	not apprec.	high, exc. for blemish free
Flat-Panel CRT	$(16.26 \text{ cm})^2$	1.5 cm	not apprec.	unknown
LED	$(0.26 \text{ m})^2$ to date	~ 10 mm	not apprec.	high
EL	$(1.63 \text{ m})^2$	≈ 5 mm	not diff.	high
Plasma Discharge	$(54 \text{ cm})^3$	≈ 12 mm	not apprec.	high, now mature technology
LC	$(30 \text{ cm})^2$	1-2 mm	not apprec.	high
EC	no known limits	1-2 mm	not diff.	high
EPID	150 x 300 mm to date	1-2	not apprec.	unknown

TABLE 5

Comparison of Power and Voltage Requirements Characteristics
of the Display Technologies

Display Type	Voltage Requirements (V)	Power Consumption	Other
CRT	several, to 15 KV	$\leq$ 100 W	several voltages, complex
Flat-Panel CRT	20-30 VDC	100 mW/cm^2	none
LED	1.5-5.0 DC	1-5 mW/element, typ.	none
EL	30-650	2-6 mA/cm^2	none
Plasma Discharge	200 DC or 200 AC	400-500 mW/cm^2	none
LC	1-8, DC	1 mW/cm^2	none
EC	1.5 DC	100 mw/cm^2	none
EPID	2.5-100 DC	0.3 mW/cm^2	none

TABLE 6

Comparison of Spectral Emission Characteristics of the Display Technologies

Display Type	Dominant Wavelength	Spectral Dispersion	No. of Discriminable Colors Available
CRT	varies with phosphor	varies with phosphor	$\leq$ 20 with 3-gun CRT
Flat-Panel CRT	varies with phosphor	varies with phosphor	$\leq$ 20 with triad dots
LED	649, 632, 590, 550, 470 nm	red, orange, yellow, blue, green (wide, continuous)	5
EL	525, 585 nm; varies with phosphor	100 nm	2-3
Plasma Discharge	585 nm(neon); others less common	varies with phosphor or gas	$\leq$ 20 with full color; $\overline{1}$ otherwise
LC	varied	narrow	unknown
EC	varied	varied	unknown
EPID	varied	unknown	unknown

TABLE 7

Comparison of Luminance Characteristics of the Display Technologies

Display Type	Maximum Luminance (cd/m^2)	Minimum Luminance (cd/m^2)	Dependent on Resolution
CRT	30,000 300 typical	1-2	yes
Flat-Panel CRT	820	1-2	yes
LED	68,500	0	no
EL	10,000	0	no
Plasma Discharge	600	0	no
LC	n/a	n/a	no
EC	n/a	n/a	no
EPID	n/a	n/a	no

TABLE 8

Comparison of Luminous Efficiency Characteristics of the Display Technologies

Display Type	Lumens/Watt, maximum	Lumens/Watt, minimum
CRT	65, P-20	0.1, P-16
Flat-Panel CRT	2 (est.)	unknown
LED	4.2 (green)	0.006 (yellow)
EL	19	0.3 (AC)
Plasma Discharge	3.4	0.05 to 0.3 (typical)
LC	n/a	n/a
EC	n/a	n/a
EPID	n/a	n/a

TABLE 9

Comparison of Element Size, Shape, and Density
Characteristics of the Display Technologies

Display Type	Element Size, Minimum, mm	Element Shapes(s)	Element Density	Maximum Matrix Size Fabricated to Date
CRT	0.07 at 2.35σ	Gaussian	variable	0.75 m, diag.
Flat-Panel CRT	0.35 (est.)	Gaussian	to 3.15/mm	$(16.26 \text{ cm})^2$
LED	0.08	round, square	to 10/mm	$(204 \text{ mm})^2$
EL	0.05	selectable	to 19/mm	$(15 \text{ cm})^2$
Plasma Discharge	0.25	double diamond	to 3.27/mm	$(54 \text{ cm})^2$
LC	0.254	selectable	to 3.93/mm	$(89 \text{ mm})^2$
EC	unknown	selectable	unknown	n/a
FPID	unknown	selectable	to 5/mm	n/a

TABLE 10

Comparison of Contrast and Dynamic Range Characteristics of
the Display Technologies

Display Type	Maximum Modulation	Dependent on Ambient Illuminance	Light Emitter or Light Modulator
CRT	98%, at low luminance and low ambient	yes	emitter
Flat-Panel CRT	98%	yes	emitter
LED	96%	somewhat	emitter
EL	92%	somewhat	emitter
Plasma Discharge	95%	somewhat	emitter
LC	96%	yes	modulator
EC	90%	yes	modulator
EPID	94%	yes	modulator

TABLE 11

Comparison of Uniformity Characteristics of the Display Technologies

Display Type	Small Area	Large Area	Image Geometric Stability
CRT	good	fair, 50% rolloff	fair
Flat-Panel CRT	fair	fair to good	good
LED	good	large-area nonuniformity	very good
EL	fair	fair	very good
Plasma Discharge	good	good	very good
LC	good	fair	very good
EC	probably good	unknown	very good
EPID	probably good	unknown	very good

TABLE 12

Comparison of Temporal Characteristics of the Display Technologies

Display Type	Rise Time	Fall Time	Inherent Memory	Refresh Requirements
CRT	1 μs to > 1 ms, depend on phosphor	1 μs to > 100 s, depend on phosphor	typically not, except for storage CRTs	varies with phosphor
Flat-Panel CRT	same as CRT	same as CRT	same as CRT	same as CRT
LED	10 ns	10 ns	none	very high, 500 Hz typ.
EL	1 ms	10 μs to 1.5 ms	no	60 Hz
Plasma Discharge	100 ns	2 μs	yes	50-60 Hz
LC	50-300 ms	100-400 ms	yes	none
EC	0.1-1.0 s	0.1-1.0 s	yes	ncne
EPID	10-100 ms	10-100 ms	yes	none

TABLE 13

Comparison of Addressing/Driving Interface Characteristics of the Display Technologies

Display Type	Defined Threshold	High Voltage Requirements	Element Drivers Necessary (?)	Matrix Compatible
CRT	no, analog driven	12-15 KV	no	yes
Flat-Panel CRT	no, analog driven	no	no	yes, but not X-Y directly
LED	yes	no	no	no, very high refresh rate needed
EL	yes	yes, 50-600	typically	yes
Plasma Discharge	yes, very	no	no	yes
LC	unstable	no	usually	yes
EC	not very	no	n/a	no
EPID	yes, stable	no	probably not	probably

TABLE 14

Comparison of Cost Characteristics of the Display Technologies

Display Type	Known or Estimated	Production Quantity Estimate
CRT	known	< $10 for monochrome; < $100 for color
Flat-Panel CRT	estimated	high, major problem
LED	known	4-8¢/element singly; $6500 for $(512)^2$ array
EL	known	$3500 to $5000 for $(512)^2$
Plasma Discharge	known	$4000 to $9500 for $(512)^2$
LC	known/estimated	$15/TV panel; $150 for TV system
EC	estimated	low
EPID	estimated	low

Large displays (e.g., 0.5 m or more on a side) have only been made in CRTs, EL panels, and plasma panels to date. Conceivably, LED arrays and LCDs could be fabricated that large, but none are presently known to exist. Thus, if the application calls for a display larger than 0.5 m on a side, the designer is restricted to CRTs, EL panels, and plasma panels. On the other hand, if display depth is a problem or limitation, as in some vehicular installations, then the CRT might be necessarily eliminated.

TABLE 15. Comparison of Display-Type Applications Characteristics of the Display Technologies

Display Type	Single Alphanumeric	Matrix (graphic)	Matrix (TV)	Large-screen, direct viewing	Large-screen, projection	Gray Scale Available
CRT	possible, but not practical	yes	yes	yes, but low luminance	yes, with light valve	yes
Flat-Panel CRT	yes	yes	yes	no	no	yes
LED	yes	available, but too expensive	no, too costly	no	no	yes
EL	yes	yes	mono-chrome only	no	no	yes
Plasma Discharge	yes	yes	yes	no	no	yes
LC	yes	yes	yes	no	yes, as light valve	yes
EC	yes	no	no	no	no	no
EPID	yes	perhaps	doubtful	potentially yes	unlikely	yes

Secondly, some applications will be limited in terms of power and voltage requirements. CRTs require high voltages, and draw a fairly large amount of power. Thus, should the application require both large displays and low power, the EL and plasma panels become most attractive. For smaller displays, EC and EPID displays offer attractively low power consumption, as long as the application requires only alphanumeric readouts, for matrix and TV capabilities have yet to be proven for these technologies.

Color displays are available with several technologies, but full color control at nonflickering rates is still best achieved with CRTs, although plasma displays will provide a challenge for those applications in which low luminance is acceptable.

Under high ambient conditions, and for monochromatic displays, acceptable candidates are CRTs, LEDs, and ELs. Under some circumstances, light modulating displays will provide the greatest contrast, since their contrast generally increases with increasing ambient illuminance. However, under low ambient illuminance, light modulating displays require an internal illuminance source. One also needs to be careful in selecting high luminance displays which have resolution that varies with the luminance level at which they are driven, such as the CRT. Displays that have fixed element geometry, however, can successfully be luminance modulated with no effect on element density or resolution.

TABLE 16. Comparison of Future Technology Projections of the Display
Technologies

Display Type	Mature Technology	Major Improvements Likely	Major Improvements Required for Wide-spread Usage	R&D Trends
CRT	very	minor only	none	better uniformity resolution
Flat-panel CRT	no	posssible	cost, color	currently no R&D
LED	yes	unlikely	cost and uniformity	cost, luminous efficiency, color
EL	moderately	yes	cost, luminous efficiency	color, cost, luminous efficiency
Plasma Discharge	yes	yes	cost, color	color resolution, cost
LC	moderately	yes	rise/fall times, angular viewing	response times, addressing
EC	no	perhaps	response times, threshold	response times, threshold
EPID	no	yes	response times, addressing	response times

Solid state displays fall into this category, although the only flat-panel display that has a luminous efficiency reasonably close to the CRT is the EL panel.

If high resolution is demanded for a given application, then attention should be given to LEDs, ELs, and CRTs. Again, however, the EL and LED element sizes are invariant with luminance level, while the CRT small spot size will bloom with luminance increases.

All technologies are capable of good dynamic range under ideal illuminance conditions, but some (e.g., CRT, LED, EL, and plasma) achieve this high modulation under low illuminance, while others (e.g., LC, EC, and EPID) achieve it under high illuminance. Care must be taken, then, to consider carefully the operating ambient.

While uniformity requirements of the human observer are not well substantiated, one does find large variability in uniformity among the technologies surveyed. At the present time, in the absence of justifiable criteria for the three types of uniformity, the designer can only use good judgment in applying this criterion to device selection.

Some technologies are clearly fast reacting, and thereby permit rapid change of displayed information. These inlucde CRTs, ELs, plasma panels, and LEDs. In fact, the LED rise and fall times are so short as to cause significant refresh problems. On the other hand, if the information change rate is relatively slow, and especially if only small amounts of

alphanumeric data are to be displayed, then refresh problems are greatly simplified by using LC, EC, or EPID displays.

The issue of driving and addressing a matrix display is complex, and beyond our intended level of discourse. Nonetheless, it is clear that some technologies are much more compatible with addressing and driving flexibility than are others. For example, CRTs are relatively easily driven in stroke, vector, or raster modes, and the techniques are well defined and proven. LEDs, on the other hand, because of their very short rise/fall time, are virtually impossible to use in a large matrix display. EL, plasma, and LC displays have been successfully used in matrix displays, for both vectorgraphics and TV applications. Further improvements are likely in each of these latter three technologies for matrix use.

Costs of surveyed technologies vary considerably, ranging from very small amounts for EC, LC, LED, and EPID alphanumeric readouts, to large amounts for LED, EL, and plasma matrix displays. At the present time, the CRT is by far the cheapest and most flexible of the matrix display types, offering inexpensive display of alphanumerics, graphics, and TV. While some displays have been sold in large quantities in a direct competition against the CRT, they have not been cost competitive.

The only technologies showing current capability or near-term promise for large-screen projection displays are

the projection or light-valve CRT and the liquid crystal light valve. Large-screen direct viewing displays are likely to be limited to CRTs and EPIDs in the near future, with EPIDs having only monochrome capabilities.

Lastly, it should be noted that the display research and development community is extremely active, and is exploring various types of improvements in all these technolgies. Not only are the flat-panel technologies developing rapidly, but manufacturers of CRTs are finding new means to improve that product and keep ahead of the flat-panel proponents. Any review, such as this, will only be meaningfully valid for a short period of time (e.g., two-three years), due to the rapid technological growth in this area.

Section 4

HUMAN OPERATOR VISUAL CAPABILITIES AND REQUIREMENTS

The evaluation of any visual display, in terms of photo-metric, geometric, and information content, must be made relative to the visual requirements of the human observer. The display device with elegant electronic switching is of no use if the human observer is incapable of obtaining the needed information accurately and quickly. Nor is a display device which is ideally designed for one specific task nec-essarily ideal or even adequate for another, different task.

This important relationship between the human's visual system requirements and capabilities and the display's char-acteristics has been the subject of much research for many years. Indeed, it has motivated the research orientation of many vision scientists, and has similarly caused many dis-play designers to become very familiar with the human visual system.

Until recently, vision scientists and display designers have had significant communication difficulties, based largely upon the limited compatibility of measurements used by the two groups. The vision scientist, whose work dealt largely with the neurophysiology of the visual system, has produced little data of interest to display hardware engi-

neers. The vision scientist working with psychophysical techniques has often developed threshold measures of contrast and acuity which could not be directly translated to the legibility or readability of displayed information, for the displayed information rarely consisted of the discs, squares, and rectangles of the classical psychophysicist. Thus, there exists a very large body of valid, empirical, carefully controlled human visual performance data which is of limited or no use to the visual display designer or to the systems/human factors engineer looking for a quantitative base from which to select a display design or configuration. Fortunately, this schism between the two communities has been bridged, to some extent, by the recent application of a common language to both vision research and display design. The mathematical techniques of linear systems analysis have been used in electronics and optics for many years, but only recently have they become familiar and useful to vision researchers.

In brief, the concept of linear systems analysis permits one to determine the extent to which any component or system of components can transmit a signal. In the transmission process, some of the signal's amplitude is often lost, due to limitations of the transmission system, and this loss is measurable if the measurement is made under the proper circumstances. The linear systems analysis concept will be summarized in this section, and its application to visual

performance in the spatial, temporal, and chromatic domains will be described.

4.1   SPATIAL DISCRIMINATION

The most important characteristic of any display is that the elements of the display which are intended to convey different information be indeed discriminable. This, in turn, infers that each discriminable information element on the display be visually discriminable in the display space; hence, we speak of the importance of spatial discrimination of the observer as the guideline to setting display resolution and contrast requirements.

4.1.1   Classical Approach

As is generally known, the ability of the visual system to resolve (or detect, at a 50% threshold criterion) depends on many attributes of the stimulus or display. Among these are the size of the target, the luminance of the target, the luminance of the background, the shape of the target, the color (hue) of the target and background, and the adaptive state of the visual system. It is beyond the scope of this report to summarize or review this voluminous data set. The interested reader is referred to other secondary sources, such as Graham (1965) and Kling and Riggs (1971). By way of example only, representative data from Blackwell (1946), illustrated in Figure 55, indicate the classic generality:

- 204 -

smaller targets must be higher in contrast to be seen, and the greater the background (or adapting luminance), the smaller the threshold target size.
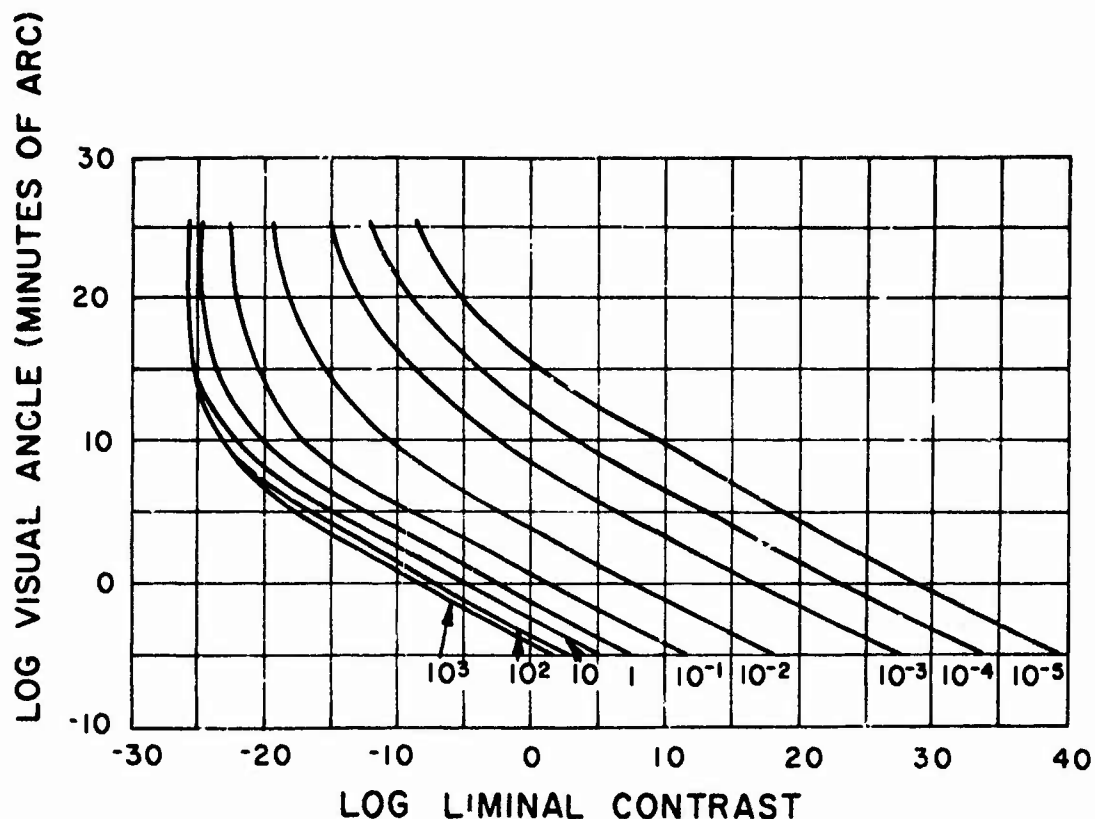
Figure 55: Resolvable target size as a function of background luminance and contrast, from Blackwell (1946)

However, these and similar data have a very limited utility when applied to the design and selection of visual displays. As a result, the concept of linear systems analysis to generate a contrast sensitivity function is developed in the next section for spatial discrimination. The contrast sensitivity function will then be applied to display design and evaluation in Section 5 of this report.

4.1.2 Contrast Sensitivity Function

The data of Blackwell (1946) point out that the human
observer can see small discs at varying luminance contrast
thresholds which depend upon the size of the disc and the
adapting luminance level.   Further,   other data could have
been cited which indicate that the threshold for small
square stimuli or rectangular stimuli also behave in a rea-
sonably similar manner, but lead to slightly different per-
formance levels due to the nature of the shape of such stim-
uli.   In fact,   one can define many variations of this type
of threshold experiment which will cause some slight changes
in threshold levels.   Thus, the person attempting to apply
these data to some real-world problem is faced with a sub-
stantial dilemma.   If he should choose to apply the classi-
cal psychological data to a particular design problem, he
must estimate some "field factor" which accounts for the
difference between the original laboratory data and the
application situation.   Such field factors have been esti-
mated to be on the order of 2.5 to 10.   If an appropriate
field factor is 5 for a given situation,   the threshold is
elevated by a factor of 5 to produce the predicted threshold
for the situation.   It is often difficult, if not impossi-
ble, to estimate the appropriate value of such a field fac-
tor, and persons attempting to use such adjustments find
that they can only do so by after-the-fact curve fitting
techniques, rather than by pre-experimental estimation.

On the other hand, one simply cannot perform the virtually infinite number of experiments which would be necessary to identify thresholds for objects of varying shapes, varying sizes, varying contrasts, and under the conditions of varying adapting luminance levels. Fortunately, in recent years the application of linear systems analysis has been applied to research in visual psychophysics, with the result that the very powerful techniques of Fourier analysis can be used to much abbreviate the requirement for such a multitude of experiments. Specifically, linear systems analysis states that any repetitive waveform can be analyzed into a number of component frequencies, with each of the component frequencies having a specific amplitude and phase relationship and that, if all the frequencies are appropriately combined with their respective amplitudes, the resulting summation is the original repetitive waveform, however complex it may have been. Figure 56 shows an example of the construction of a nearly square-wave luminance distribution by a series of sine waves of varying frequencies and amplitudes (Cornsweet, 1970).

The key to the usage of linear system analysis spatial discrimination, of course, is that the waveform must in fact be repetitive and non-varying over a reasonable length of time (the stationarity requirement). The analysis procedure by which a given repetitive waveform is broken down into the particular set of sine waves that must be added to produce
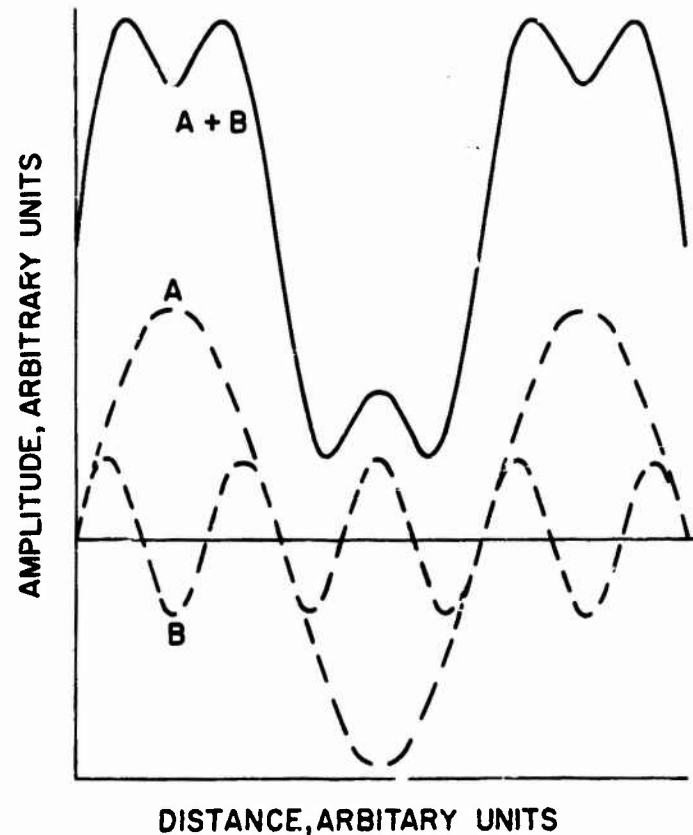
Figure 56: Synthesis of sinusoids varying in frequency and amplitude to form complex, repetitive waveforms, from Cornsweet (1970)

the waveform is called Fourier analysis, and the addition of sine wave components to reconstitute the original complex waveform is called Fourier synthesis. As can be seen from Figure 56, a true square wave would contain a fundamental sine wave (A) of a frequency equal to the fundamental of the square wave, along with additional sinusoidal components of higher frequencies (3F, 5F, 7F, ...), where F is the fundamental frequency, and each of the succeeding wavefo ns has a respective amplitude of one-third A, one-fifth A, one-seventh A, ..., where A is the amplitude of the fundamental

waveform. As more and more such higher frequency waveforms, all sinuosidal, are added together, the synthesis of these waveforms more closely approximates the square wave.

Another way of looking at this is to realize that a perfect square wave, because its rise and fall time are infinitely small, must be composed of some extremely (infinitely) high frequency waveforms to achieve the sharp rise and fall times of the square wave. Thus, a square wave pattern can only be realized if in fact the frequency response of the system producing it is extremely high, and a perfect square wave can only be realized if the system frequency response reaches in effect infinity, which is of course not possible.

Of considerable importance in the design of any display system is the fact that the high spatial frequency information must be preserved if the high frequency information is critical to the performance of the task by the observer. Thus, a Fourier analysis of the displayed information can be used to determine if the necessary high frequencies are present, and at what amplitudes they are represented. If their amplitudes exceed the observer's threshold, then the information is detectable and potentially useful, as will be seen later.

Typically, as the frequency of some input to a display system increases, the amplitude of the resulting image will tend to be reduced, much as the frequency response of a good

high fidelity audio system falls off as the frequency of the input signal is increased. Thus, one can plot the amplitude of the displayed information relative to the amplitude of the object plane information to determine the degree to which a given imaging system can transfer the spatial frequencies contained in the object plane to the image plane of the display. When the relationship is expressed in the form of modulation (see Section 3) and the transfer function is described in ratio form, then one has the basis for a technique known as modulation transfer function analysis.

The displayed modulation is the ratio between the difference, peak to peak, of some sinusoidal signal as displayed, over the sum of the maximum and the minimum of that signal. The modulation transfer factor, the ordinate of Figure 57, is the ratio of the modulation (M) out of the system to the modulation into the system. That is,

$$\text{Modulation Transfer Factor} = (M_{out})/(M_{in}). \tag{17}$$

When the display is incapable of "passing" a spatial frequency without attenuating it, $M_{out} < M_{in}$ and the modulation transfer factor is less than unity.

The abscissa of Figure 57 is plotted in spatial frequency units, such as cycles per unit distance, cycles per display width, or, in the case of the visual system, cycles per degree of arc (at the eye).
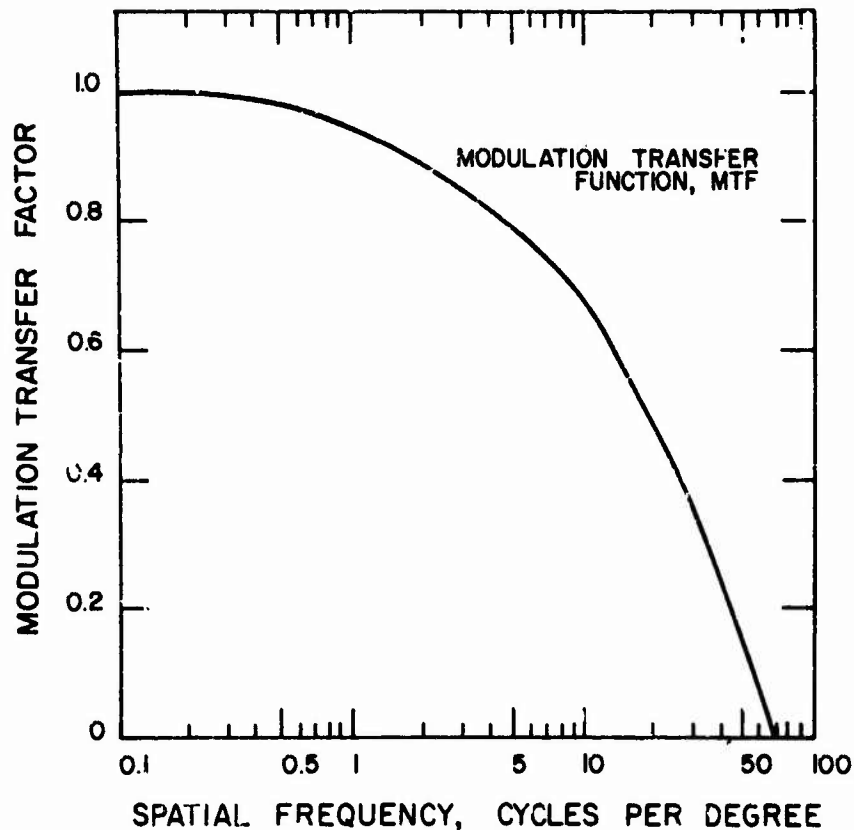
Figure 57: Modulation transfer function

Connecting the modulation transfer factor values for all spatial frequencies forms a continuous function, the modulation transfer function, or MTF (Figure 57). The MTF concept is widely used in communication systems analysis, optical systems, acoustic systems, etc. For a thorough discussion of the MTF in optical systems, see Gaskill (1979).

Thus far the MTF concept has been applied to displays, and not to the visual system. However, if we consider the visual system to have an input, a spatially varying sinusoidal grating, and an output, the perception of that sinusoid, then we can explore the notion of the MTF of the visual system in terms compatible with that of display devices.

To determine experimentally the MTF of the visual system, it is necessary to know (1) the input modulation, and (2) the output modulation. Two general experimental techniques can be used. In the first type of technique, the input (stimulus grating) modulation is held constant, and the output (subject's visual response) modulation is measured as the perceptual response strength, by a psychophysical scaling approach such as direct magnitude estimation. The second, more popular, technique holds the output modulation constant (at the observer's absolute threshold) and adjusts the input to produce a threshold response. The inherent assumption is that a threshold stimulus yields a constant "output" (i.e., threshold) response across all spatial frequencies. Thus, with this second technique, the MTF (or M_{out}/M_{in}) is inversely porportional to M_{in} because M_{out} is constant. In the first technique, M_{in} is usually held constant, and the MTF is directionally proportional to M_{out}, the magnitude of the observer's response.

The second, constant-threshold technique yields data of threshold modulation as a function of spatial frequency. A plot of these threshold modulations is commonly termed the "contrast sensitivity function" or CSF. A typical CSF, including 90% population estimates, is shown in Figure 58. The reciprocal of the CSF, if normalized to 1.0, is the MTF because MTF is inversely proportional to M_{in} (or CSF) for a "constant" M_{out}. (Normalization to 1.0 is necessary because the subject's "output" is dimensionless, though constant.)

The CSF has a minimum value in the region of 3-6 cycles/degree, with increasing modulation required at both higher and lower spatial frequencies (Figure 58).
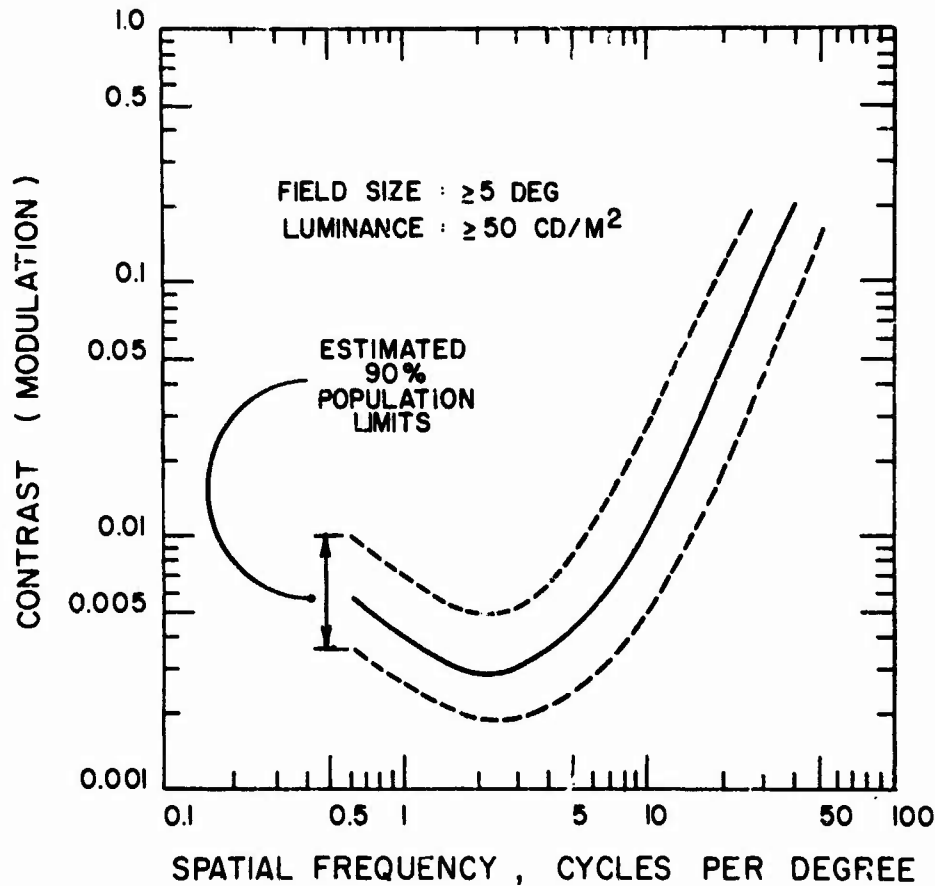
Figure 58: Visual contrast sensitivity function

In spite of the relatively large experimental technique differences used by various investigators, the fact that some used artificial pupils while others did not, the matching technique required of a subject, etc., there is a strong similarity among the results of the various empirical CSF studies. In general, the maximum sensitivity of the eye appears to be in the mid-range of spatial frequencies,

- 213 -

between approximately 3 and 6 cycles of the sine wave grating per degree of visual angle. As the spatial frequency of the sine wave increases, the eye requires greater modulation or contrast for a threshold response. Similarly, and perhaps surprisingly, as the spatial frequency of the sinusoidal grating decreases, the eye also requires greater modulation for a threshold response. Expressed in electrical analog terms, the eye appears to have a limited DC response, and a finite AC bandwidtn. This general relationship is shown in Figure 59, which combines normalized MTF data from several experiments using different experimental techniques and measurements, but all of which suggest the same general relationship. As will be shown in the following sections of this chapter, the peak sensitivity will shift with various stimulus and visual considerations, as will the relative sensitivity among spatial frequencies. In nearly all cases. however, it can be demonstrated that this contrast sensitivity function (or MTF) is decidedly nonmonotonic, and that the sensitivity decreases as one approaches very large gratings (low spatial frequencies) or very small gratings (high spatial frequencies).

The CSF, as illustrated in Figure 58, has become the bas's for the quantitative analysis of display quality, as will be discussed in Section 5. However, the CSF is altered by various stimulus conditions pertinent to display design and usage. In the following paragraphs, this systematic
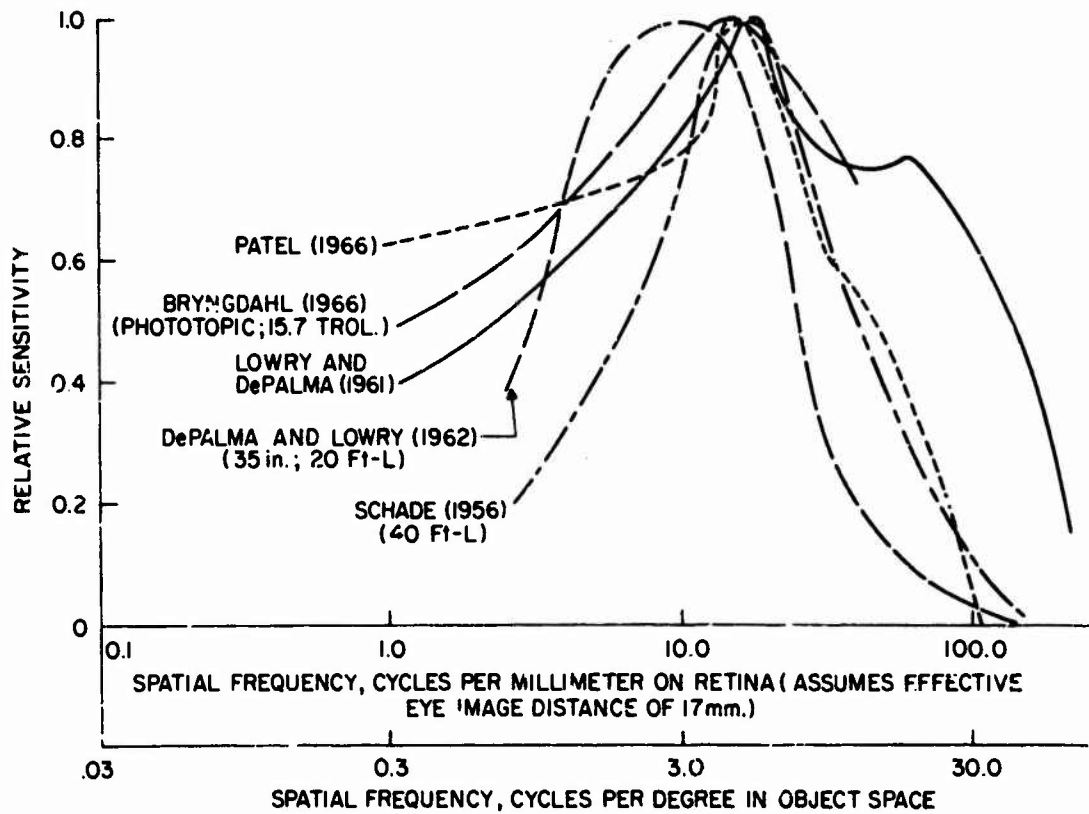
Figure 59: Representative visual system sensitivity functions

shifting of the CSF with various display conditions is described.

4.1.2.1   Effect of display luminance or retinal illuminance

Several experiments have investigated the effect on the human visual sensitivity function to sine wave stimuli when either the display luminance or the retinal illuminance is varied.

Both Patel (1966) and Van Nes and Bouman (1967) have demonstrated that modulation sensitivity is a nonmonotonic function of spatial frequency for relatively high levels of

retinal illuminance, but tends to become decreasingly nonmonotonic with lower retinal illuminance levels. As illustrated in Figure 60, the sensitivity increases with increases in spatial frequency up to approximately 3 cyc/deg if the retinal illuminance is on the order of 90 or 900 trolands, but that the peak sensitivity is at approximately 1 cyc/deg for retinal illuminance of only 10 trolands. If the retinal illuminance is reduced to approximately 1 troland, then the sensitivity decreases consistently from about 0.3 cyc/deg to a limiting resolution (at 100% modulation) of approximately 25 cyc/deg.
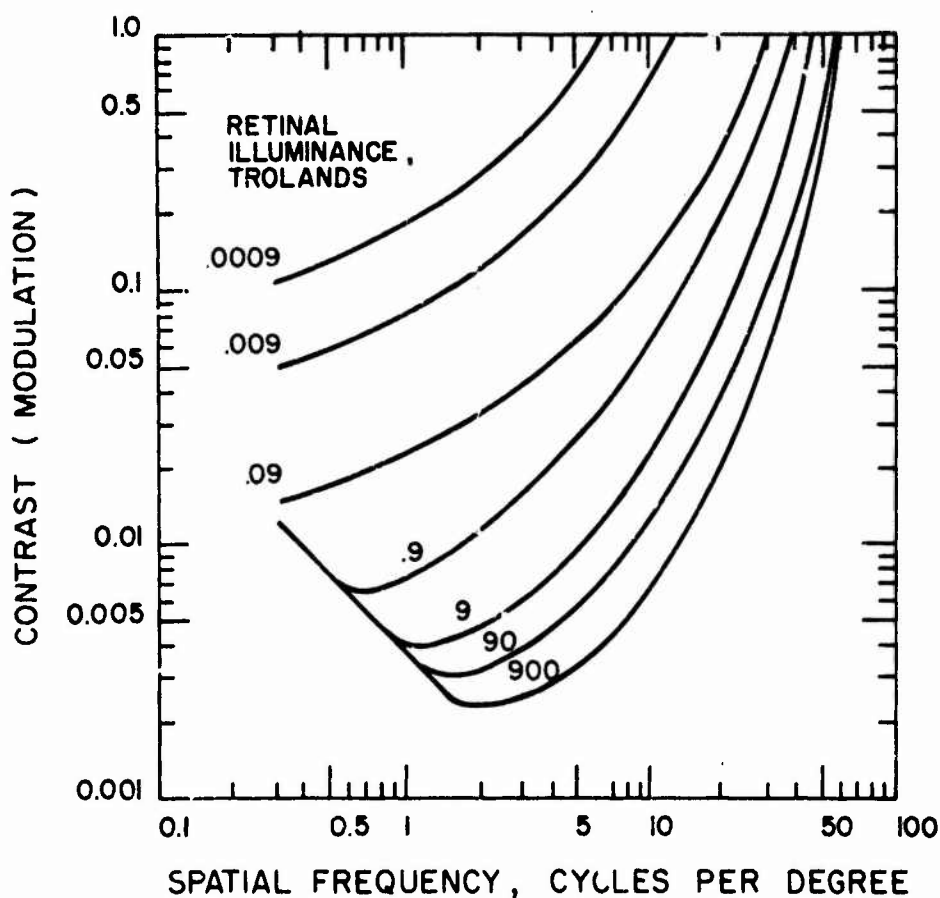
Figure 60: Effect of retinal illuminance on sensitivity

- 216 -

These results are not due entirely to retinal illuminance under artificial pupil conditions, but are also obtainable using direct viewing without an artificial pupil, as has been demonstrated by Van Meeteren, Vos, and Bongaard (1968). Their results, illustrated in Figure 61, show the slight differences for two subjects between vertical and horizontal grid orientations. The results of Van Meeteren et al. show the expected curve shapes, although the peak values are somewhat lower in spatial frequency than are those of Patel and of Van Nes and Bouman. For example, at 10 cd/m^2 (equal to about 31 trolands), the peak values for vertically oriented gratings in the Van Meeteren et al. paper occur at approximately 2 cyc/deg, whereas in the data of Patel, the interpolated value for 31 trolands would occur at approximately 15 cyc/mm on the retina, or 4.5 cyc/deg. Again, an interpolation from the data of Van Nes and Bouman, suggests that the peak sensitivity value for 31 trolands would occur at approximately 13 cyc/deg, or 43 cyc/mm on the retina.
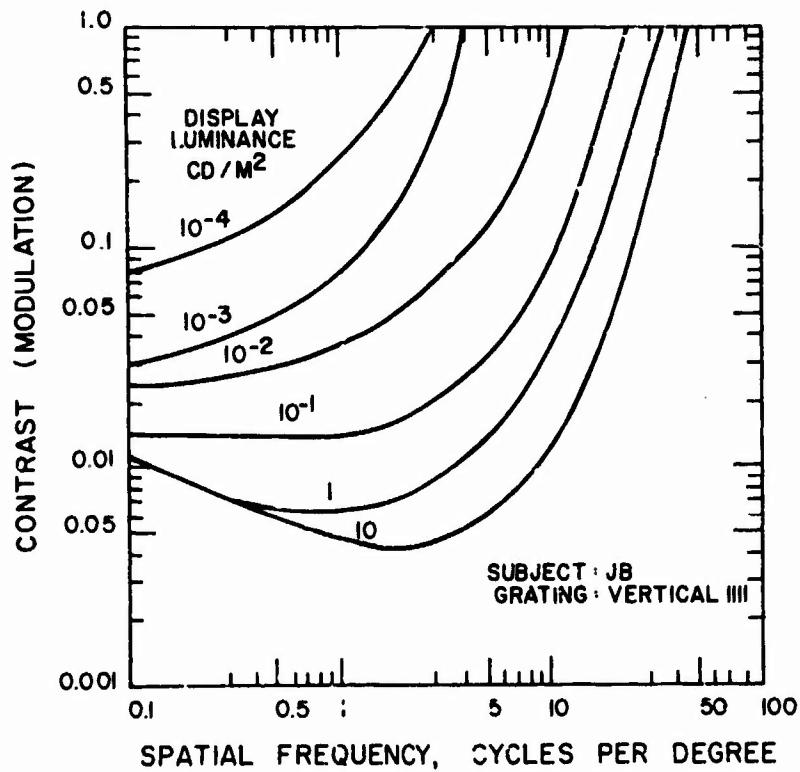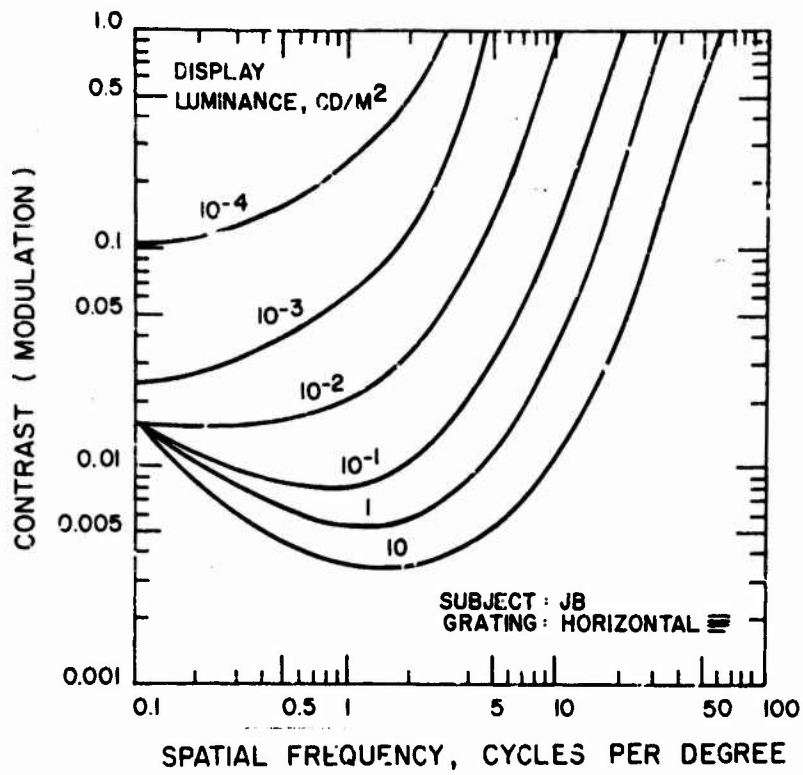
Figure 61: Contrast sensitivity for two subjects and two
orientations, from Van Meeteren et al. (1968)
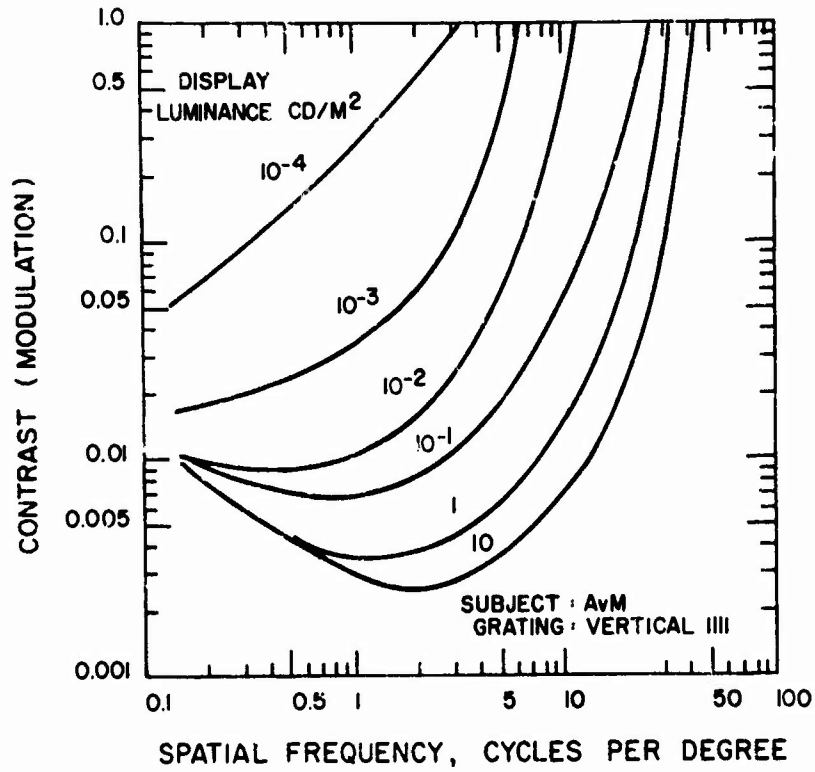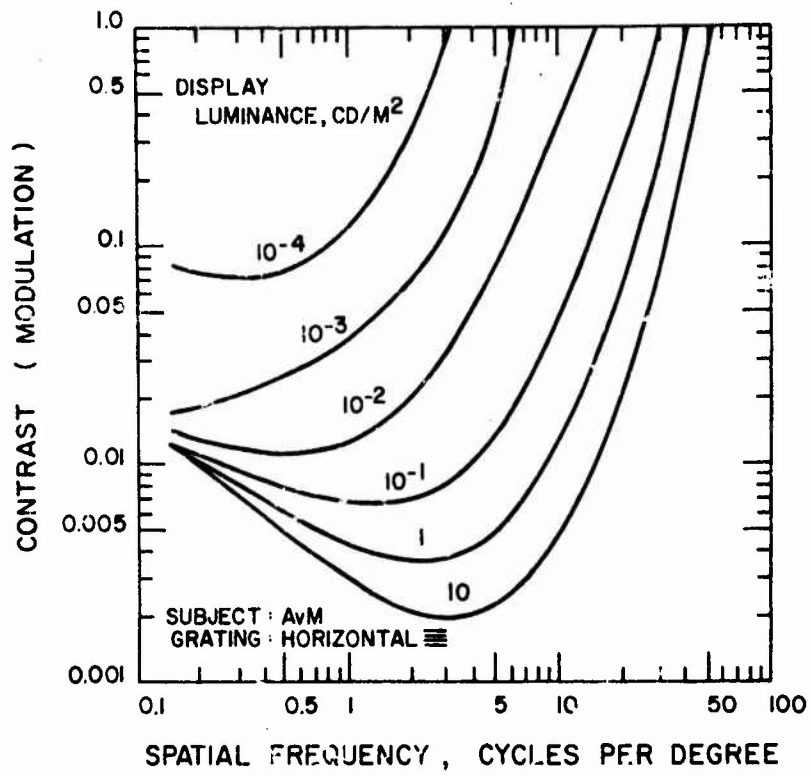
Figure 61, continued

- 219 -

Thus, these data are quite consistent, taken as a whole and considering the variety of techniques used to obtain them. In general, the peak sensitivity at high illuminance is on the order of 5 cyc/deg, and falls off in either direction. This peak is slightly lower in spatial frequency as either the retinal illuminance or the field luminance is decreased, with the function becoming monotonic in the neighborhood of approximately 0.1 to 3 trolands of retinal illuminance.

As we shall see later, this general relationship is extremely important in the design of visual displays for two reasons. First, the contrast required for a threshold response, and therefore for a given quality of image above threshold, is directly related to the display luminance or, equivalently, directly proportional to the retinal illuminance. Secondly, as the display luminance is decreased, the required size of the objects to be recognized on a given display must generally increase. This decrease is important not only in that the minimum size object which can be resolved must decrease in accordance with the limiting spatial frequency of Figures 59 through 61, but also in that reduced retinal illuminance conditions cannot take advantage of the peaking of the sensitivity function in the region of 3 to 6 cyc/deg, and therefore must depend even more upon larger displayed images for recognition performance and subjectively high image quality.

4.1.2.2 Effect of grating modulation on suprathreshold sensitivity curve

Among the several assumptions of linear systems analysis is the linearity of the system irrespective of the input level. This linearity assumption states that an increase in the overall (average) luminous intensity of the input sinusoidal pattern will result in a proportional increase in the output. Stated another way, in a linear system the output spectrum will never contain frequencies which are not present in the input spectrum, and all amplitudes across the pertinent frequencies will remain proportional from output to input independent of the change in input amplitudes.

In the case of the visual system, the linearity assumption applies to both the overall level of the input signal as well as to the modulation of that input signal. Bryngdahl (1966) specifically addressed the question of the linearity of the visual sensitivity function as he varied the input modulation of the stimulus. In his experiment, the subject was asked to match the overall luminance of a large comparison field to, alternately, the minimum and maximum luminance of the sine-wave grating. Figure 62 illustrates the subjective-to-object contrast ratio which the subject created by this response. Fundamentally, the subjective-to-object contrast ratio is the modulation of the sine-wave grating as determined by the luminance matching technique divided by the modula-

tion of the sine-wave grating as measured by a slit photometer. Thus, it is effectively the "gain" in modulation exhibited by the visual system. In Figure 62, this gain is plotted as a function of spatial frequency for four different object modulation levels. As can be seen, an increase in gain is obtained as the stimulus modulation is decreased. That is, as the input modulation becomes less, the visual system attempts to amplify this modulation, resulting in a higher subjective modulation than that which actually exists in the stimulus. For the assumption of linearity to be met, these curves should be identical, rather than varying in both slope and level. Thus, Bryngdahl's data show that the visual system does not meet the assumption of linearity at these particular suprathreshold levels.

4.1.2.3 Effect of retinal location

There is considerable evidence for the existence of specific spatial frequency detectors in various concentrations across the retina, with those detectors being most sensitive to high spatial frequencies located closer to the fovea, and those detectors most sensitive to lower spatial frequencies being most dense farther away from the fovea. If this is true, then one might expect a varying subjective contrast gain as a function of retinal location. Specifically, one might expect that the more dense the specific spatial fre-
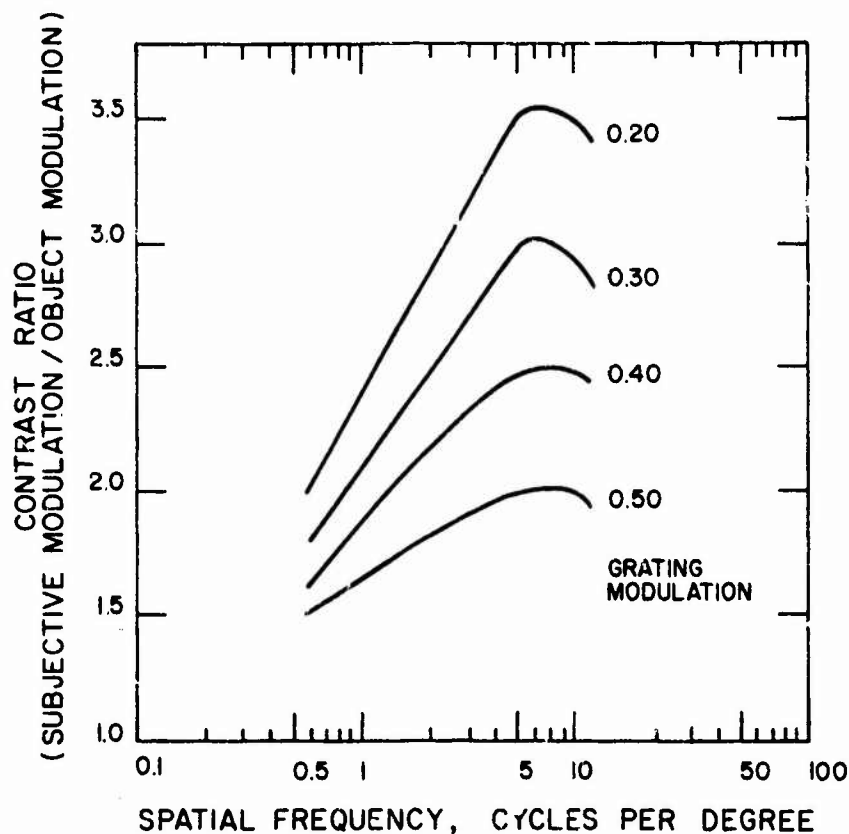
Figure 62: Evidence for nonlinearity of suprathreshold
sensitivity function, from Bryngdahl (1966)

quency detectors in a given retinal area, the more contrast
gain that would be obtained at that spatial frequency to
which the detectors are sensitive. The data obtained by
Bryngdahl (1966) have been recalculated and plotted in the
form of contrast gain as a function of grating spatial fre-
quency. Figure 63 illustrates that the greatest contrast
gain for high spatial frequencies is at the fovea, and the
contrast gain decreases in the fovea as the spatial fre-
quency of the grating decreases. Conversely, as one moves
farther away from the fovea, at least to 10 deg above the
fovea, the maximum gain occurs for lower spatial frequen-

- 223 -

cies. Thus, for a spatial frequency of about 1.2 cyc/deg, the maximum gain occurs about 10 deg above the fovea, while for a spatial frequency of 3 or 4 cyc/deg, the maximum gain occurs on the order of 5 deg above the fovea. The magnitude of contrast gain, as plotted on this figure, can be considered similar to sensitivity, so that maximum sensitivity at the fovea is for higher spatial frequencies, while maximum sensitivity outside the fovea applies to lower spatial frequencies as the distance from the fovea increases. This result is in congruence with the often obtained relationship which shows that visual acuity decreases as distance from the fovea increases.

4.1.2.4   Effect of grating orientation

It has long been known that the visual acuity for non-periodic objects, such as single and broken lines (vernier acuity), is greater for lines which are oriented vertically or horizontally than it is for lines having some other angular orientation. Using sinusoidal gratings, it is possible to orient the direction of the "bars" at any angle and to determine either the limiting spatial frequency (the spatial frequency at which the subject can no longer resolve the grating contrast) or the contrast sensitivity curve by having the subject vary the modulation for any given spatial frequency. The results are essentially the same with either technique. The CSFs for four angular orientations are indi-
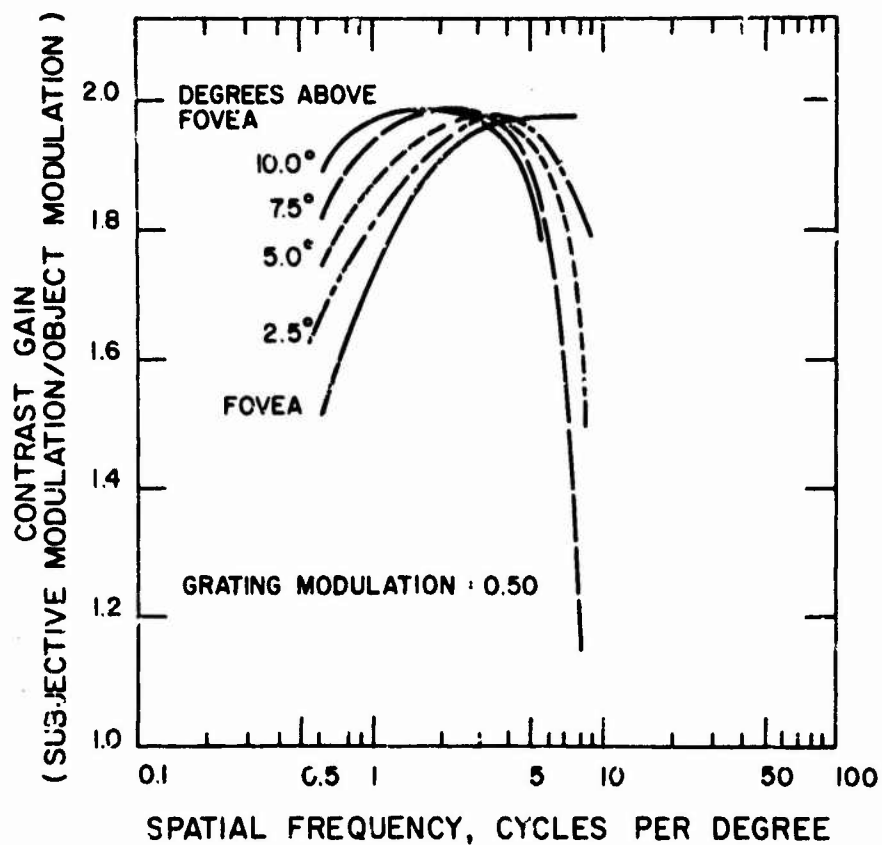
Figure 63: Contrast gain of the visual system as a function
of retinal location, from Bryngdahl (1966)

cated in Figure 64. The visual system is most sensitive at
0 and 90 deg, with a maximum reduction in sensitivity at the
45 and 135 deg points.

These results, of course, apply to the existence of a
sinusoidal grating in a non-structured visual field. When
such gratings are presented in the presence of field struc-
ture, such as on a television display which has a line
raster oriented typically in the horizontal direction, then
the threshold response can be expected to behave differ-
ently, and will obviously interact with the spatial fre-
quency of the raster background. For fundamental data on
the interaction between two sinuosidal patterns in an

- 225 -
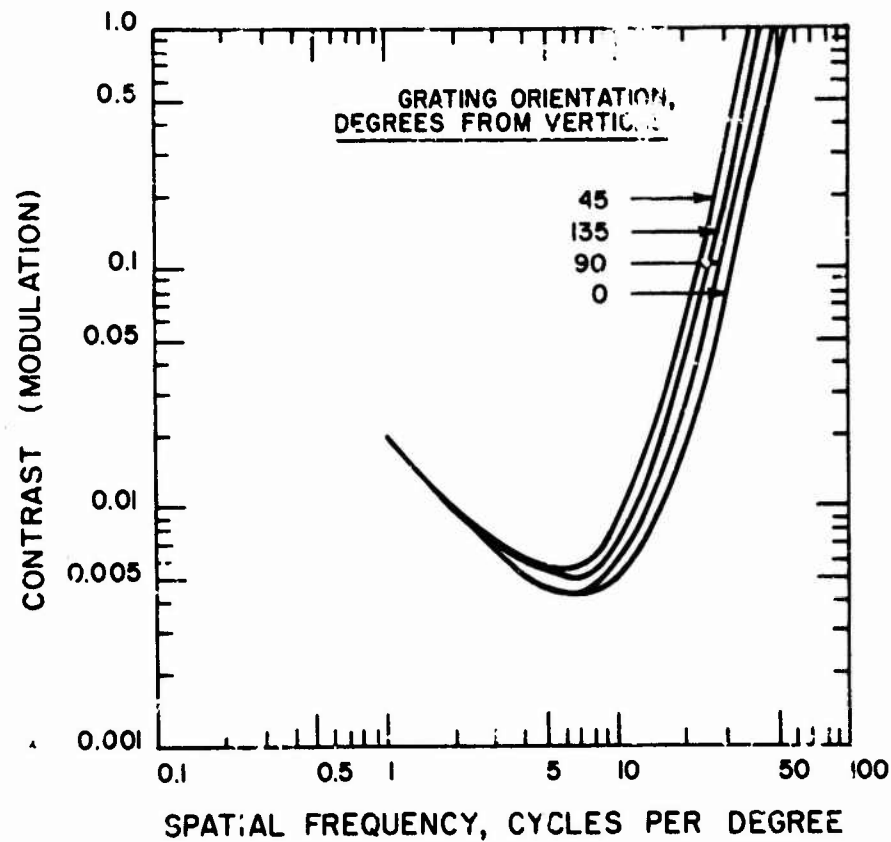
Figure 64: Changes in the resolvable spatial frequency for different orientations, from Campbell, Kulikowski, and Levinson (1966)

otherwise unstructured field, see Campbell and Kulikowski (1966), and for a more empirical demonstration of the masking and interference effect of a raster line grating upon the contrast sensitivity function, see Keesee (1976). This latter effect will be discussed in Section 5 as it affects the image quality of television displays.

4.1.2.5   Effect of display motion

It is well known that the visual acuity obtained for a static stimulus is much better than it is for a stimulus

that has motion either in a plane perpendicular to the line of regard or with radial motion toward or away from the observer (Snyder and Greening, 1966). Most of the existing data on this subject (see Ludvigh and Miller, 1958) have demonstrated that an exponential function nicely fits the data, and if the target motion contains both radial and axial motion, then a complex function can best fit the data.

More recently, techniques have been developed by which a sinusoidal grating can be moved perpendicular to the line of regard, with the subject having control of the modulation of the test pattern such that a threshold measurement can be made as a function of both spatial frequency and grating velocity of movement. Figure 65 illustrates this effect, as reported by Watanabe, Mori, Nagata, and Hiwatashi (1968). Drifting gratings generally elevate the CSF, but this elevation is not consistent with angular velocity. The inconsistency of these results across angular velocities points out the need for further research in this area. For the present, however, the CSF elevation shown in Figure 65 for 6.9 and 12.2 deg/s can probably be used as a design rule of thumb. Higher (19.2) and lower (2.2) drift rates in the Watanabe et al. (1965) study may have suffered from artifactual flicker due to the fixed TV frame rate in their apparatus. This decrease in sensitivity is fairly consistent for all spatial frequencies sampled, from 1.25 to 0.08 cycles/degree of arc. Of course, these spatial frequencies

are quite low, and the need is obvious for replication of this experiment at higher spatial frequencies, compatible with the information content of most displays.
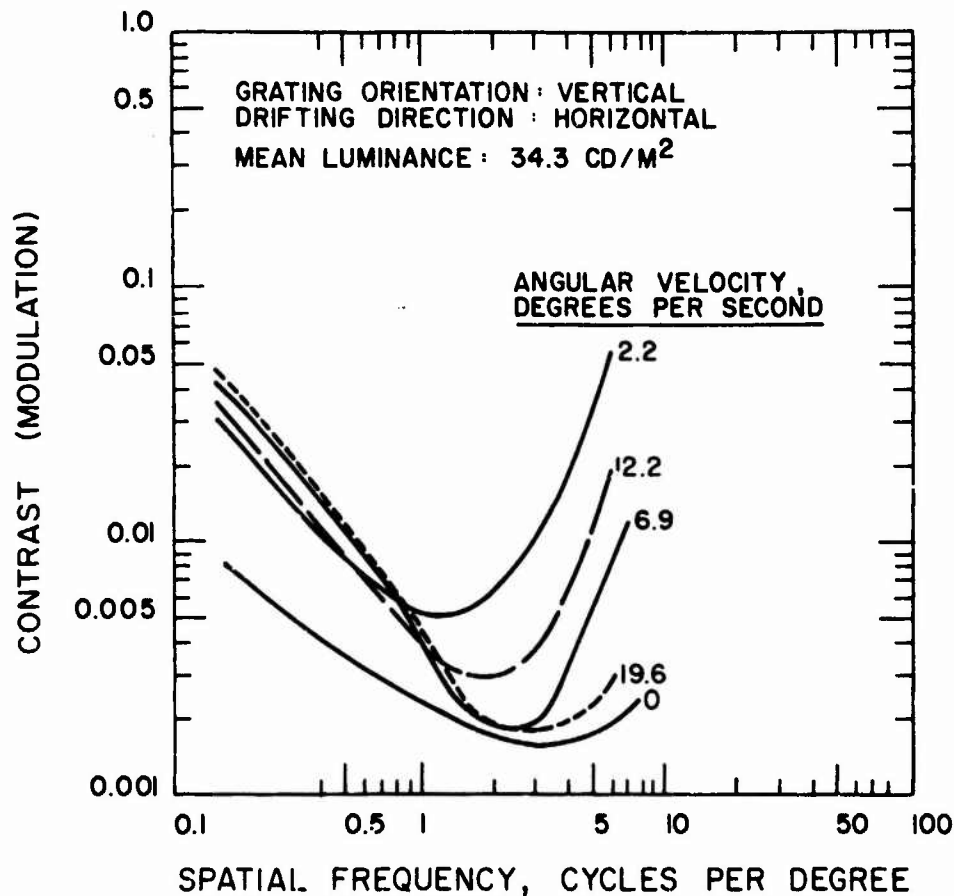
Figure 65: Effect of grating rate of motion upon contrast sensitivity function, from Watanabe et al. (1968)

4.1.2.6   Effect of viewing distance

As the target grating to be viewed by the subject changes distance from the observer, there is necessarily a change in the focal length of the lens, which in turn alters the

extent of various spherical and astigmatic aberrations. In general, as the lens is forced to focus at a shorter distance, the MTF of the lens is reduced. This effect can be seen most clearly in Figure 66, which shows that the contrast sensitivity for any given spatial frequency grating increases as viewing distance increases. A plot of limiting resolution as a function of viewing distance, based upon the work of De Palma and Lowry (1962), illustrates essentially the same effect (Figure 57). Of practical importance is the fact that the "knee" of this curve occurs at approximately 46 to 50 cm, but of equal importance in some situations is the fact that improvements in limiting resolution continue to occur beyond 305 cm. This general result has been found by both De Palma and Lowry (1962) and Watanabe et al. (1968).

4.1.2.7 Comparison of sine-wave and other waveform thresholds

For a variety of reasons, investigators have chosen to use either sine-wave targets or, in some cases, square-wave targets. The convenience of generation of a square-wave target is fairly obvious. One can cut opaque materials into strips and locate them on a transparent base to produce square waves; existing patterns can be purchased directly in many sizes and densities; and relatively simple laboratory equipment necessary to generate
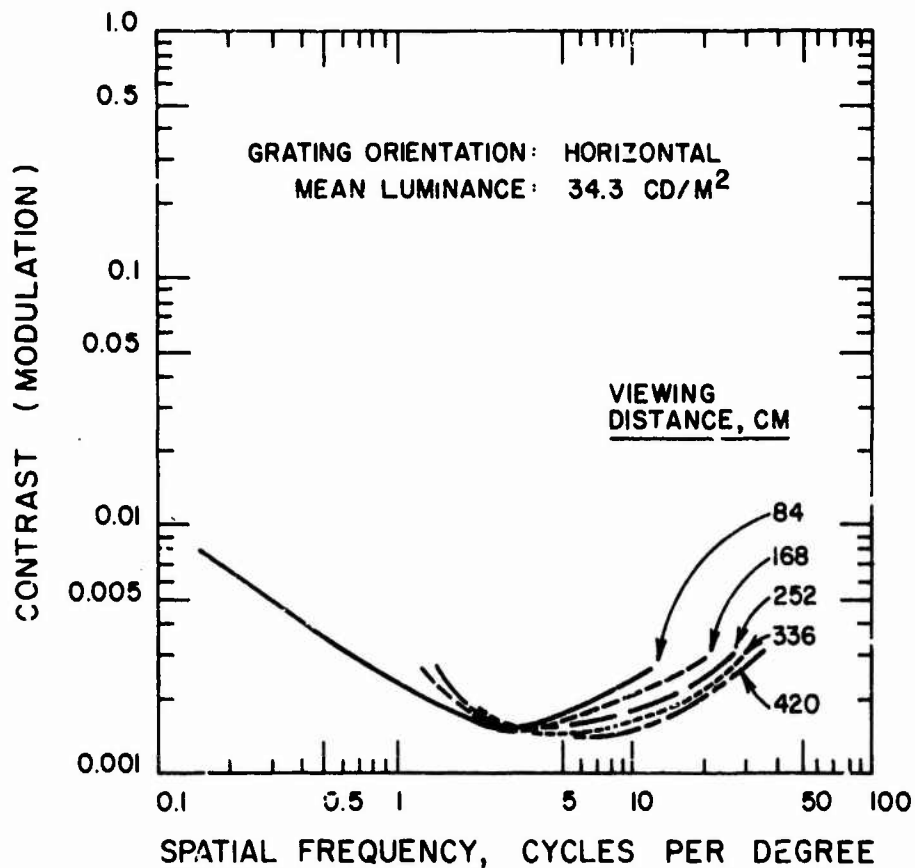
Figure 66: Effect of viewing distance on contrast
sensitivity function, from Watanabe et al.
(1968)

such either photographically or optically is both inex-
pensive and easily understood. By comparison, sine wave
gratings are extremely difficult to generate and, even in
the best cases, often result in a few percent harmonic
distortion. Thus, although the ease of fabrication of
the test pattern favors the use of square-wave gratings,
the linear systems analysis requires the use of sinusoi-
dal gratings. It is perhaps easiest to generate a sinu-
soidal grating on a cathode ray oscilloscope, athough
irregularities in the phosphor, as well as non-lineari-
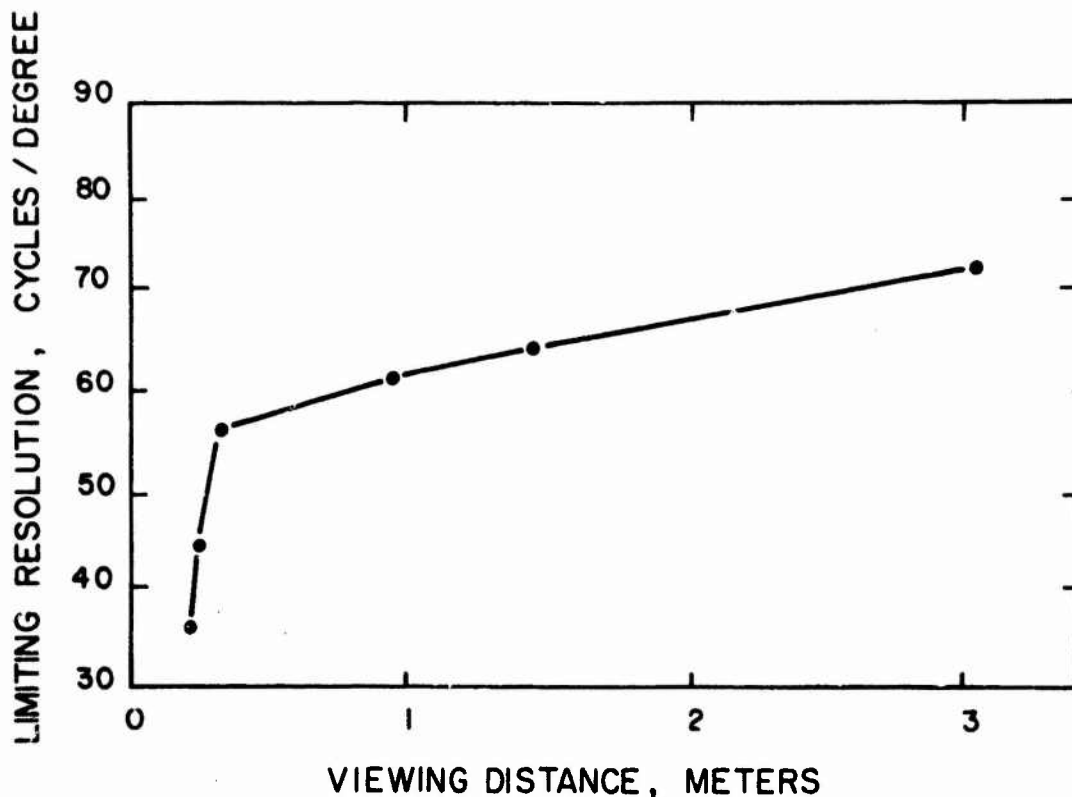
Figure 67: Effect of viewing distance on limiting spatial frequency of sine-wave grating, from De Palma and Lowry (1962)

ties of the luminance versus voltage transfer function may cause slight distortions in the sine-wave patterns. It is thus pertinent to inquire whether the more easily fabricated square-wave gratings can be used in place of the sine-wave gratings for experimental purposes, and subsequently to modify the data in some predictable fashion in order to make use of the analysis techniques based upon sine-wave gratings.

In 1962, De Palma and Lowry obtained threshold contrast functions for both sine-wave and square-wave grat-

ings ranging from 1 to approximately 30 cycles/deg. Their results indicated that the threshold sensitivity for square-wave gratings is greater than the threshold sensitivity for sine-wave gratings for spatial frequencies up to approximately 6 cyc/deg. Above 6 cyc/deg, the square-wave threshold is slightly lower than that for sine waves, although this difference may not be statistically significant. No statistical data were presented. A very similar result by Campbell and Robson (1968) obtained the predicted difference at low spatial frequencies, but effectively no differences at spatial frequencies above 7 cyc/deg. The results of Campbell and Robson are presented in Figure 68.

To understand the difference between sine- and square-wave gratings, concepts of linear systems analysis can be used. Fourier theory shows that a square wave can be synthesized by combining a number of sine wave components, the frequencies of which are odd multiples of the fundamental frequency of the square wave. Thus, the odd multiples have amplitudes of one-third, one-fifth, one-seventh, etc., for frequencies of 3F, 5F, and 7F, where F is the fundamental frequency. All even harmonics have zero amplitude. Because the contrast sensitivity function for the sine-wave grating falls off rapidly at spatial frequencies above approximately 3 cyc/deg, it might be therefore expected that a square-wave grating which

- 232 -
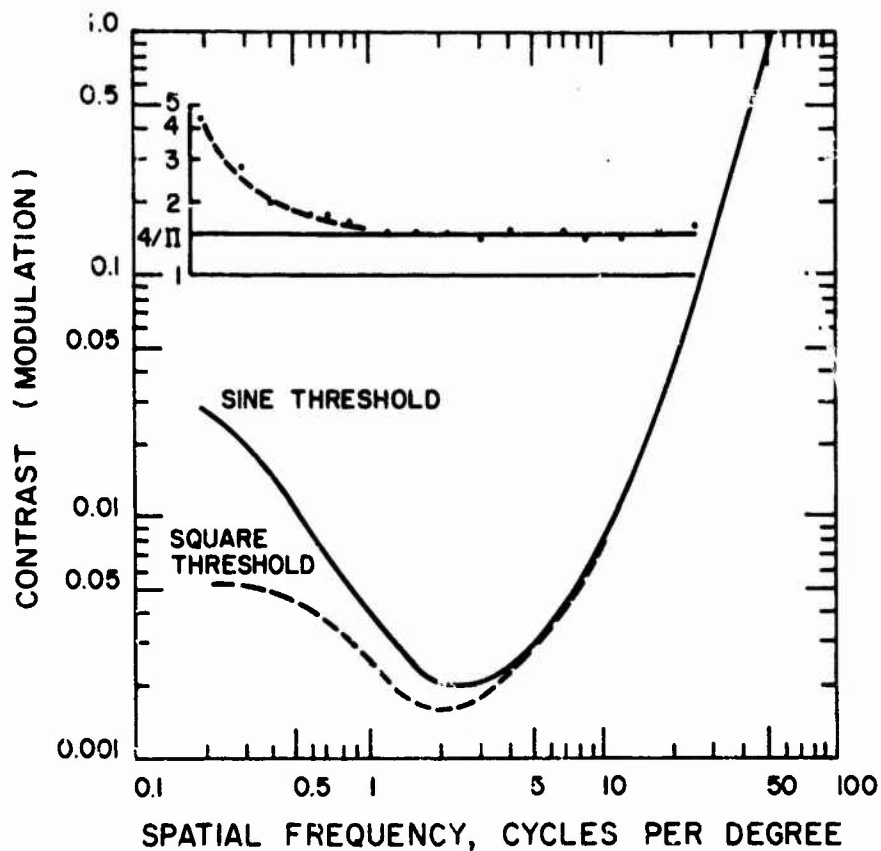
Figure 68: Comparison of sine- and square-wave
sensitivity function, from Campbell and
Robson (1968)

has a fundamental frequency in excess of 3 cyc/deg would
be nearly indistinguishable from a sine-wave grating;
that is, its higher order harmonics would contribute very
little to its visibility. As the spatial frequency
increases above 3 cyc/deg, the harmonics of the square
wave would contribute so little to its visibility that
the square-wave and sine-wave thresholds should be
approximately the same.

Below 3 cyc/deg, there is enough amplitude in higher
harmonics of the square-wave fundamental so that these

- 233 -

harmonics might contribute substantially to the threshold response to the square wave. To determine this contribution of the harmonics, one must examine the amplitudes of the harmonic components. The amplitude of the fundamental or first harmonic of a square-wave grating having contrast modulation M is $4M/\pi$, and the amplitudes of the third, fifth, and subsequently higher harmonics are $4M/3\pi$, $4M/5\pi$, etc. Note that the amplitude of the first harmonic (fundamental) component is actually greater than the contrast M of the square wave, by a factor of $4/\pi$ or 1.273. Thus, if a square wave and a sine wave of the same fundamental frequency have the same peak-to-peak amplitude, the amplitude of the first harmonic of the square wave is actually greater than is that of the sine wave. Accordingly, one would expect to find greater sensitivity for the square wave than for the sine wave for those spatial frequencies to the point where the eye becomes decreasingly sensitive, or about 3 cyc/deg. Figure 68 illustrates this concept at the top of the figure, where a horizontal line is drawn intersecting a ratio equal to $4/\pi$, or 1.273. The plotted points along this line represent the ratio of square-wave to sine-wave sensitivity at the various spatial frequencies (Campbell and Robson, 1968), and the dashed line is a best-fit line drawn by eye through these data points. This ratio does not deviate substantially from 1.273 until the spatial frequency of the pattern is less than about 0.8 cyc/deg.

At 0.8 cyc/deg, the third harmonic is 2.4 cyc/deg which, although having an amplitude of only one-third the fundamental, still has a frequency below the frequency (3 cyc/deg) at which the visual system sensitivity begins to decrease for a pure sinusoidal waveform.

In summary, the eye appears to behave as a Fourier analyzer of square waves, with the contrast sensitivity significantly affected by the higher order components as long as those higher order components are of a spatial frequency to which the eye is reasonably sensitive. Because the sensitivity begins to decline rapidly beyond about 3 cyc/deg, the contribution of higher order components is negligible for fundamental frequencies above approximately one-third of this modal value of the sensitivity curve, or for fundamental frequencies above approximately 1 cyc/deg. Stated another way, the sensitivity to the square-wave grating is greater than is that to the sine-wave grating if the third harmonic of the square-wave grating is sufficient in amplitude to exceed the sinusoidal threshold at the frequency of the third harmonic. Thus, if the third harmonic of the square wave is above the sinusoidal threshold at that spatial frequency, then the square wave is perceived as a square wave, rather than as a sine wave. Otherwise, the observer cannot distinguish between square waves and sine waves.

Campbell and Robson (1968) also provided data for the relative visibility of other waveforms, including rectangular and saw-tooth gratings, with results that further support the concept of sensitivity being determined by thresholds levels of higher order harmonics. These results, which have subsequently been replicated by other researchers, form the very important basis for comparison of Fourier analysis of displayed information with the ability of the visual system to see that information.

4.1.2.8 Stimulus duration effects

It is common to find that tachistoscopically presented stimuli are seen either more easily or more correctly as the duration of the presentation time increases. Because the eye integrates information over time, as has been demonstrated repeatedly since Blondel and Rey (1911), then there is every reason to expect that some temporal integration would occur for sinusoidal or square-wave gratings as well as for other stimuli. This assumption has been tested several times, and the results are consistent with other data and as expected. Figure 69 illustrates the results obtained by Schober and Hilz (1965), who used square-wave gratings varying in spatial frequency from approximately 0.24 cyc/deg to approximately 18.0 cyc/deg. Their stimulus field had a luminance of 110 cd/m^2, and was viewed from a distance of 1 m. The nonmonotonic reversal of the CSF

occurs for all exposure times from 1000 to 1.5 ms, although the nonmonotonicity is certainly more pronounced for greater exposure time.
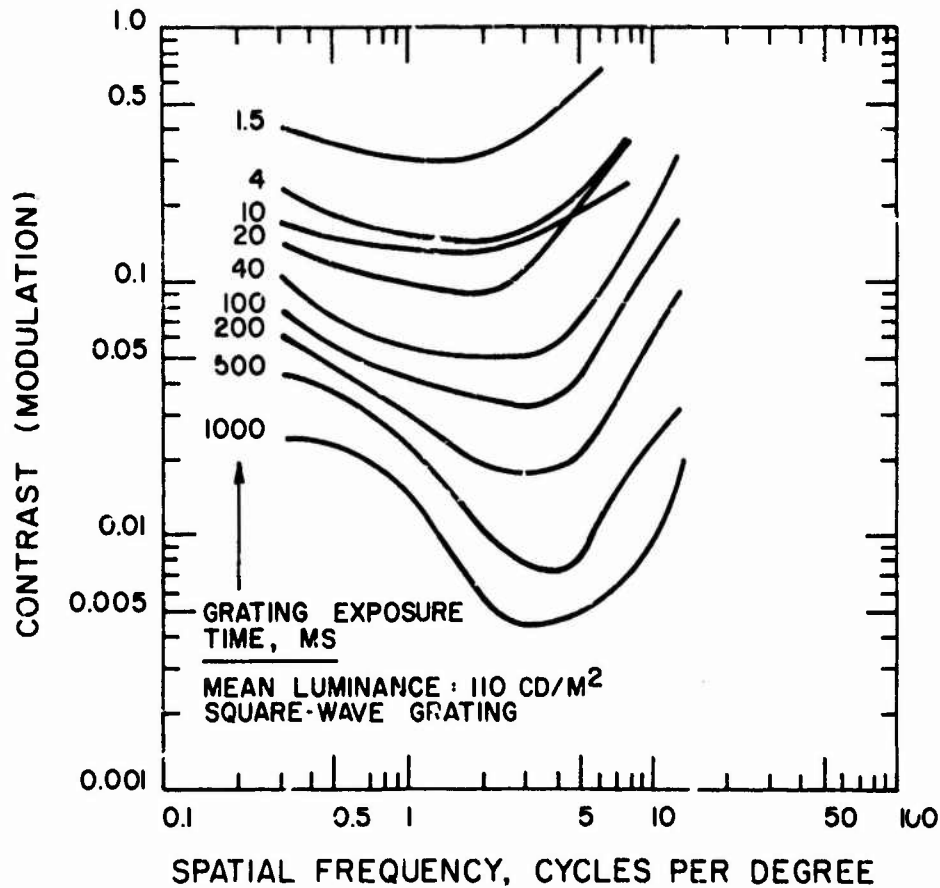
Figure 69: Effect of grating exposure time on contrast sensitivity function, adapted from Schober and Hilz (1965)

Thus, the data related to limited stimulus exposure time achieve essentially the same type of contrast sensitivity function as those for which a stimulus is exposed as long as the subject is required to make a response. The peak sensitivity occurs somewhere between approximately 1.8 and 3 cyc/deg. As the exposure time is decreased to fractions of

a second, the sensitivity curve becomes more flat and greater contrast or modulation is required to achieve a threshold response. There is also a suggestion that the modal value (peak) of the curve may shift slightly toward a lower spatial frequency, but this effect is certainly not very pronounced.

4.1.2.9   Effect of wavelength on luminance CSF

Stiles (1949) found that the eye had a slightly lower acuity to blue light than to green and red light. To test this result using sinusoidal gratings, Watanabe et. al. (1968) used monochromatic television picture tubes having red, green, and blue phosphors. Luminous contrast sensitivity was checked at both liminal and supraliminal contrast levels. Their results indicate no substantial difference among the three colors, except for perhaps a slight reduction of sensitivity to blue light. Thus, their data generally support the slight reduction in sensitivity obtained by Stiles, but suggest that luminance contrast sensitivity appears to be, for all practical purposes, independent of the wavelength, provided that the luminances are matched for the appropriate photopic viewing conditions. This obviously has importance for the design of visual displays if one assumes that such displays might contain various colored elements and one is concerned with the relative sensitivity to detail (spatial frequencies) for elements of various colors.

4.1.2.10   Summary of spatial CSF

The preceding paragraphs have described some of the basic research on the spatial CSF, perhaps in more detail than needed for present purposes. This background discussion was included so that the reader could develop an understanding of the CSF concept and how the CSF varies with situation- and display-specific variables.

In making use of the CSF in display design and evaluation, the system designer must set display parameters to "match" the CSF of the observer for the given viewing conditions. The way in which this "match" is obtained is discussed in Section 5. At this point it is pertinent to point out that the basic CSF (Figure 58) was shown to vary systematically with the following situation and display conditions:

1. display luminance, Figure 61,

2. grating input modulation, Figure 62,

3. retinal (off-axis) location, Figure 63,

4. grating orientation, Figure 64,

5. grating angular velocity, Figure 65,

6. viewing distance, Figure 66, and

7. grating exposure time, Figure 69.

Any use made of the CSF should therefore use the CSF variant which most closely represents the viewing conditions of concern.

4.1.2.11 Linearity: is it important?

In much of the above information of this section, it was pointed out that the human visual system does not contain the properties required for accurate analysis of pure linear systems, and that perhaps the concept of the modulation transfer function or CSF is inappropriate for a description of performance by the visual system.

It is not surprising, on the other hand, that a linear systems model does not fit the properties of the visual system. Rarely does a linear model ever describe biologically dynamic systems with great accuracy; in fact, rarely does a linear model ever describe inanimate systems with great accuracy over their entire operating range. Nonetheless, linear systems analysis has been and will continue to be applied to describe both animate and inanimate systems in spite of the lack of appropriateness of the assumptions. That is, linear systems models can be very useful approximations and the advantages obtained by using such may well outweigh the inaccuracies produced. As we shall see in Section 5, the use of this linear model for the human visual system has the distinct advantage of placing in a systems model context the relative sensitivity of the visual system

to various display parameters and permitting the calculation of an overall system response (including both the display and the visual system) in the same set of equations and terminology.

Other portions of this report demonstrate that use of the linear systems model, although certainly inaccurate and only approximate at times, provides a practically useful description of total system performance that would otherwise be unattainable were one not to use such a linear model. Until more sophisticated nonlinear systems models are available to describe the spatial characteristics of the visual system in its various operating ranges and modes, the use of a linear model is probably better than any other approximation. In fact, because the linear systems model more closely approaches validity as we restrict the operating range of the visual system, it will be seen that in many cases the fit of the linear model is not very bad at all. The reason for this is that in most cases of display design, the design problems occur at near threshold operating ranges, or certainly over restricted operating ranges. When the permissible and usable operating ranges are relatively large, there is typically no problem in the display design. That is, the adaptability of the eye and the flexibility of the display design more than account for any great need for accurate prediction of performance.

On the other hand, when it is necessary to minimize the display size, minimize its power, and therefore reduce the number and size of individual characters and their luminance levels, then the criticality of each design parameter becomes apparent. However, at these critical levels, the operating range of the eye is similarly restricted and defined, such that the linear model has greater accuracy and applicability.

4.2 TEMPORAL DISCRIMINATION

When the display designer, faced with the task of designing a compatible and efficient dynamic display, begins to allocate his input bandwidth to the system, he obviously must divide the available bandwidth into both spatial and temporal contributions. That is, if he chooses to use a large number of x,y specific resolution components, he is then clearly limited, assuming some finite bandwidth, by the maximum update or refresh rate of the display. Conversely, should he decide that he must refresh the display quite often, then it is equally obvious that the bandwidth (number of data points per unit time) will be allocated in greater proportion to the refresh rate or temporal considerations than to the spatial considerations. Figure 70 illustrates this general tradeoff, and points out the importance of the relationship for a generalized display. For example, if it is necessary that the eye does not experience a flickering

display to comfortably and efficiently resolve each data point, and if the minimum refresh rate for the given display is at 50 interruptions per second, then it is clear that the display must have a bandwidth of 5 MHz to contain 100 K elements. This is a fundamental tradeoff curve which the designer of any dynamic display uses in the very early stages of the design cycle.
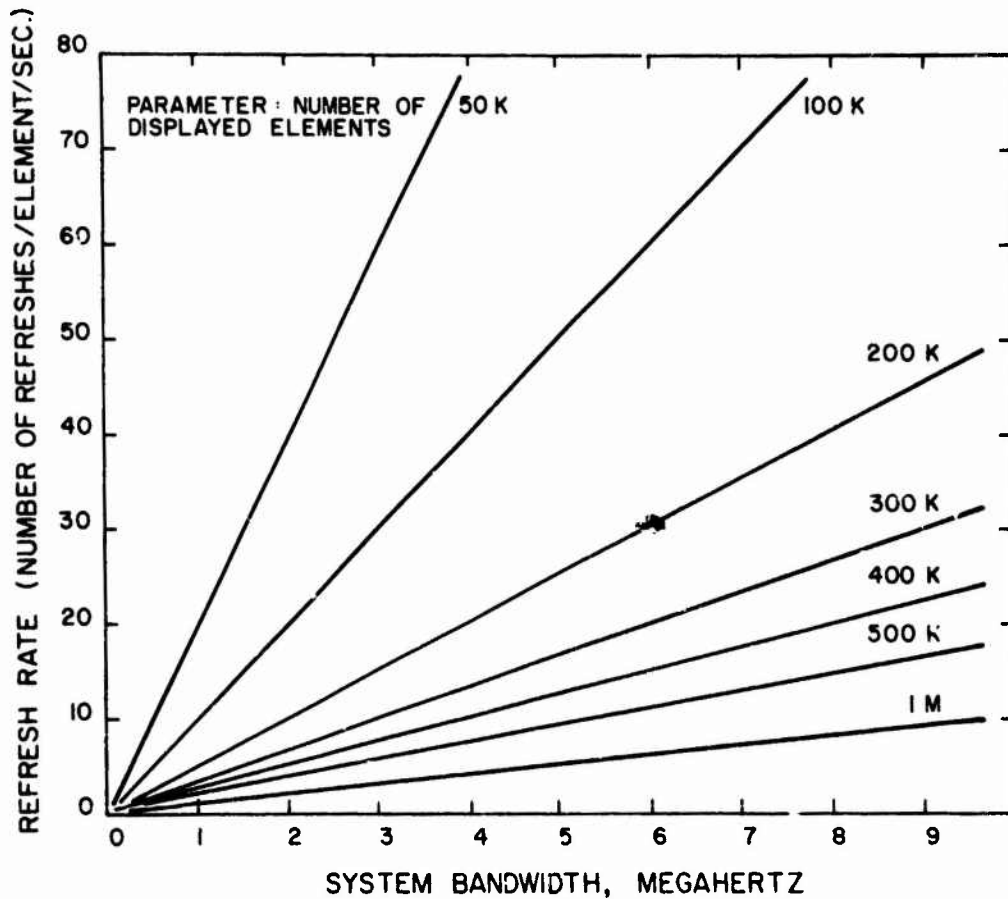
Figure 70: Relationship among available bandwidth, display refresh rate, and number of displayed elements

By comparison, should the display not have elements that require continuous refreshing in order to avoid flicker, but

rather elements that remain "on" until turned "off," such as some of the more recent matrix displays and gas discharge panels, then the designer no longer must contend with the flicker problem, but can concentrate more on the spatial relationships and contrast requirements.

It is likely the case, however, that dynamic displays will always retain a place in most display/control systems, simply because many systems require a continual update of changing information in order that the operator may take appropriate action on a timely basis. As long as this information change rate is reasonably high, there will continue to be displays in which the luminance information decays rapidly if the display is not refreshed periodically.

4.2.1   Critical Fusion Frequency Approach

There is a large volume of literature dealing with research on the sensitivity of the human visual system to interrupted pulses of light. Much of this research takes the form of a stimulus which is sequentially turned on and off in rapid succession, with the observer typically asked to determine the minimum on-off rate at which he or she no longer experiences a flickering sensation. This on-off rate, at which flicker is no longer viewed, is conventionally termed the critical fusion frequency (CFF) or the flicker fusion frequency (FFF). In all experimental situations, a threshold measure is defined as the frequency at which flicker (or

fusion) is just barely eliminated. The experimental situation may permit the observer to vary either the frequency of the displayed element or the difference between the "on" and "off" portions of the cycle. In either case, one can plot the threshold CFF versus other parameters of the display. For an excellent review of the great volume of literature in this area, the reader is referred to Brown (1965). Unfortunately, this extensive literature has its inherent ambiguities and disagreements. Although the existing literature nicely points out the effects upon the CFF of variables such as stimulus size, retinal location, spectral distribution of the stimulus, etc., the data are generally complex and occasionally result in gross disagreement. For all practical purposes, these data are not necessarily of the most importance to the functional display designer. As in the case of spatial capabilities of the visual system, the temporal capabilities of the visual system are also amenable to analysis and specification by the contrast sensitivity function approach.

4.2.2   Temporal Sensitivity Function Approach

One can take a train of on-off information and analyze it in the temporal frequency domain much as one analyzes spatial information in the spatial frequency domain. As the content of a complex but periodic sound wave can be analyzed using Fourier analysis, so can the content of a complex but peri-

odic train of light energy in the same temporal frequency domain. In recent years, it has become popular to conduct experiments such that Fourier analysis can be applied to temporally varying stimuli in the visual spectrum, and the temporal analysis capability of the visual system has been likened to a Fourier analyzer. As we shall see, this approximation is reasonably appropriate, and a fair understanding of the temporal sensitivity and discrimination capability of the eye can be obtained in this manner. Further, this temporal CSF is very useful in display design and evaluation.

In 1958, de Lange conducted an experiment to compare the effect of various temporally modulated waveforms upon the critical flicker threshold. His results, shown in Figure 71, indicate the sensitivity of the visual system to the retinal illuminance level and simultaneously its insensitivity to the nature of the waveform of light. At a reasonably low modulation amplitude (difference between light and dark portions of the light energy), such as at 5 percent modulation amplitude, the CFF for a low retinal illuminance level of 4.3 trolands is approximately 11 Hz, increasing to approximately 27 Hz for a retinal illuminance level of 430 trolands. As the modulation amplitude increases to 100 percent, the CFF similarly increases, requiring an interruption rate of approximately 47 Hz for a retinal illuminance level of 430 trolands.
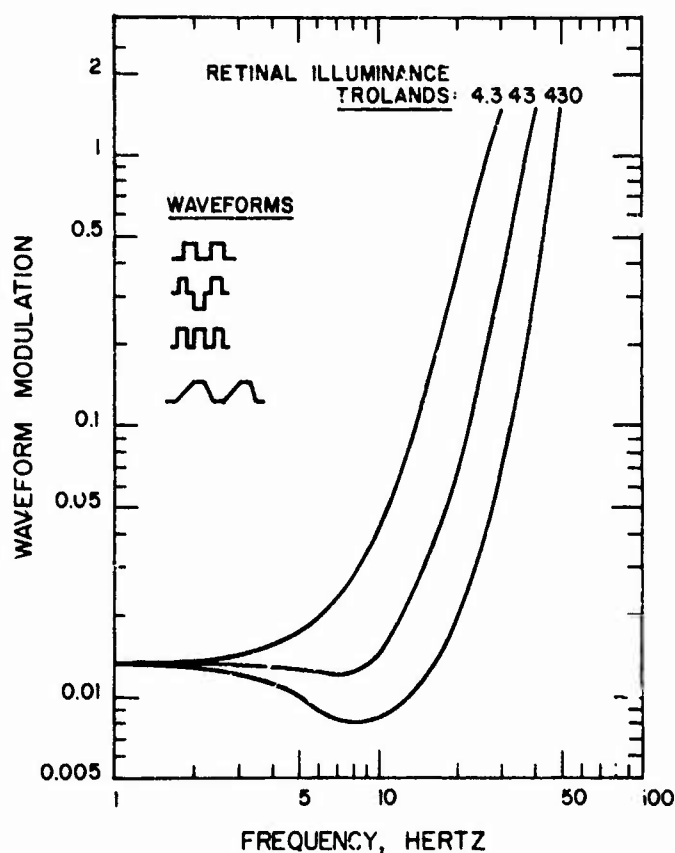
Figure 71: Temporal contrast sensitivity function, for fundamental compo..ent of four waveforms, from de Lange (1958)

Of key importance in the results of de Lange is the fact that the CFF curve for each of the three retinal illuminance levels is quite independent of the shape of the waveform. Thus, the various waveforms indicated in Figure 71 apparently had no effect upon the CFF at a given retinal illuminance and modulation level. One might ask the reason for this, since we have already seen that there is a difference in spatial sensitivity of the visual system between square-wave and sine-wave sensitivity. That is, if one were to perform the necessary Fourier analysis of the waveforms

- 247 -

illustrated in Figure 71, would not the eye be predicted to show differential sensitivity to those different wave forms? In 1959, Kelly questioned de Lange's data and concluded that de Lange's results could be attributed to the fact that he used a constant luminance surround at the edge of his central flickering stimulus, thereby forming a sharp edge bordering the flickering stimulus against the large surround. Because the contrast conditions at the edge of the flickering stimulus formed a reasonably high, but alternating contrast, flicker under these conditions could be presumed to relate to not only the temporal variation in the central stimulus field but also to the spatial contribution of the edge of the field. Kelly tested this hypothesis by using a circular test field of uniform luminance in a large adaptation field, but tapered off the edge of his central field. Using this blurred disc, Kelly obtained thresholds as a function of retinal illuminance ranging from 0.06 to 9300 trolands, and found that his thresholds were considerably more nonmonotonic than those of de Lange. These results are shown in Figure 72, where it is obvious that Kelly's threshold at very low frequencies indicated the eye to be much less sensitive than did de Lange. That is, the difference between Kelly's edgeless field curve and de Lange's curve can be attributed to the contribution of spatial frequencies to de Lange's data and not to Kelly's.
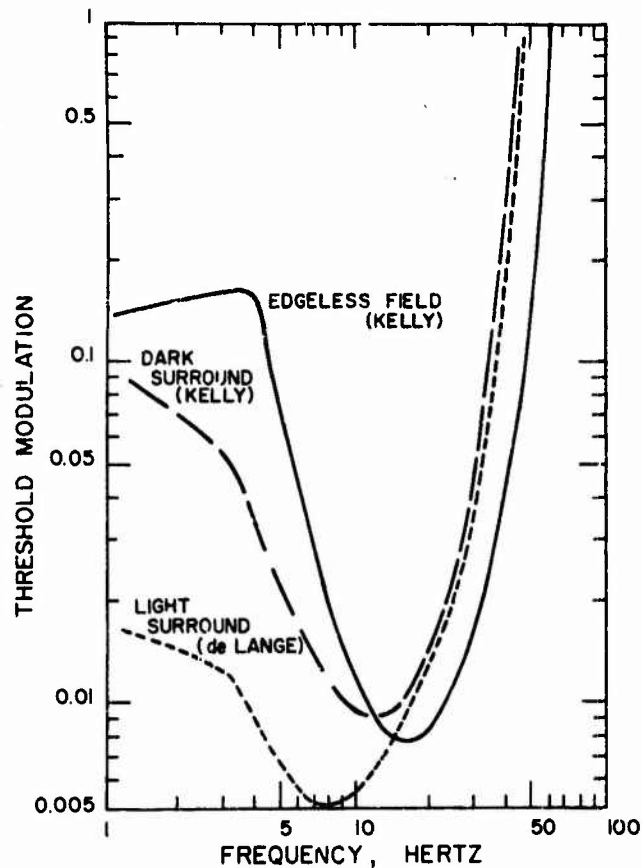
Figure 72: Contrast sensitivity function for three
different surround conditions, from Kelly
(1959)

Kelly only explored temporal frequencies to slightly
below 2 Hz. Almagor, Farley, and Snyder (1979) extended the
temporal sensitivity curves further, to 0.01 Hz, and found
that sensitivity continued to decrease to this very low fre-
quency. Therefore, the eye shows a sensitivity to temporal
stimulus variation much as it does to spatial stimulus vari-
ation. It has a mid-frequency range to which it is most
sensitive, with a reduction in sensitivity falling off as
one either increases or decreases the temporal frequency.
Perhaps this general relationship is most interestingly

plotted in Figure 73, also taken from an experiment by Kelly (1961). The top figure illustrates the absolute modulation amplitude required for a threshold flicker response as a function of temporal frequency. The five highest retinal illuminance levels form a common curve at the high frequency end, indicating that the high frequency response of the visual system to flicker is independent of the actual retinal illuminance level and is dependent only upon the depth of modulation of the sinusoidally varying stimulus. At the low temporal frequency end, however, the sensitivity to absolute depth of modulation is clearly dependent upon retinal illuminance, with greater sensitivity (less modulation) obtained for lower retinal illuminance levels. This is a matter of key importance for the display designer because it means that, at any high temporal frequency, when the observer will be just able to detect an intensity modulation of a given amplitude, increasing the display luminance while maintaining a constant absolute modulation will stop the apparent flicker. However, only a small increase in refresh rate will also stop the flickering because the high frequency end of this combined response is very steep and common to all luminance levels.

The bottom half of Figure 73 normalizes the results in terms of average retinal illuminance, plotting on the ordinate the relative sensitivity or modulation, in the traditional form. Thus, the ordinate now represents not the
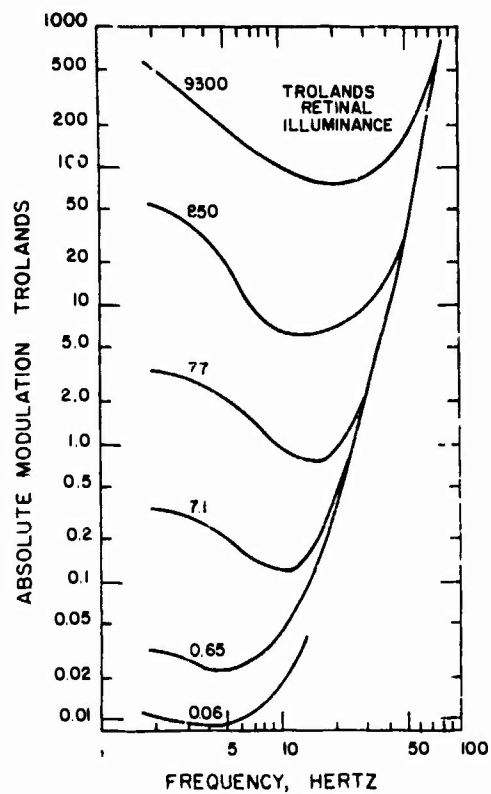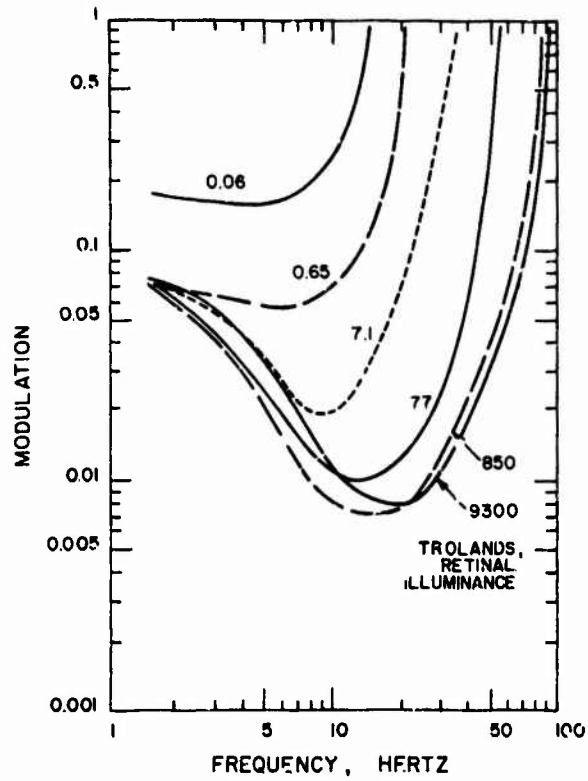
Figure 73: Results of Kelly's (1961) experiment, plotted as
contrast sensitivity function (top) and as
absolute modulation (bottom)

absolute amplitude of the modulation itself, but rather the relative change of this modulation to its average value, which is the usual CSF. As this figure shows, when plotted in this way, the curves are largely equivalent at low temporal frequency, but vary considerably from one another at high temporal frequency. Operationally, this means that if low frequency modulation is approximately at the threshold amplitude, and then one adds a veiling constant illuminance to the display (as in turning on room lights), the modulation will drop below the flicker threshold irrespective of the display luminance level. Thus, displays which are designed to be approximately at the flicker threshold for low frequency temporal variation and low luminance levels will be decidedly non-flickering if either a constant level of illuminance is added to the display, or if the present modulation is slightly decreased, thereby pushing the modulation below the threshold curve in Figure 73 for low frequency conditions.

As shall be seen in the next section, however, the low temporal frequency effects are not independent of spatial frequency. Thus, lowering the modulation at low temporal frequencies may eliminate flicker for some spatial frequency information, but not for other. While this interaction may seem to complicate design criteria, it is typically of little consequence as the more practical problems (and real display designs) exist at the higher temporal frequenices, not the lower ones.

4.2.3 Application of Temporal CFF to Nonsinusoidal Stimuli

Assuming that the CSF for the temporal characteristics of the visual system is essentially as that shown in Figure 73, a critical question is whether this concept can handle non-sinusoidal time-varying stimuli.

This question was addressed in the experiments of Brown and Forsyth (1960), who varied the three components of a periodic stimulus as illustrated in Figure 74. Basically, the duty cycle within each of the A, B, and C portions of the stimulus train was kept at 50%, and the relative periods of A, B, and C were varied. A and B were set at a specific level and the subject set C at the threshold value. Figure 75 shows the general results obtained in this series of experiments.

It can be seen that there are often three frequency values of C which produce a threshold CFF, depending upon the specified values of A and B. These functions are largely symmetrical, indicating that the order of presentation of A, B, and C is unimportant in the wave train, but rather that the subject is responding largely to the relative periods of A, B, and C, in context. Using a Fourier analysis model, Forsyth (1960) took the Fourier components of all such threshold points for the entire wave train and found, as illustrated in Figure 76, that a consistent prediction of threshold response could be obtained from the Fourier fundamental. In Figure 76, all combinations to the right of the

lines appear fused, while those to the left appear to flicker. Thus, fundamental amplitude seems to be an excellent predictor of CFF. The eye appears to behave as a Fourier analyzer of such pulse trains and the threshold can be predicted from the Fourier spectrum.
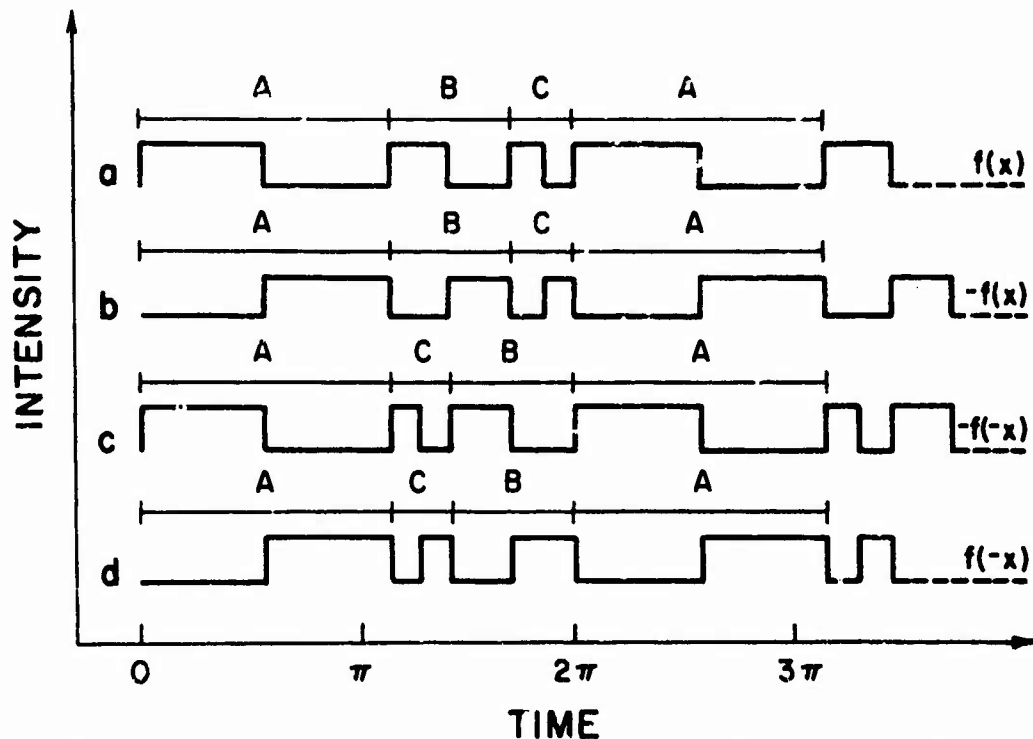
Figure 74: Stimulus trains used by Forsyth (1960)

Additional studies by Levinson (1960) and de Lange (1961) have also concluded that a Fourier analysis approach leads to reasonably good prediction of the sensitivity of the eye to flicker. As we shall see subsequently in Section 5, it is possible to apply the Fourier transform to various types of complex decay functions of representative displays and still retain a reasonable approximation to prediction of the threshold flicker frequency.

- 254 -
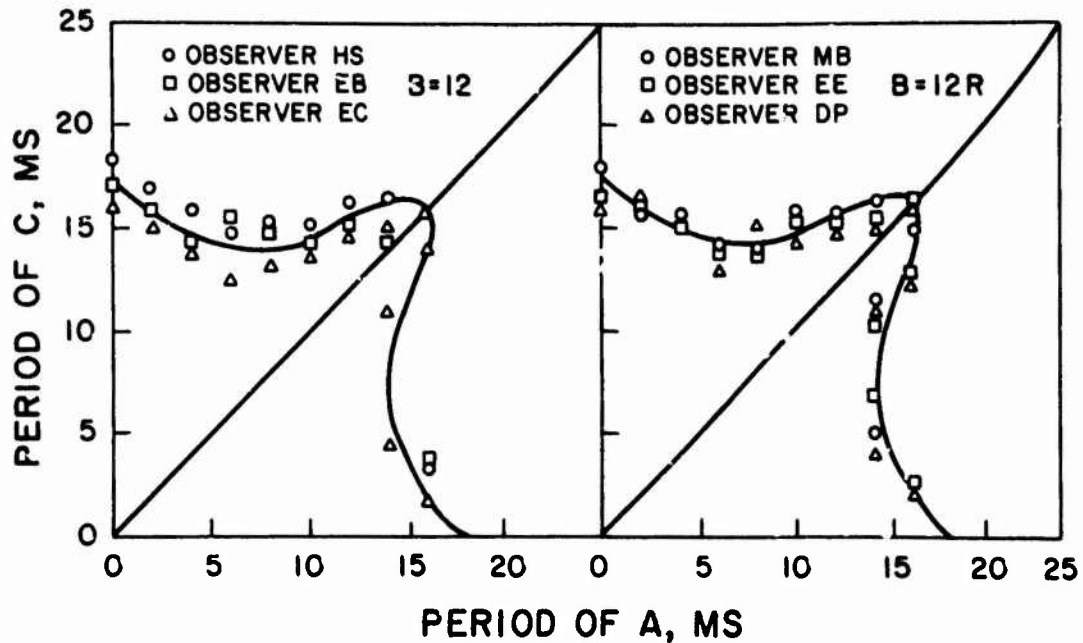
Figure 75: Fusion contours of Forsyth (1960)

At this time, the data pertaining to the effect of other stimulus and field parameters upon sinusoidal temporal CFF values have not appeared. Thus, it is hoped that, in the near future, we will find reported studies of the effect of stimulus variables such as retinal location, retinal field size, stimulus movement rate, chromatic field characteristics, chromatic contrast, and others upon the temporal MTF of the visual system.

4.3   SPATIOTEMPORAL INTERACTION

It was indicated above that there is a typical spatial contrast sensitivity curve for the eye, and that this curve will shift in absolute value as a function of several display parameters. Similarly, the immediately preceding por-
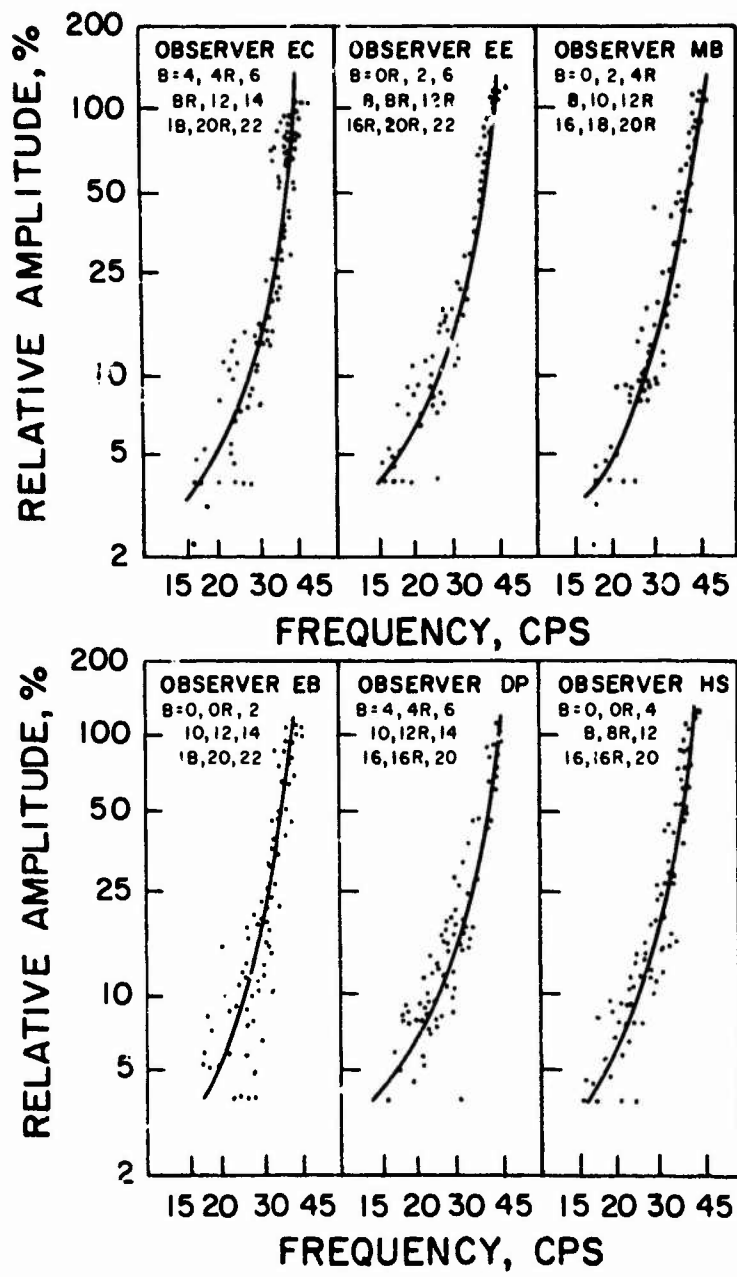
Figure 76: Relative amplitude versus frequency of the
Fourier fundamental, from Forsyth (1960)

tions of this section have indicated that there is a typical

temporal CSF, which has a shape very similar to that of the

spatial function.

One is then led to question the independence of these two functions, and how one might choose to combine them for practical applications. The following information indicates clearly that these two characteristics of the human visual system are not in fact independent of one another.

Kelly (1966) described the subjective phenomena which resulted from concomitant variation in spatial and temporal frequency. His results, illustrated in Figure 77, indicate that the spatial frequency appears to double if the temporal frequency is in excess of approximately 7 Hz and the spatial frequency is less than approximately 3 cyc/deg. If the temporal frequency is between approximately 3 and 7 or 8 Hz, then the display has apparent motion to it, with this motion persisting if the temporal frequency is increased as well as if the spatial frequency is increased. Finally, below approximately 3 Hz, the observer views a slow flicker phenomenon. For the particular conditions of observation in Kelly's experiment, the CFF ranged from 20 Hz for a spatial frequency of approximately 20 cyc/deg to 50 Hz for a spatial frequency on the order of 0.3 cyc/deg. The spatial CSF and the temporal CSF are certainly not independent of one another, and the display designer must clearly take into account this nonindependence in his allocation of the finite bandwidth of any given display system to the spatial and temporal characteristics.
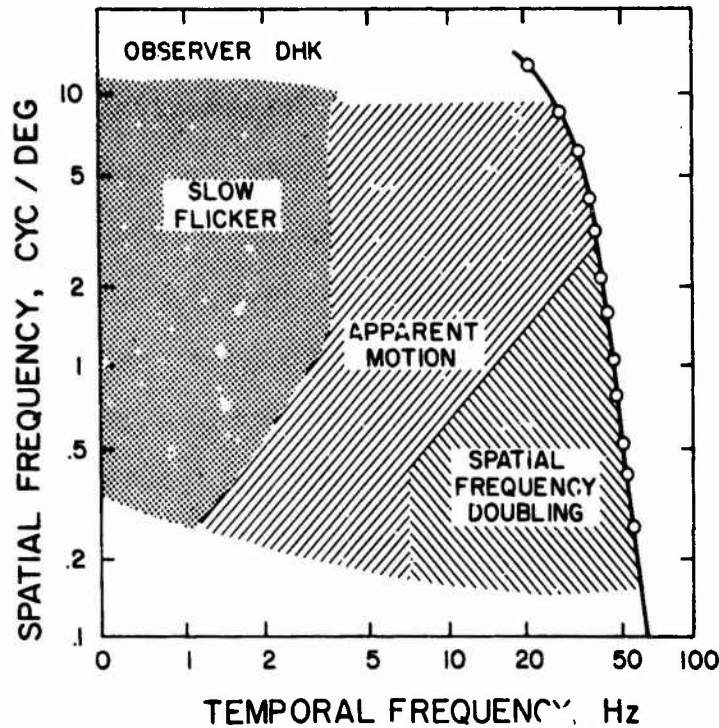
Figure 77: Approximate map of subjective phenomena encountered at various spatial and temporal frequencies, from Kelly (1966)

Perhaps of equal importance to Kelly's subjective phenomena, are the relative sensitivity data presented by Robson (1966). As shown in Figure 78, the falloff in sensitivity at high spatial frequencies is independent of the temporal frequency; similarly, the falloff in sensitivity at high temporal frequencies is independent of the spatial frequency. This independence would obviously be desirable from a practical standpoint were it not the case that the independence breaks down at low frequencies in both the spatial and temporal domains. For example, the falloff in sensitivity at low spatial frequencies occurred only when the temporal frequency was low, and the falloff in sensitivity at low temporal frequencies occurred only when the spatial frequency was low.

- 258 -
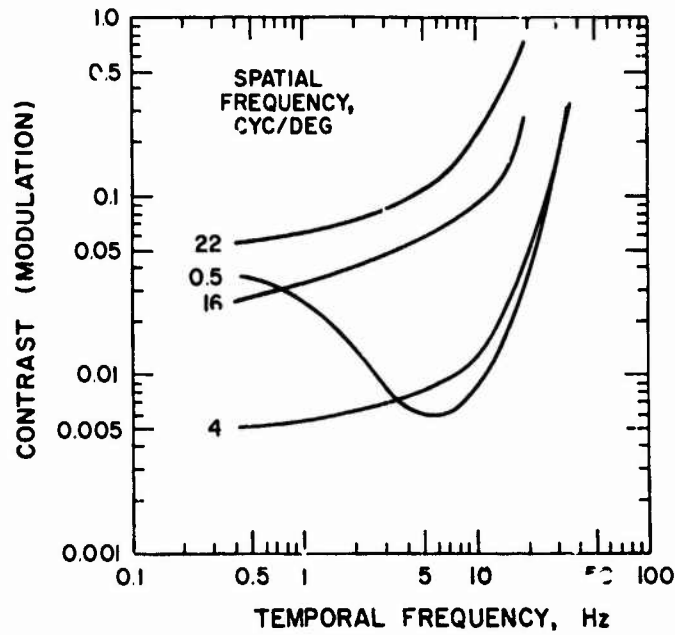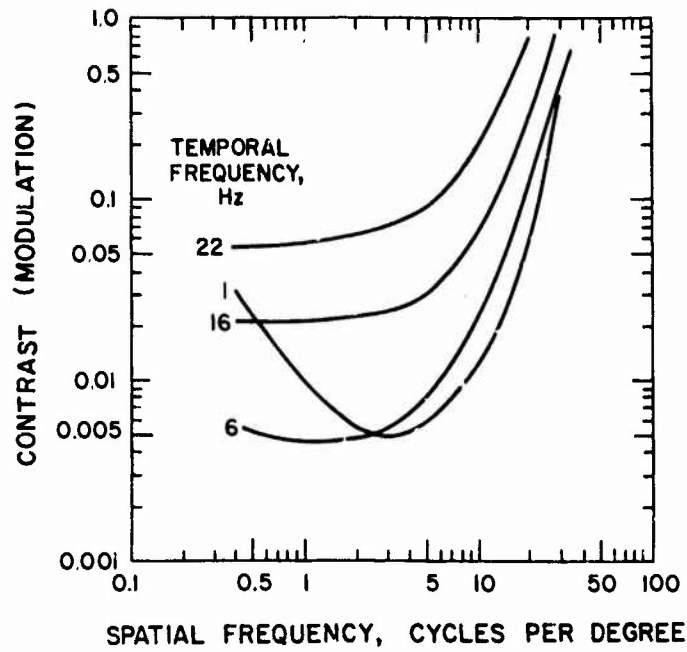
Figure 78: Robson's (1966) spatiotemporal thresholds

Robson's data cause us to consider possible applications
of the information. For example, referring to Figure 78,
one concludes that objects with a spatial frequency of 10
cyc/deg are more readily seen at a temporal frequency of 1

Hz (decidedly flickering) than at 22 Hz. Conversely, if the object of concern has a spatial frequency of 0.7 cyc/deg, it is more readily seen at a temporal frequency of 6 Hz than at either a higher or a lower spatial frequency. In fact, referring to Figure 78, one is forced to conclude that since it is not necessarily possible to control the spatial frequency of all information presented on a display, then the best single temporal frequency at which one can drive a display is at approximately 6 Hz, because at this frequency one does not suffer the reduction in contrast sensitivity as a function of temporal frequency for the low spatial frequency objects. If, on the other hand, one had total control over the spatial frequency of any item to be presented on the display (as in a computer-driven alphanumeric display), then the greatest sensitivity is obtained with a spatial frequency of 4 cyc/deg and a temporal frequency of approximately 0.5 Hz.

Apparently, one is faced with the rather awkward conclusion that the greatest sensitivity of the visual system, and therefore perhaps the best point design for a display system, is one at which the spatial frequency of the presented information is quite reasonable (4 cyc/deg), but that the information must be presented at a very slow refresh rate, so that the display is decidedly flickering to the eye. Though there is some evidence to indicate that the greater the displayed information is above threshold sensitivity,

the greater the utility of the displayed information or its recognizability (Snyder, Keesee, Beamon, and Aschenbach, 1974), one seriously questions whether this generalization holds for temporal frequencies as it does for spatial frequencies. To date, there has been only one research study that related human operator performance to this combined spatial/temporal sensitivity function. Those results were decidedly negative (Silman, 1978).

One experiment (Williams and Erickson, 1974) has attempted to relate contrast sensitivity to spatial frequency, temporal frequency, adaptation luminance, and retinal position. These results are what one would expect from such a study, although the data are highly irregular and inconsistent in terms of the smoothness of the curves. In general, at any given temporal frequency, the contrast sensitivity was at a maximum at some middle spatial frequency, usually about 3.7 to 7.4 cyc/deg. For the lower spatial frequencies, maximum sensitivity was obtained at a middle temporal frequency, typically about 7 Hz. For the higher spatial frequencies, however, sensitivity tended to increase as temporal frequency decreased. These relationships are somewhat similar to those of Robson (1966), except that Robson's data were considerably smoother in form. Although the data of Robson did not contain the typical peak spatial frequency sensitivity at about 3 to 10 cyc/deg, the data of Williams and Erickson (1974) do contain such a peak value,

which agrees with the more consistent results in the experimental literature.

No reason is known for the lack of Robson finding the nonmonotonic nature of the sensitivity curve as a function of spatial frequency.

4.4 CHROMATIC SENSITIVITY AND DISCRIMINATION

The previous discussion has been concerned primarily with the sensitivity of the human visual system to various levels of luminance, both at threshold and above threshold, for a variety of viewing conditions. Except for preliminary mention of the visual stimulus as having a specific wavelength and subjective color, no attention has been paid to the concept of color vision. This relative lack of emphasis has been deliberate, primarily because most display problems and designs in the past have typically been of a monochromatic nature, and in many cases of an achromatic nature. More recently, however, emphasis has shifted toward the use of color coding of displays and to the presentation of both natural color and pseudo color images of literal scenes. Extensive computer graphics hardware and software have hastened the development of this emphasis. Accordingly, this section summarizes human visual sensitivity to color (chromatic) displays in a form compatible with visual sensitivity to luminance and temporal information.

4.4.1 Colorimetry

Prior to 1931, the National Physics Laboratory used the RGB System for specification of color stimuli. The primaries in the RGB System were 700, 546.1, and 435.8 nm. The last two of these primaries are spectral lines of the mercury arc lamp. However, it was realized for many years that the combination of these three primaries required, in some cases, negative coefficients to generate all visible colors. For this reason, the CIE (Commission Internationale de l'Eclairage) in 1931 adopted the international XYZ system, which specified the Standard Observer for Colorimetry. As illustrated in Figure 79, the standard observer chromaticity diagram for the XYZ system has its side XY tangent to the spectral locus at the red end, and has its side YZ very nearly tangent to the spectral locus at approximately 503 nm. Chromaticity values lying outside the spectral locus represent imaginary colors, and all real colors are contained within this spectral locus. Thus, using the convention that X, Y, Z are the primaries of this system, with their respective quantities labeled x, y, and z, it can be shown that the sum of $y + y + z = 1.0$, and that all visible colors can be defined in terms of some combination of components of X, Y, and Z.

A further convenience of this system is that Y is defined specifically as the photopic luminosity function, as shown in Figure 80, which also describes the tristimulus values
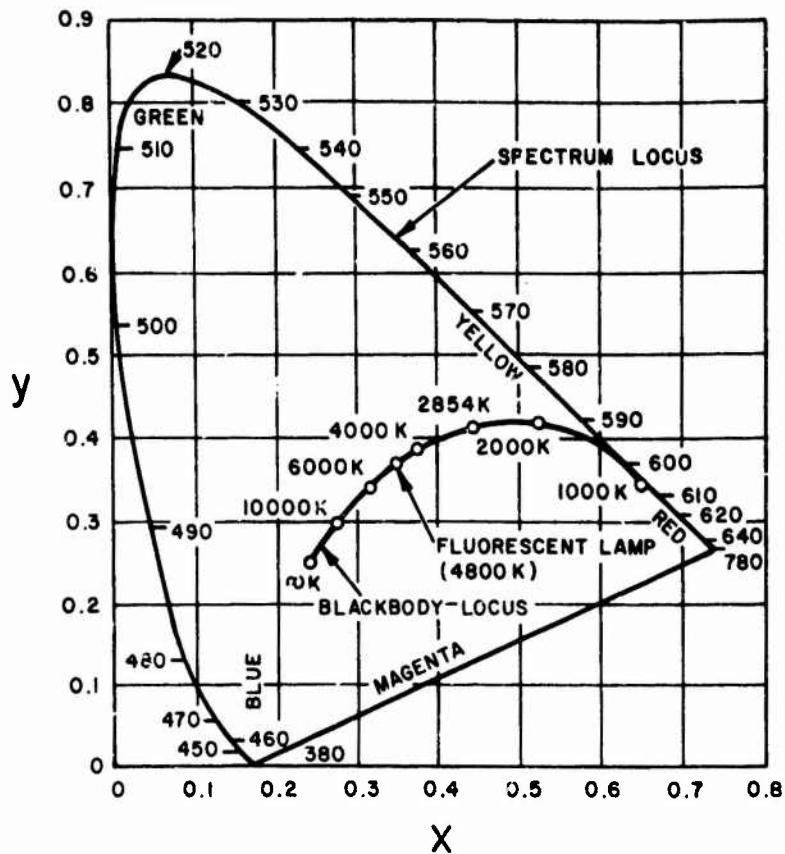
Figure 79: CIE chromaticity diagram for the XYZ system

for an equal energy spectrum within the XYZ system. Note that the distribution of X is bimodal, having major components at both the blue and the red end of the spectrum. Because this system is defined such that the sum of x, y, and z is unity, one needs only to plot any two of these primary coefficients to fully define a color. (Although Y is the photopic luminosity function, one should not equate y to "brightness," the subjective correlate of luminance.)

Another valuable property of the XYZ system is that energy source E is at the center (x=y=z). Thus, the excitation purity of a stimulus is the proportion of that stimulus from E to the spectral locus.
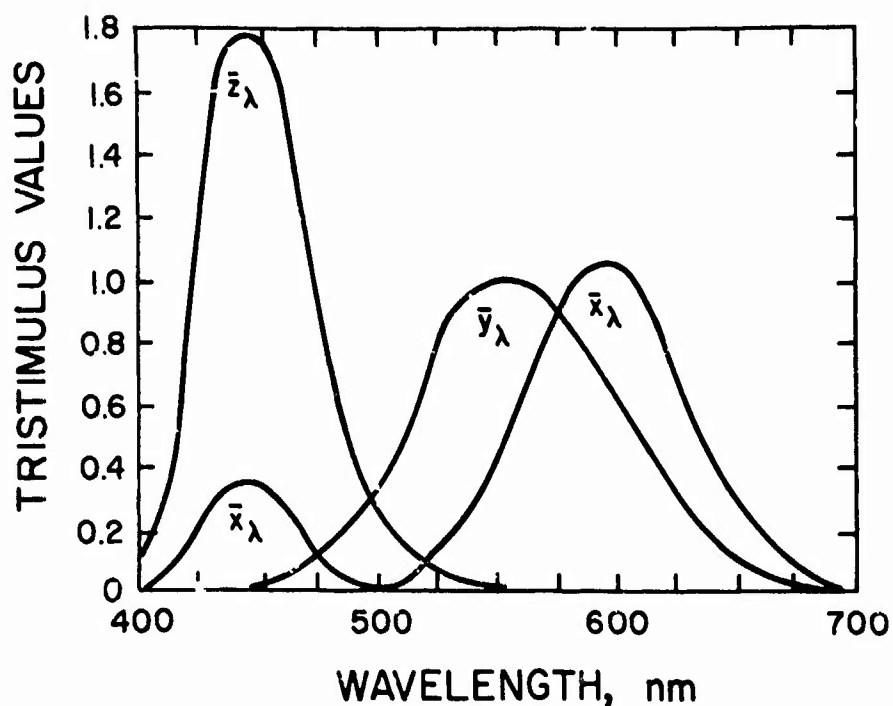
Figure 80: Tristimulus values for equal energy spectrum

4.4.2 Chromaticity Discrimination

In most practical applications, the designer is concerned with using color coding to make discrimination among various items as easy as possible for the observer. Thus, he attempts to choose colors which are maximally discriminable, with the exact specification of such discriminable colors highly dependent upon the number of colors or items which he must separately code. In part, the early work of MacAdam (1949) suggests the relative discriminability of persons for various locations within the visible spectrum. As illustrated in Figure 81, the discriminability is greatest in the extreme blue and red ends of the spectrum, and is relatively poorest in the green, bluish-green, and yellowish-green regions. The ellipses in Figure 81 are approximately 10

- 265 -

times the standard deviation of settings for chromaticity matches, and therefore represent approximately 10 times a measure of the just noticeable difference. These data, then, form a basic source of information regarding the extent to which either the purity or wavelength of a color must be changed to obtain a just noticeable difference, or a barely discriminable difference between two colors. (This argument deals only with <u>threshold</u> discrimination, not suprathreshold differences.)
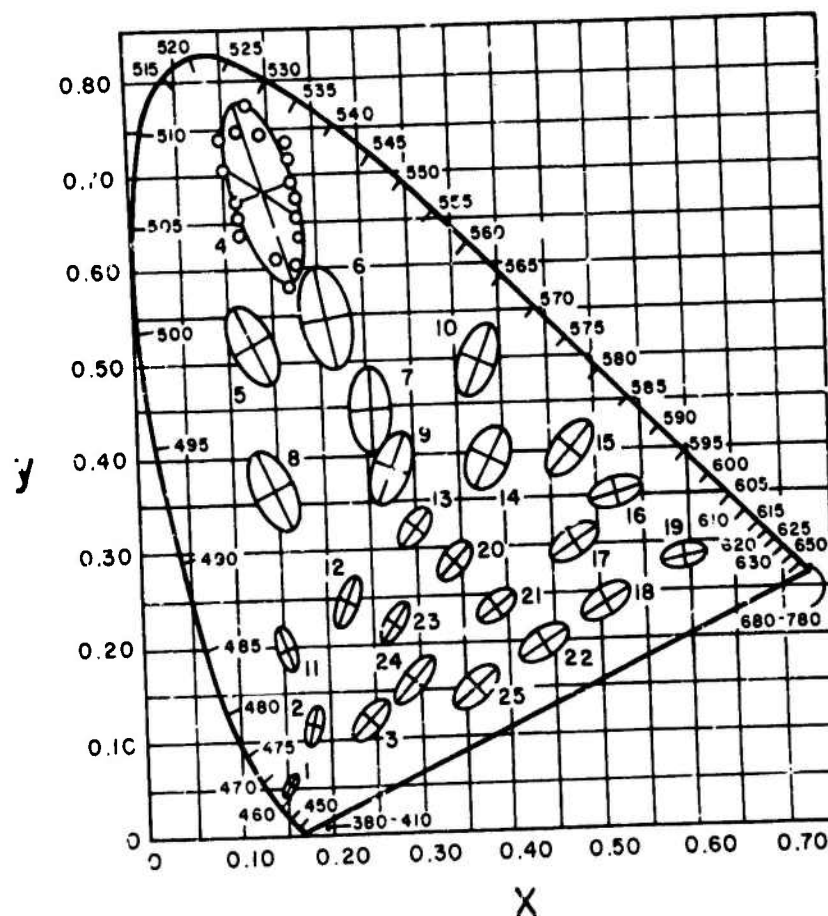
Figure 81: The MacAdam (1949) ellipses

4.4.3   Chromatic Sensitivity Function Approach

In 1958, Schade reported that the human chromaticity response to spatially displayed sine-wave patterns of constant luminance was very similar to that observed for sine-wave gratings varying in luminance only. Specifically, Schade's results for chromaticity modulation showed a distinct resonance at medium-low spatial frequencies, much like the resonance obtained for luminance variation in which maximum sensitivity occurs at a spatial frequency on the order of 3 cyc/deg. Subsequently, studies by van der Horst, de Weert, and Bouman (1967) and others used square-wave patterns rather than sine-wave patterns and did not find Schade's sensitivity decrement at low spatial frequencies. The only reported study which used sine-wave patterns was that of van der Horst (1969), who also failed to find the effect obtained by Schade. The van der Horst (1969) experiment was based upon a variation in chromaticity modulation as a function of spatial frequency. Figure 82 illustrates the variability which he induced into a television monitor, which served as a visual stimulus. Sinusoidal variation in chromaticity was presented along the axes BEyy, located at point E, or along the axes bgER, but also at point E. The point E is a standard white, or equal energy point in the spectrum. van der Horst's results are illustrated in Figure 83, which indicates that the threshold sensitivity (measured in percent purity, which varies inversely with sensitivity)

is greatest for a square-wave and least for a triangular wave. The sine-wave sensitivity is approximately midway between the square wave and triangular wave sensitivity.
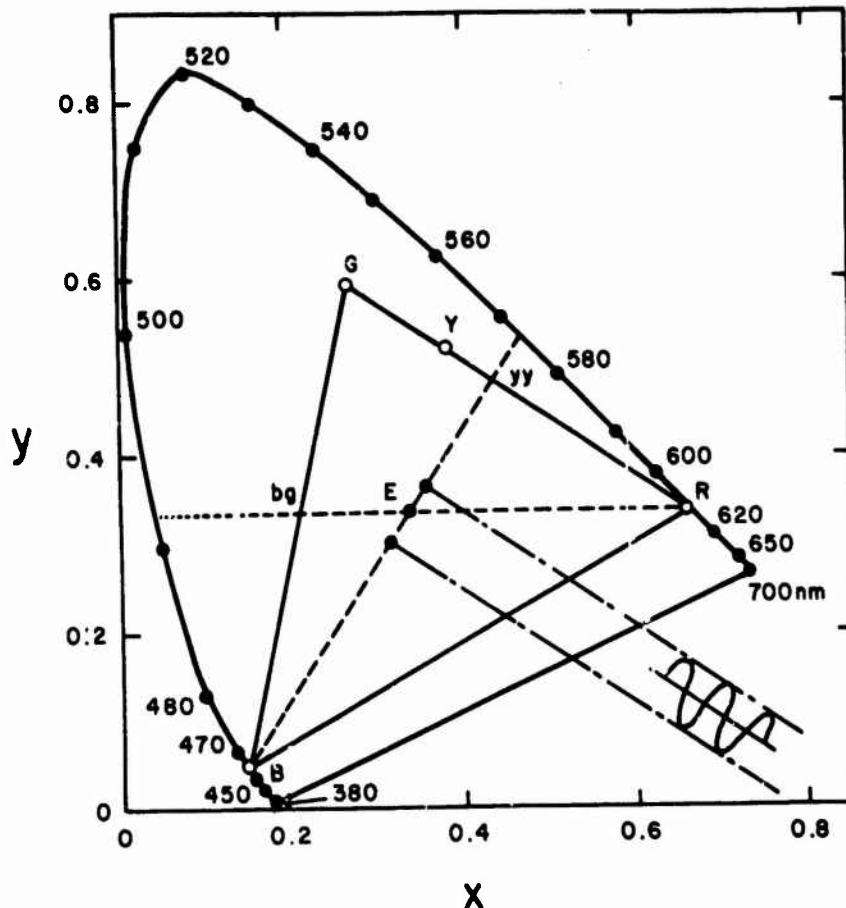
Figure 82: CIE chromaticity diagram stimuli of van der Horst (1969)

If one computes the first harmonic of these three waveforms, the harmonics are of the order of 1; 1/1.27; 1/0.81, for the sine-wave, square-wave, and triangular-wave gratings, respectively. It was indicated above that the Fourier spectrum of the square wave is given by
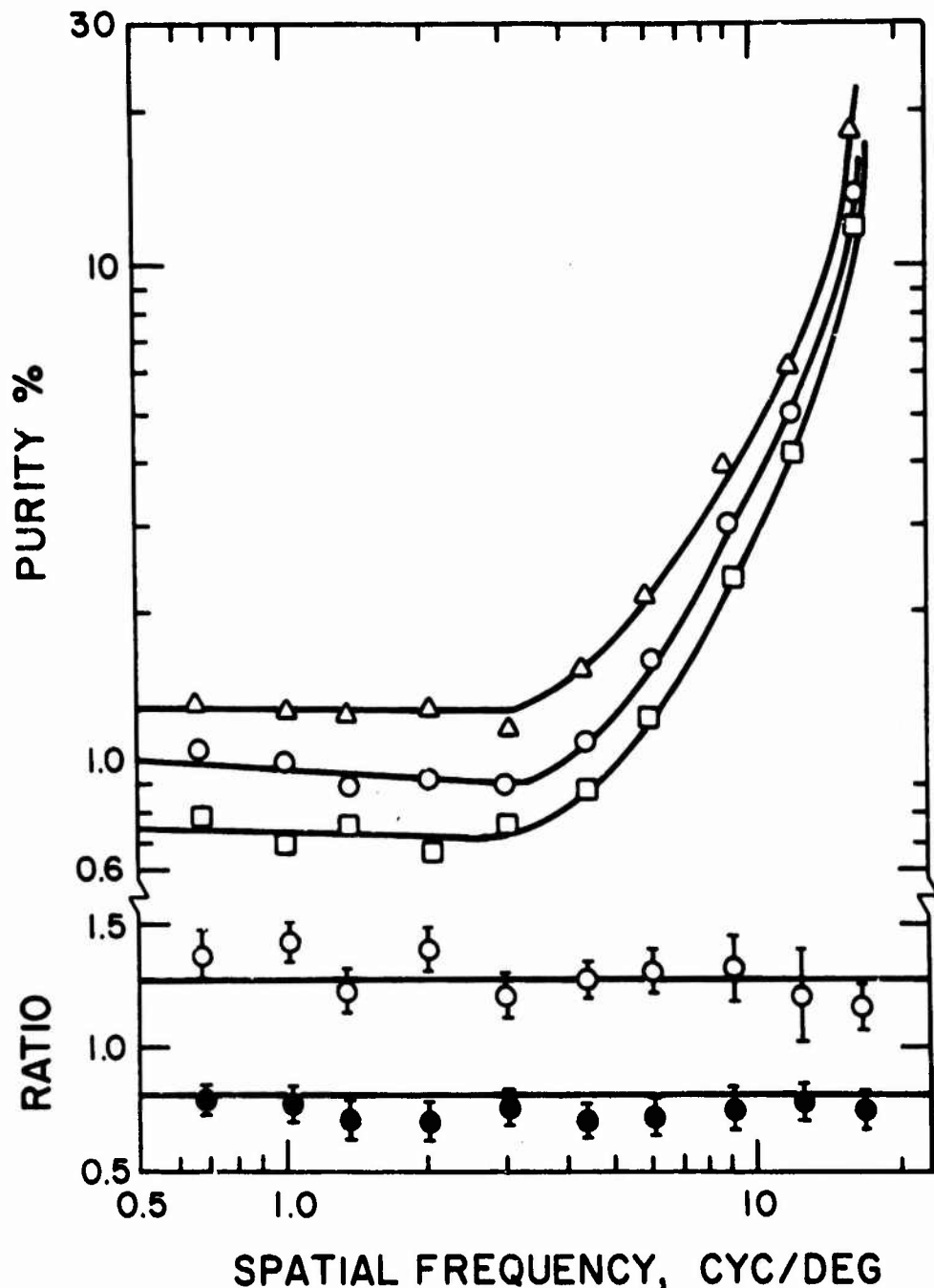
Figure 83: Chromatic threshold-contrast modulation: square
wave, sine-wave, triangular wave gratings, from
van der Horst (1969). Purity is in percent for
a dominant wavelength of 492 nm. The ratios of
sine wave to square wave (O) and sine wave to
triangular wave (●) are plotted at the bottom.
Lines drawn are at 1.27 and 0.81.

$$f(x) = 4/\pi \; (\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} \; ...), \qquad (18)$$

noting that $4/\pi = 1.27$.

For the triangular wave, the Fourier components are

$$f(x) = 8/\pi^2 \; (\cos x + \frac{\cos 3x}{9} + \frac{\cos 5x}{25} \; ...), \qquad (19)$$

noting that $8/\pi^2 = 0.81$.

In the bottom of Figure 83, the actual ratios of sine-wave to square-wave, and sine-wave to triangular-wave sensitivities are plotted. These ratios show good agreement with the expected ratio of the fundamental amplitudes of 1.27 and 0.81. Thus, the visual sensitivity to chromatic gratings, at this particular locus E of the CIE chromaticity diagram, is predicted well by a Fourier analysis of the grating waveform, and there is essentially no sensitivity to higher order harmonics of the fundamental wave. That is, the visual system appears to have very little nonlinearity at the low spatial frequency end, and responds almost exclusively to the Fourier fundamental of the waveform without any degradation at the low spatial frequency end.

These results were replicated by Granger and Heurtley (1973), who also concluded that there was no appreciable reduction in sensitivity with spatial frequencies as low at 0.1 cyc/deg, thereby extending the data of van der Horst to

an even lower spatial frequency. Of further interest in the data of Granger and Heurtley (1973) is the fact that they were unable to obtain a pure chromatic response above approximately 3 cyc/deg. All their subjects reported an induced luminance-contrast component that they could not null out by adjustment of the modulation of either primary color of their monitor. Thus, in excess of approximately 3 cyc/deg, subjects reported a luminance confounding with the actual chromatic difference in the display. This result suggests that the visual system becomes unequally sensitive to various colors above approximately 3 cyc/deg, at least along the red-green axis which they were using. The chromaticity coordinates of the red and green phosphors (x,y) were, respectively, (0.621, 0.349) and (0.289, 0.590). From a practical standpoint, these data suggest that color coding at spatial frequencies in excess of 3 cyc/deg may well be viewed by the subject as having a luminance coding component, which may or may not enhance detectability of such coded symbols. Conceivably, this perceptual confounding could reduce detectability of color coded elements having a fundamental spatial frequency greater than 3 cyc/deg.

A further study by van der Horst (1969) directly compared the luminance threshold contrast modulation with the chromatic threshold contrast modulation, as indicated in Figure 84. His results also indicate that the chromatic modulation function is monotonic, with a much lower bandpass than the

luminance modulation function, which he also found to be nonmonotonic and with a higher passband than that for chrominance.
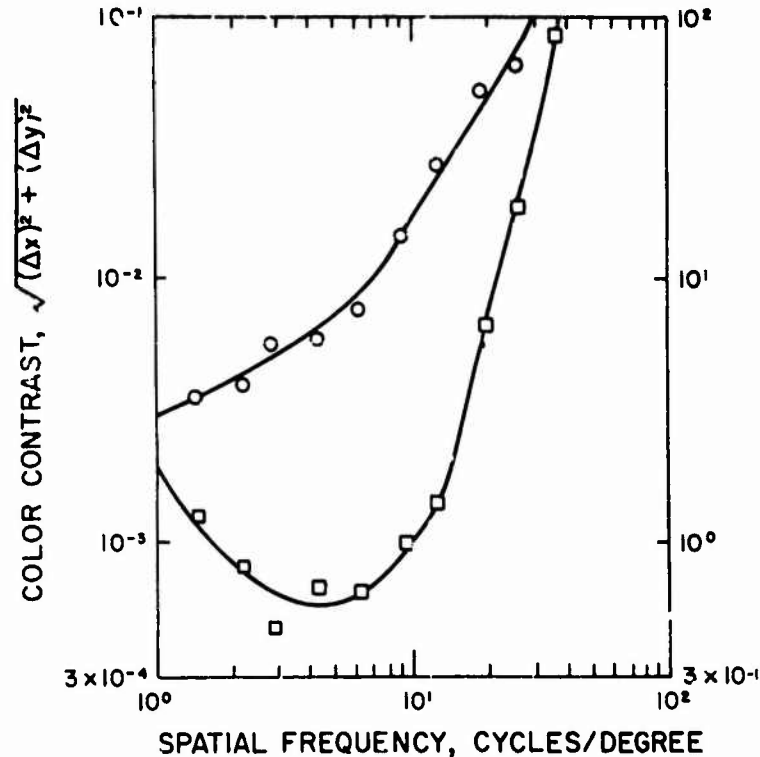
Figure 84: Luminance and chrominance CSF, from van der Horst (1969)

4.4.4   Temporal Modulation of Chromaticity Gratings

It is also interesting to compare the sensitivity for chromatically varying gratings with the sensitivity functions obtained with luminance gratings. van der Horst and Bouman (1969) varied chromatic gratings in both spatial frequency and temporal frequency, using a red-bluish-green modulation around a standard white source, E, and determined the percent purity (modulation) required at the threshold.

- 272 -

percent purity (modulation) required at the threshold. Further, their data also investigated the effect upon this threshold of the overall retinal illuminance. As illustrated in Figure 85, the chromatic threshold sensitivity decreases (purity increases) with decreases in retinal illuminance from 160 to 0.30 trolands. For all retinal illuminance levels, however, there is no suggestion of an inversion in the threshold function at low spatial frequencies. That is, as retinal illuminance decreases, the equivalent passband of the eye also decreases, but the visual system still behaves as a low frequency filter.
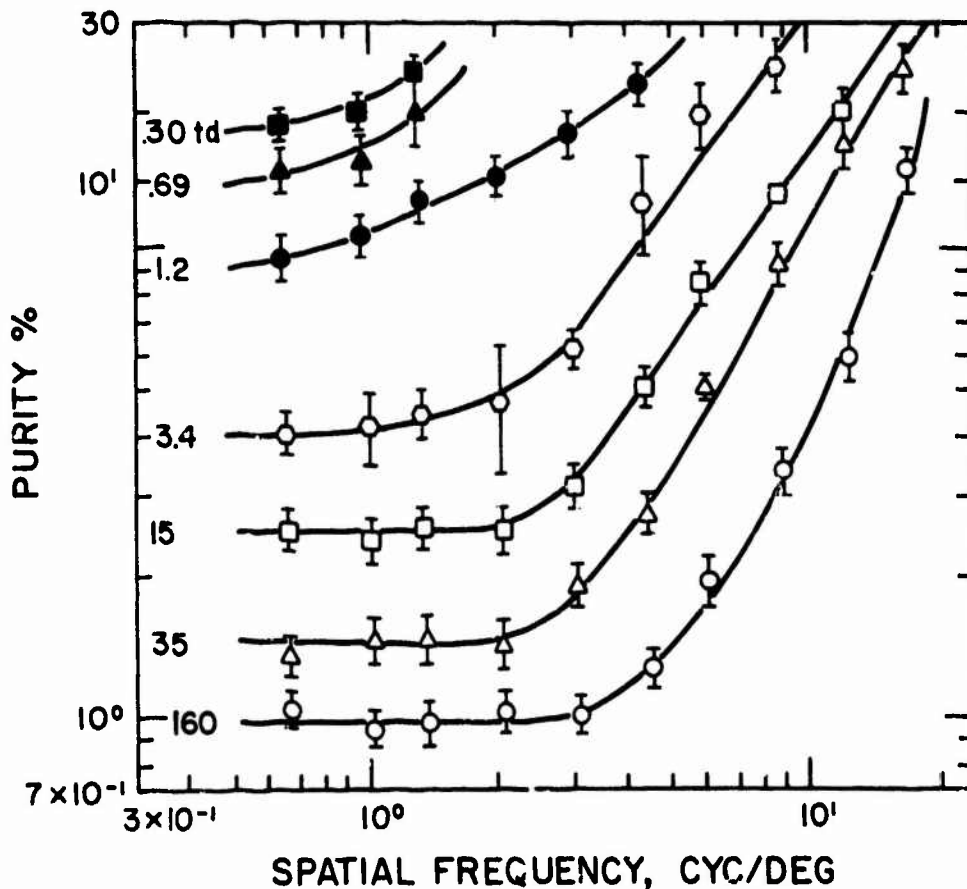
Figure 85: Chromatic threshold-contrast modulation
dependence on the spatial frequency of the sine-wave grating, from van der Horst and Bouman (1969)

van der Horst and Bouman (1969) also investigated the
combined variation in spatial frequency and temporal fre-
quency using the same dominant wavelength of 492 nm. They
found that contrast sensitivity decreases (purity increases)
as either the temporal frequency increases or the spatial
frequency increases, and that there is essentially no inter-
action between these functions. That is, the effect of tem-
poral frequency at all spatial frequencies is monotonic, as
is the effect of spatial frequency at all temporal frequen-
cies. There is no evidence in either set of curves for a
low frequency decrease in sensitivity.

4.4.5 Summary

The eye is less sensitive to color modulation than it is
to luminance modulation, as can be seen by comparing the
data from this section with that of the previous section.
However, the visual system behaves in a similar manner for
both luminance and chrominance as retinal illuminance
changes. With increases in retinal illuminance, increases
in contrast sensitivity are obtained.

The fundamental difference between chrominance sensitiv-
ity and luminance sensitivity lies in two characteristics.
First, the luminance channel appears to have a resonance
around 3 to 7 cyc/deg, whereas the chrominance channel
behaves as a low bandpass filter; the chrominance channel
has no inversion of the sensitivity function at very low

spatial frequencies. Secondly, the passband of the chrominance channel is somewhat less than that of the luminance channel, having a much lower frequency cutoff. Thus, the eye is less sensitive to chromatic differences at high spatial frequencies or viewing small detailed objects. If it is necessary for the eye to detect small features in its environment, then chrominance differences have very little effect upon such detectability. On the other hand, if it is necessary for the eye to detect very large objects, then there is every indication that detection can be made more sensitive by using chromatic coding rather than (or in addition to) luminance coding, due primarily to the reduction in luminance sensitivity and the nonreduction of chromatic sensitivity at very low spatial frequencies.

Section 5

VISUAL REQUIREMENTS AND DISPLAY EVALUATION

This section attempts to relate (1) the existing experimental data pertaining to the known effects of display parameters upon operator performance to (2) the previously described quantitative visual characteristics of the human operator. Following the organization of Section 4, we first discuss spatial (5.1), temporal (5.2), and chromatic (5.3) display parameters, relating them to visual system characteristics. These sections are followed by a summary of attempts to combine these quantitative data into display figures of merit, or unitary measures of image quality (5.4). Finally, Section 5.5 deals briefly with the more intepretive aspects of information display, those pertaining to information encoding and the design problems and parameters which relate to this area of inquiry.

5.1   SPATIAL PARAMETERS

Probably the most important design variables for any visual display are those dealing with the spatial density, organization, and luminance characteristics. The first question typically asked by the designer is "How much resolution is necessary?" and this is generally followed by "How much con-

trast is required?" As indicated previously, this type of question is not as precise as desirable, and furthermore, the interactions between the design variables of "contrast" and "resolution" and other display variables are of considerable importance. Of course, the very concepts of "contrast" and "resolution" are not adequately precise, and further quantitative definition is needed to reach an adequate design specification.

Fortunately, it can be shown that the spatial sensitivity characteristics of the visual system relate fairly consistently and logically to human performance as it is affected by spatial display variables. This section summarizes a small, but carefully selected, portion of the vast literature on this subject, and relates it to the previously developed concept of the contrast sensitivity function.

5.1.1 Resolution/Density

The number of picture elements (pixels) that can be contained on a display is constrained by the size of the display, the element size, the spacing between elements, and occasionally the element shape. In an analog (nondiscrete) display, such as a CRT, there is, in addition, some overlap betwen pixels which can cause heterogeneity of pixel sizes and shapes. Furthermore, until recently, most of the human performance data were obtained with CRT displays which were not addressed in discrete pixel form.

For analog, or continuous image, devices such as photographs or television, it is more meaningful to talk about the general concept of "resolution" than to speak of pixel density. On the other hand, for discretely addressed X-Y matrix displays, the term density is more meaningful. Each is defined below, and representative behavioral data are presented to show typical effects of these design variables.

"Resolution" is used inexactly by many persons. In some contexts, it is intended to mean the number of pixels per unit display distance, which is a valid, precise definition for density in a digital display. More typically, it refers to the number of "resolvable" elements per unit dimension, where the resolvable element is measured either subjectively or photometrically. Because there are at least eight or nine measures of resolution in the literature, and also because many reports are imprecise in defining the selected measure, one must be particularly careful in relating these measures to observer performance. Since the context in which these measures have usually been made is that of a CRT display, this discussion assumes a CRT image. However, it should be realized that the concepts may be applied to some other display types as well.

5.1.1.1   Measures of resolution/density

Sherr (1979, p. 9) defines resolution as "the smallest discernable or measurable detail in a visual presentation."

Indeed, many of the resolution measures are made by one or more persons "discerning" the smallest detail in a far from controlled psychophysical procedure. As a result, disagreement in "resolution" occurs much too frequently. Photometric measurement, followed by specific mathematical analysis, is much preferred.

The most often used measure of resolution for a television image is that of _television limiting resolution_. It is typically measured with a test pattern such as the "Indian Head" or RETMA test pattern. The limiting resolution is simply the spatial frequency (usually expressed in TV lines per picture height or width) at which the observer can no longer discriminate the light and dark bars of the image. It is assumed that the observer's MTF exceeds that of the display system, so that there is no observer limitation to the measurement. Generally, it is also assumed that the "contrast" is about 3% at this observer threshold (_RCA Electro-Optics Handbook_, 1974, p. 119). "Contrast" in this case is defined as the luminance difference between dark and light bars divided by the background luminance. Thus, TV limiting resolution should be equal to the spatial frequency at which the CSF is equal to $0.03/(4/\pi) = 0.03\ \pi/4 = 0.024$. (The $4/\pi$ correction compensates for the square-wave image on the RETMA chart, whereas the CSF is based on a sine-wave grating.)

If a CRT image is formed with a Gaussian spot size, then two adjacent "white" lines will be barely discriminable from an interspersed black line if the line spacings, center-to-center, are 1.18σ, where σ is the standard deviation of the Gaussian distribution. Figure 86 illustrates several measures of the Gaussian distribution, as well as the overlap of lines needed to produce the 0.024 square-wave contrast at TV limiting resolution.

At ± 1σ, the luminance of the spot is equal to 60.7% of its maximum value. At one-half its maximum luminance value, the "width" of the spot is ± 1.18σ, or 2.35σ wide.

Resolution measurements made using this "half amplitude" are called 50% amplitude measurements. Alternating black and white lines separated by 2.35σ will produce a modulation of approximately 40%, as also illustrated in Figure 86.

If we assume a Gaussian spot distribution, as is reasonably accurate for CRTs, then it is possible to relate these and other measures of resolution to the MTF curve, as suggested by Slocum (1967) and illustrated in Figure 87. Conversion values are given in Table 17.

5.1.1.2   Observer performance and resolution/density

In general, the extensive literature relating CRT resolution to observer performance shows that performance increases with increases in resolution, to some high resolution level at which observer performance reaches a practical
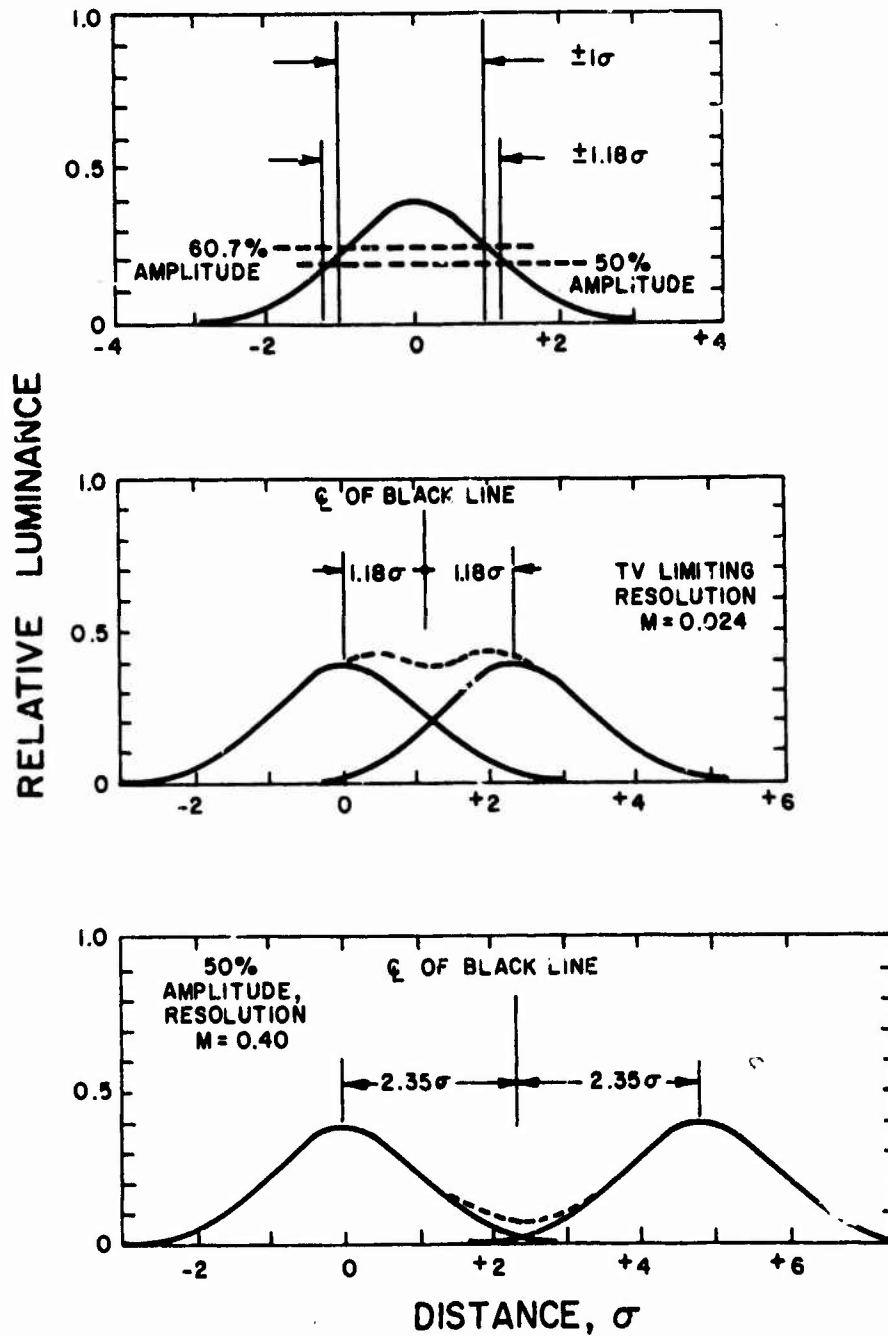
Figure 86: Overlap of Gaussian spots

asymptote. This asymptote is usually caused by either a "ceiling" effect of the performance variable, such as per- cent correct approaching 100%, or by the display resolution exceeding the spatial sensitivity of the observer. There is
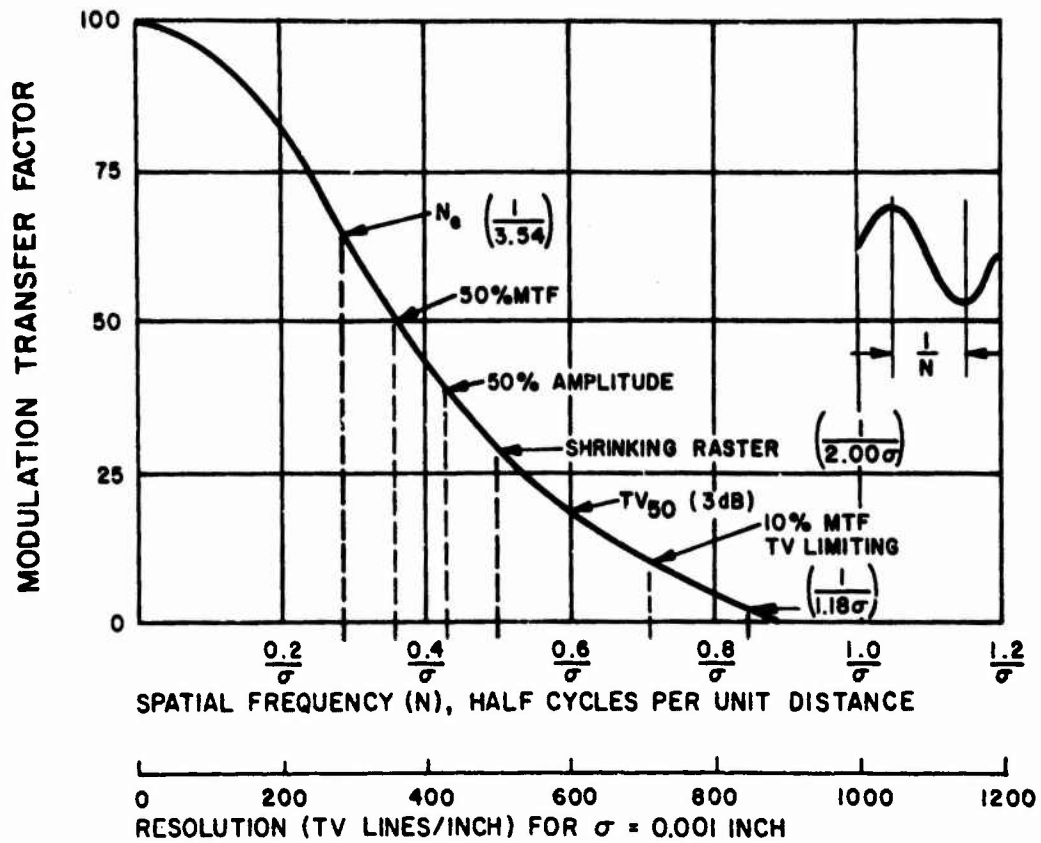
Figure 87: Relative modulation transfer function, after Slocum (1967)

also a task dependency in the literature, with some (typically simple) tasks producing observer performance asymptotes at lower resolution than that required for other (typically more complex) tasks.

It is well beyond the intent of this report to summarize this extensive literature, but a couple of examples are perhaps worthwhile. In an extensive research program on air-to-ground target acquisition via television, Humes and Bauerschmidt (1968) found that there was little difference among 525, 729, and 1024 line raster TV systems for finding large targets, using recognition latency as the performance

TABLE 17. Conversion Table of Various Measures of Display Resolution, from Slocum (1967)

From:		TV Limiting	10% MTF	TV_{50}	Shrinking Raster	50% Ampli.	50% MTF	Optical	Equiv. Passband
							To:		
TV Limiting	1.18 σ		0.80	0.71	0.59	0.50	0.44	0.42	0.33
10% MTF	1.47 σ	1.25		0.88	0.74	0.62	0.55	0.52	0.42
TV_{50} (3 db)	1.67 σ	1.40	1.14		0.84	0.71	0.62	0.59	0.47
Shrinking Raster	2.00 σ	1.70	1.36	1.20		0.85	0.75	0.71	0.56
50% Amplitude	2.35 σ	2.00	1.60	1.40	1.17		0.88	0.83	0.66
50% MTF	2.67 σ	2.26	1.80	1.60	1.33	1.14		0.94	0.75
Equivalent Passband (N_e)	3.54 σ	3.00	2.40	2.10	1.77	1.50	1.33	1.25	

measure (Figure 88). Performance reached asymptote (95%) in 12 seconds. For the smaller, more difficult, targets, no asymptote was reached in 22 s, the maximum performance was 78%, and a larger number of scan lines produced better performance. While no MTF measurements were made in this study, it is quite reasonable to assume that a larger number of scan lines would yield greater resolution (and a larger MTF value at most spatial frequencies) in the direction perpendicular to the raster. Similar results have often been obtained when video bandwidth has been increased to increase resolution in the direction parallel to the raster (Snyder, 1973; 1976).

In a simpler task, Erickson, Linton, and Hemingway (1968) have demonstrated that recognition of alphanumerics increases as either the number of scan lines intersecting the alphanumeric symbol increases or the number of scan lines on the display increases. Figure 89 illustrates these results.

While one might cite numerous CRT-related studies of this nature, there is little advantage in doing so, for it is difficult to generalize from such results to a different application or task, and also because the "resolution" effects often interact with the effects of other display/system variables. In addition, such data fail to yield a safe generalization to the primary subject of this report, the flat panel, matrix display.
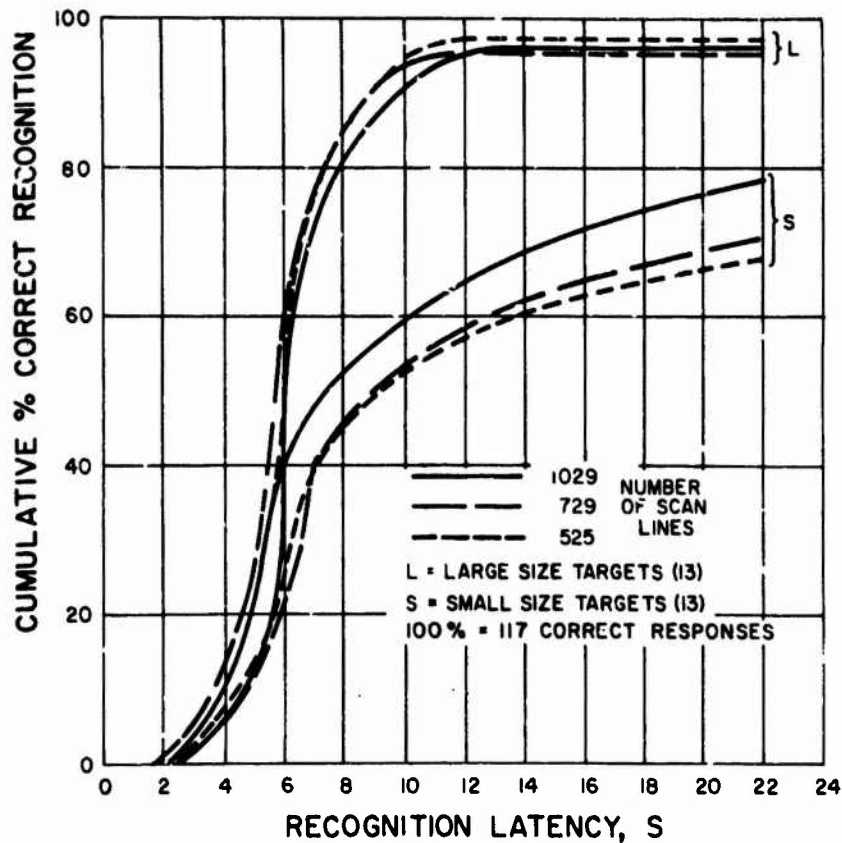
Figure 88: Recognition latency for 525, 729, and 1024 line television systems, from Humes and Bauerschmidt (1968)

The resolution or density of the fixed element (or pixel) flat panel display is determined by the element size, element shape, and element edge-to-edge spacing. Unlike with the CRT, changes in flat panel image content have no appreciable effect on element density. Further, the element density of flat panel displays is essentially invariant everywhere on the display, whereas the resolution of the CRT varies with display position, luminance, focus, and a few other parameters.
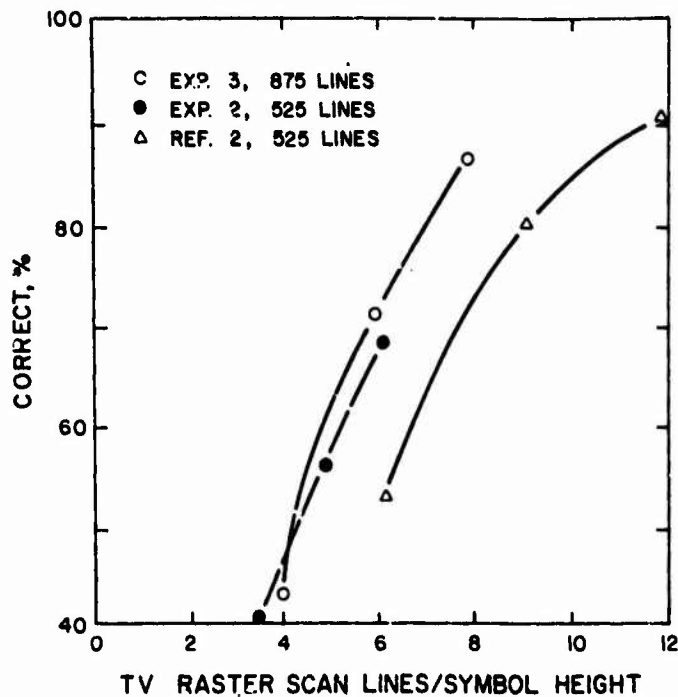
Figure 89: Effects of display raster lines and number of
raster lines intersecting target on symbol
recognition, from Erickson et al. (1968)

Accordingly, it is more meaningful to include data relat-
ing observer performance to flat-panel element density in
the next sections, in which element size, shape, and spacing
are discussed.

5.1.2   Element Size, Shape, Luminance Distribution

The size, shape, and interelement spacing variables of
dot matrix displays are illustrated in Figure 90 (Snyder and
Maddox, 1978), which also indicates how these variables were
independently simulated on a Tektronix 4014-1 computer ter-
minal by multiple addressable point selection.   An experi-
ment described by Snyder and Maddox (1978)   which investi-
gated the effects of element size and element shape is

outlined in   Figure   91,   and   Figure   92 shows the   exact
dimensions of   the size and shape levels investigated.   Per-
formance was   measured in both reading and search tasks.
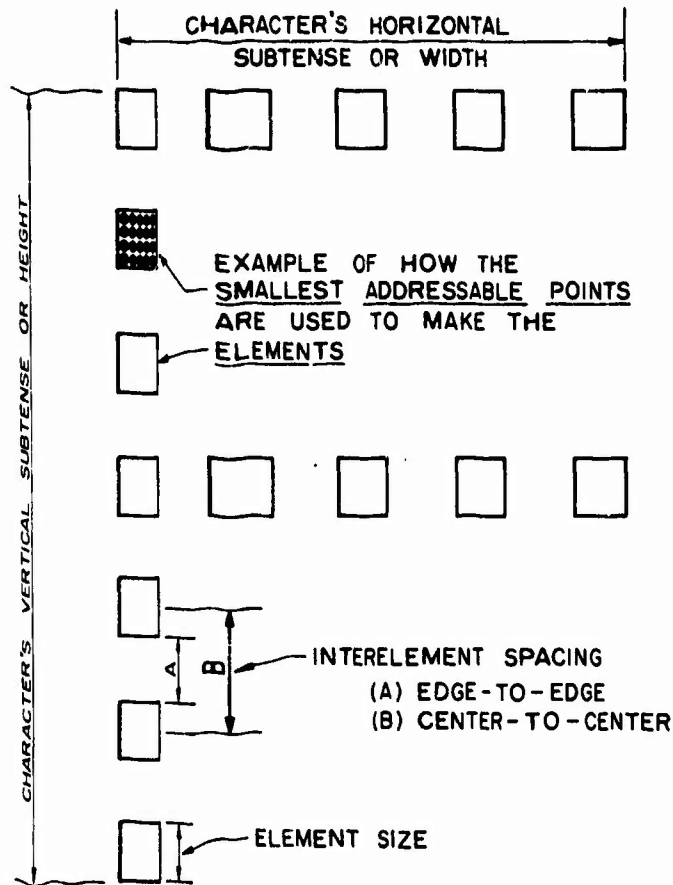
Figure 90: Size, shape, and interelement spacing of dot
            matrix character, from Snyder and Maddox (1978)

Their results, which are also described in more detail by
Burnette (1977), consistently showed that the square element
shape is better than an elongated element shape for either
reading or search (Figures 93 and 94).   Further, when repli-
cated using S-elf-Scan panels having either square or round
elements, the   square elements proved superior for both read-
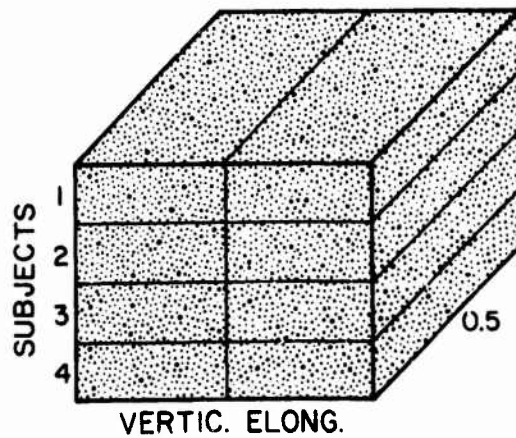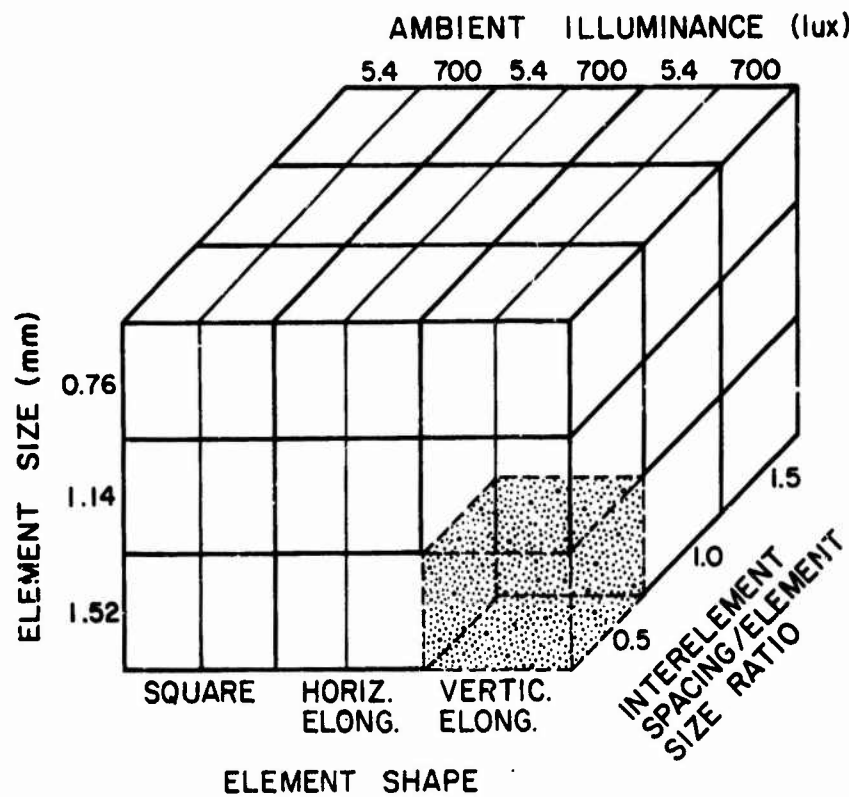
Figure 91: Experimental design, from Snyder and Maddox (1978)

ing and search. Thus, the existing data support the conclusion that the more square the display element, the better the observer's performance.
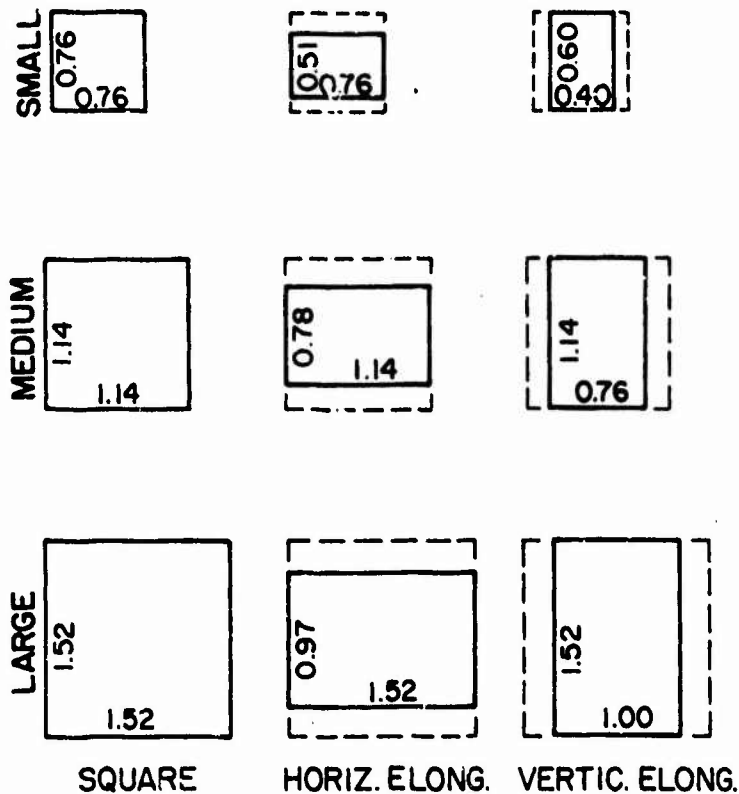
TYPICAL DIMENSIONS OF ELEMENTS (mm)

Figure 92: Element size dimensions, from Snyder and Maddox (1978)

Optimization of element size, however, is a more complex problem. For reading tasks, as in contextual alphanumeric paragraphs, it is possible to make the element size too large, as shown in Figure 95. For search tasks, however, large elements lead to more rapid identification of symbols and characters (Figure 96). Apparently, reading efficiency is obtained by using smaller, compact characters which minimize the number of eye fixations, whereas search, a task requiring detection in the visual periphery, requires larger elements (and hence larger characters). Certainly, these generalizations can be overextended, but within the 0.75-to-1.50 mm element size, they appear consistent.
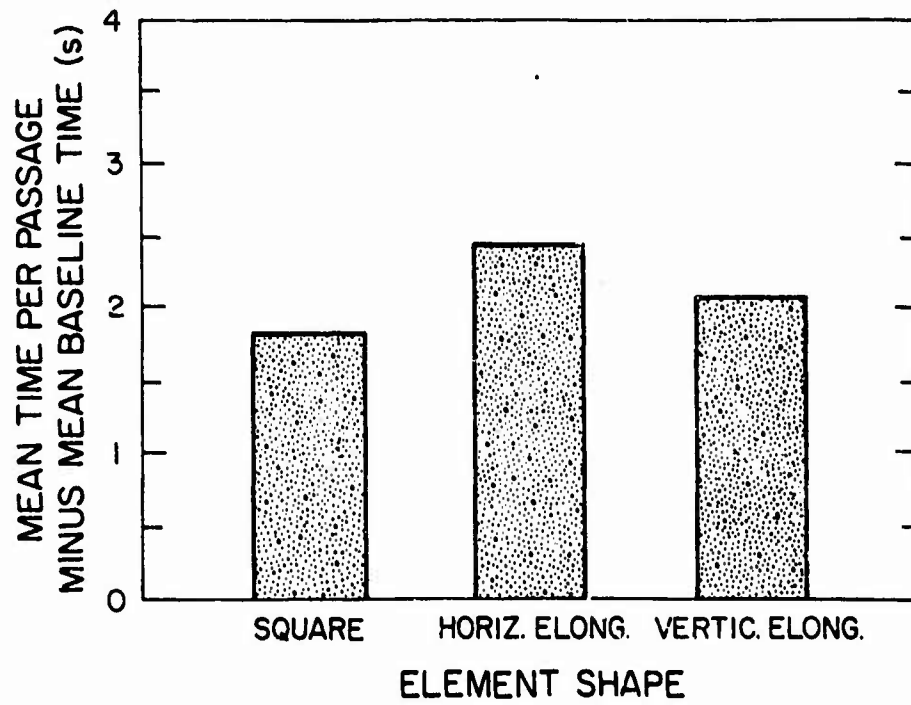
Figure 93: Effect of element shape on reading time, from Snyder and Maddox, 1978)
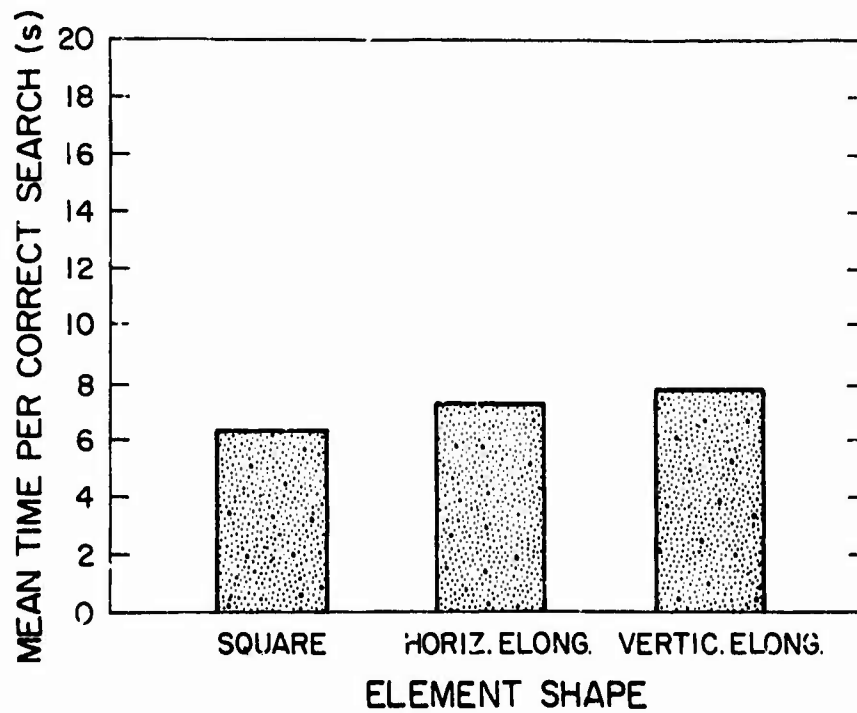
Figure 94: Effect of element shape upon random search time, from Snyder and Maddox (1978)
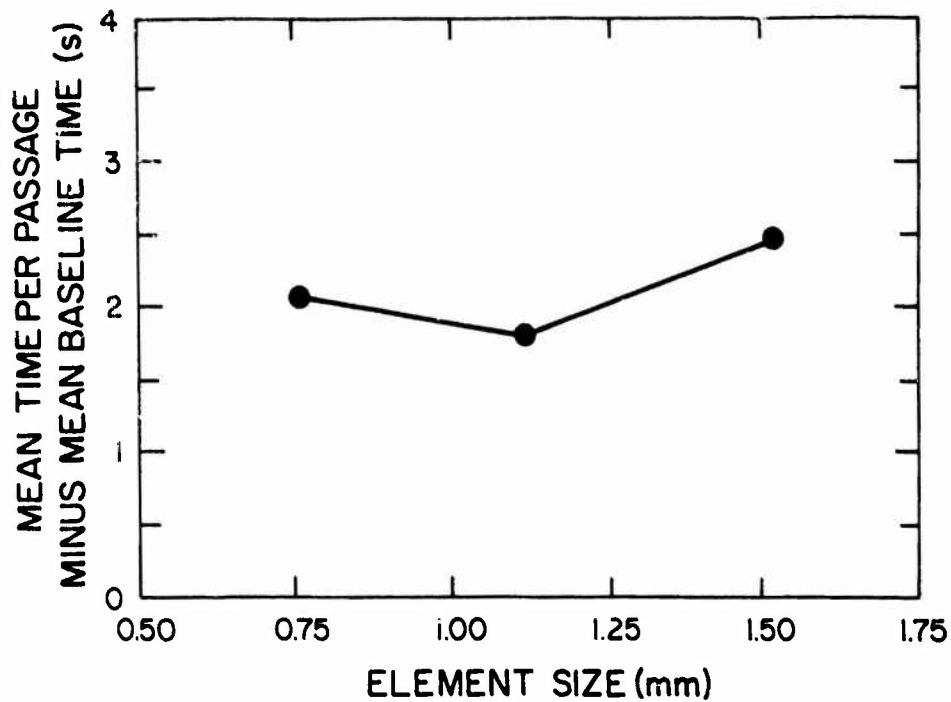
Figure 95: Effect of element size on reading time, from Snyder and Maddox (1978)
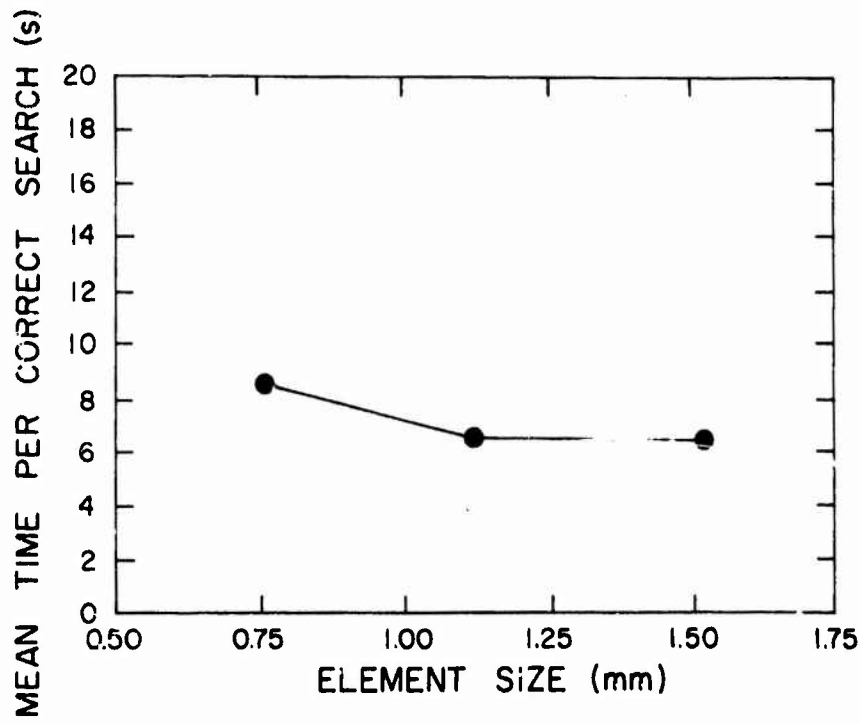
Figure 96: Effect of element size on random search time, from Snyder and Maddox (1978)

Although observer performance data have been obtained for displays which have irregular luminance distributions across the individual elements (e.g., DIGIVUE AC plasma) by Snyder and Maddox (1978), no direct comparison can be made between this display and one having uniformly luminous elements because comparable element sizes have not been studied. Thus, we have no empirical evidence on the effect of the distribution of luminance across the display element. The only means by which such an effect can be estimated is analytically, as will be discussed in Section 5.4.2.

5.1.3   Element Spacing, Continuity

Several studies have demonstrated that the closer a dot-matrix stroke or line approximates a solid line, the better the legibility and readability. Vanderkolk, Herman, and Hershberger (1975) showed that the greater the percent active area, the better the performance. More recently, Stein (1980) found that percent active area is not related to character recognition performance under good viewing conditions, but that under "stressed" conditions (decreased contrast, increased reading distance) performance fell off precipitously as percent active area fell below 45% (Figure 97). Similarly, Snyder and Maddox (1978) found a correlation between percent active area and reading time of $r = -.77$ (p < .0001), between percent active area and menu search time $r = -.53$ (p = .004), and between percent active area and random search time $r = -.30$ (p > .05).

- 292 -
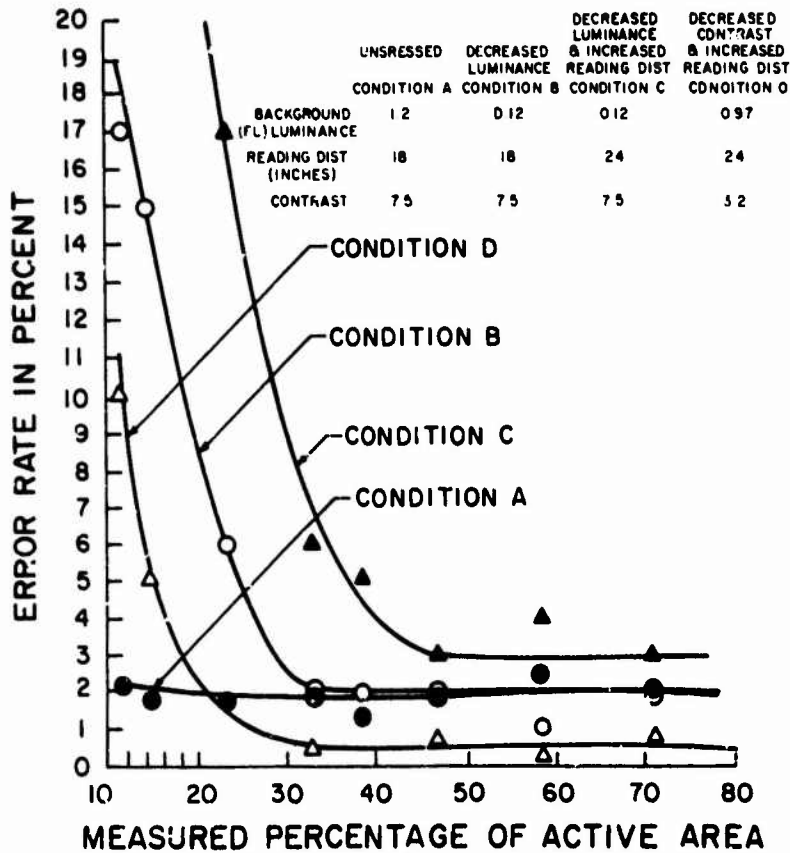
Figure 97: Effect of percent active area on character recognition, from Stein (1980)

The effect of percent active area is clearly not linear, as shown by Figure 97. In fact, it is probably best interpreted as the ratio of the element size to the between-element spacing. As this ratio increases, the line begins to appear more continuous, and the alphanumeric character is more easily read or found, as illustrated in Figure 98.

This is in keeping with what is known about raster line visibility in line-scan displays (e.g., television, see section 5.4.2) and suggests that inter-element spacing should be no more than 50% of the element width (44% active area) and preferably much less. Not all matrix element technologies currently meet that criterion.
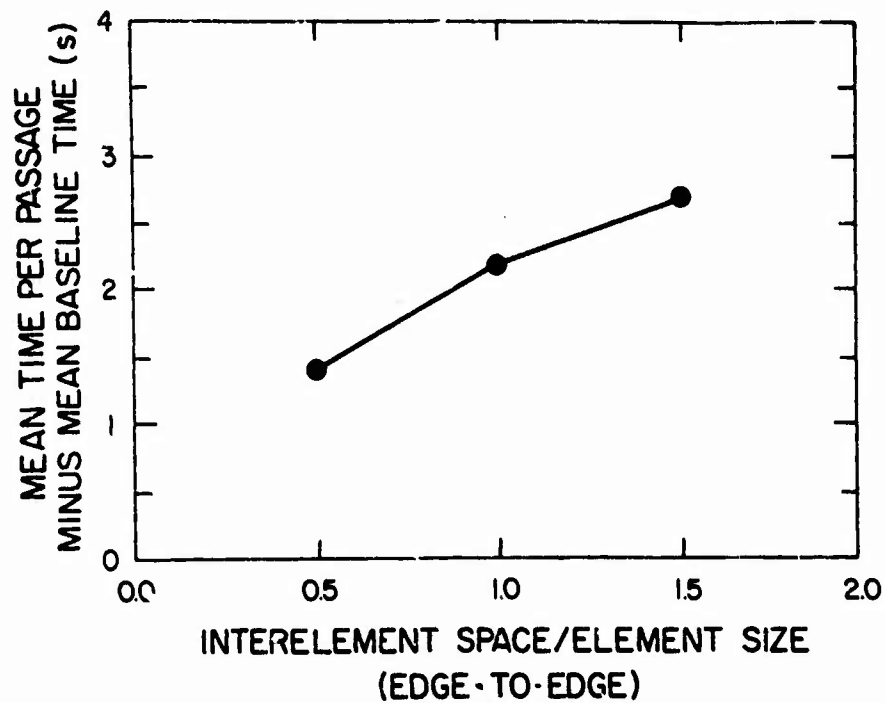
Figure 98: Effect of element size to element spacing ratio
on reading time, from Snyder and Maddox (1978)

5.1.4   Uniformity, Noise, Failure

Of the many ways in which an image can be displayed with
less than perfect fidelity, three are due to random proc-
esses uncorrelated with the commanded display content. A
display can suffer from both small- and large-area nonuni-
formity, as indicated in section 3.1.8, and these nonuni-
formities may have the effect of reducing the level of oper-
ator performance. There may also be random noise injected
into the display, which may be either correlated or uncorre-
lated with the commanded image. Lastly, cells or lines of
the display may fail, in either the "on" or "off" position,
resulting in operator performance reductions.

- 294 -

The data which indicate the effect of large-area nonuniformity on observer performance are scarce and somewhat subjective in nature. Farrell and Booth (1975, p. 3.2-60) quote two technical reports that claim "a linear drop in luminance from center to edge of a rear projection display of two thirds was tolerable" and that "a gradual brightness fall off of 50 percent will normally appear quite uniform" (p. 3.2-60). No performance data are provided, and it is implied that the referenced reports did not contain any performance data. As a result, Farrell and Booth (p. 3.2-48) recommend that luminance variation across normally used portions of the display be limited to 50 percent.

The design criteria data required for small-area nonuniformity are essentially nonexistent. While one can attempt to generalize from high spatial frequency visual thresholds, this seems particularly risky in view of the fact that many displays have aperiodic small-area nonuniformities, the nonuniformities are mixed with meaningful image content, and the nonuniformities may have a luminance level unrelated to the overall display luminance. In the complete absence of suitable data, Farrell and Booth (1975) recommend limitations of 10% across small portions of the display surface, acknowledging that their recommendation is merely an educated guess.

The effect of random noise on a display is much better understood, although this variable has been studied largely

- 295 -

for static film images and dynamic television images.
Because of the concern of noise content in the subjective
quality of and performance with television and film dis-
plays, this research had tended to translate the noise vari-
able into a relative measure of signal-to-noise ratio, which
is usually defined as:

$$S/N = \text{Peak-to-Peak Signal (volts)/Root Mean Square Noise (volts)} \qquad (20)$$

The studies in this area generally demonstrate the fol-
lowing:

1. Improvement in performance or in subjective image
   quality results from increases in S/N up to about
   35 dB;

2. Noise spatial frequencies in the range of target
   spatial frequencies have the greatest masking
   effect on the target;

3. Lower spatial frequency noise is generally more
   harmful than higher spatial frequency noise, if
   noise power is kept constant;

4. Targets can be detected at very low S/N levels, on
   the order of 2.5; and

5. Difficult tasks are more sensitive to noise than
   are easier tasks.

Unfortunately, it is difficult to generalize these results from the analog, dynamic CRT display to the same type of image on a solid-state, matrix-addressed display, such as the DIGIVUE, even if the image is refreshed at the same rate and has the same dynamic range. While one would expect the same results to obtain, the noise passband is no longer continuous because of the discrete spatial sampling of the display, the two-dimensionally discrete image, and the difficulty of photometrically measuring noise on a matrix display. As a result, while one might safely estimate that the same S/N of 35 dB would be desired to avoid noise effects upon operator performance, the frequency content effects are less certain.

It is perhaps more meaningful to think in terms of two different sources of noise for matrix-addressed displays. The first source would be display elements that do not respond properly to the commanded input level, thus exhibiting small-area nonuniformities, either static or dynamic. The subject of small-area nonuniformity was discussed previously. The second manner of looking at noise on a matrix-addressed display is to measure the incoming analog signal, prior to analog-to-digital conversion, and then represent the signal in the usual S/N form. While there is no evidence to suggest that this measure will follow the S/N generalization given above, the extrapolation seems reasonable.

- 297 -

Finally, one must be concerned with individual line and cell failures, as they are unique to matrix-addressed displays. The author knows of no research that has related line failures (on or off) to either subjective image quality or to observer performance. However, one study has investigated the effect of discrete element failure upon the legibility of 5 x 7 dot-matrix alphanumerics. Although only five letters were used in the sample set, Riley and Barbato (1978) concluded that there was no difference in legibility, using a tachistoscopic presentation, among deletion of dots, addition of dots, or a combination of the two. Increases in the total number of failed elements, of course, reduced legibility. The authors suggest that additional research on element failure is strongly needed.

5.1.5 Size, Scale

The display designer is always constrained by the available space in which to install his display, the number of elements deemed necessary on the display surface, and the scale of the displayed image. Research on these variables has been extensive over the years, often with conflicting results. In general, however, the following conclusions appear warranted:

1. For a fixed element size, increasing the display size can be beneficial because it permits the presentation of more information, at an increase in cost;

2. If the element size is scaled up with increases in
    display size, there is a maximum beyond which
    decreasing performance results; this maximum
    occurs when the individual elements become visi-
    ble. Elements are visible when they exceed the
    visual contrast sensitivity threshold, which
    depends upon spatial frequency and modulation; and

3. If displayed information is scaled up with display
    size, benefits can be obtained as the spatial fre-
    quency of the information approaches the most sen-
    sitive spatial frequency range of the visual sys-
    tem for the given viewing conditions.

This latter point is most critical to good display
design. For example, if a 10-cm wide display contains image
information which has most of its power at 20 cyc/deg, scal-
ing the display size to 25 cm will decrease the spatial fre-
quency of the image, at maximum power, to 8 cycles/deg,
which is nearer the peak of the visual contrast sensitivity
function, thus making the image content more visible. One
must be careful that the unmodulated element spatial fre-
quency, which is actually a static noise source to the vis-
ual system, does not also become proportionally more promi-
nent by this scaling. If it does, to the same degree, then
nothing is gained.

A quantitative way to evaluate the optimum display size is by considering the inverse of the contrast sensitivity function as a figure of merit. As shown in Figure 99, the optimum display size varies with the viewing distance in order to keep the target (or displayed information) spatial frequency in the most sensitive region of the "figure of merit" curve. Changes in display luminance, target spatial frequency, element spatial frequency, etc., would alter the position of these figure of merit curves consistent with the shifts in the contrast sensitivity function as described in section 4.1.
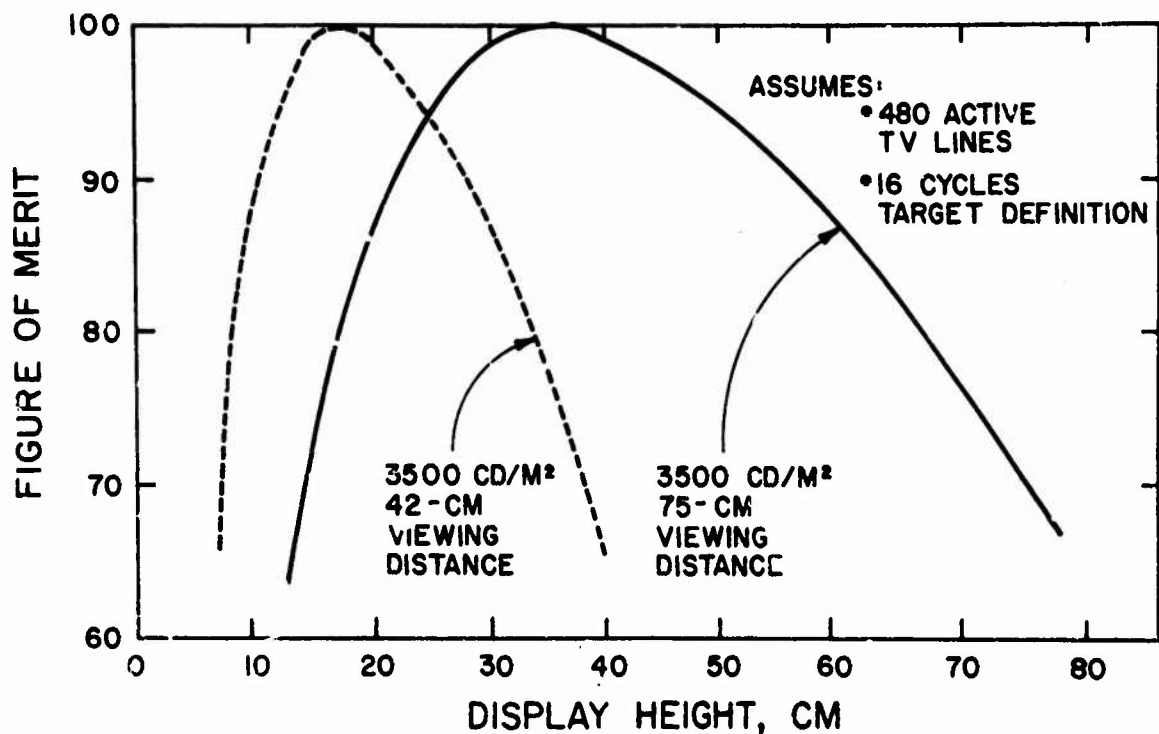
Figure 99: Figure of merit for display size for two viewing distances, from Carel, Herman, and Hershberger (1976)

5.2   TEMPORAL PARAMETERS

In Section 3.2, the temporal constrast sensitivity function of the human visual system was discussed, and it was indicated that knowledge of this transfer function adequately predicted the sensitivity of the eye to varying common nonsinusoidal pulse trains of light energy. Clearly, then, it would be desirable to be able to invoke these concepts and data to predict the frequency at which images presented on a display will fuse, rather than flicker. Because of the complexity of information that can be presented on a variety of displays, and the multitude of formats in which it can be presented, coupled with the numerous conditions of display such as phosphor, rise time, refresh rate, persistence, display size, adapting luminance levels, etc., there is little possibility of a sufficiently parametric experiment being conducted to determine the critical fusion frequency for all meaningful conditions. Thus, analytical extrapolation from existing data is necessary.

5.2.1   Rise and Fall Times

While the basic research on flicker sensitivity has been conducted with sine-wave, square-wave, or other known waveforms of stimulation using linear devices such as glow modulator tubes, the world of real displays is unfortunately less orderly. As we have seen in Section 3.2, various display devices have different rise times and fall times.   To

complicate matters further, the rates and shapes of the luminance changes (rising and falling) vary from one device to another, indeed from one phosphor to another. As a result, we again require a systematic application of theoretical data.

To evaluate the application of the temporal contrast sensitivity function, one should first examine the existing data for flicker perception on cathode-ray tube displays. In 1966, Turnage published data on the CFF for a variety of phosphors, and compared the CFF with known rise and decay characteristics of the phosphors. His experiments included ambient light conditions of 100 lux to simulate the typical display room, a sharp-edged spot of light 4 mm in diameter on the display as the target to simulate the typical alphanumeric chacters and symbols in current information displays, luminance levels from 17 to 35 cd/m^2, and contrast ratios ranging from 5:1 to 10:1.

He further chose phosphors in the medium-short to long persistence region, having high visual efficiency, such as P1, P4, P7, P12, P20, P28, and P31. Modulation of the target was sinusoidal for the first set of experiments, and then pulsed for the subsequent and final set of experiments. For the pulsed modulation, a duty cycle of 2% was selected as representative for information displays.

He then categorized the seven phosphors as having either exponential or power-law decay characteristics and deter-
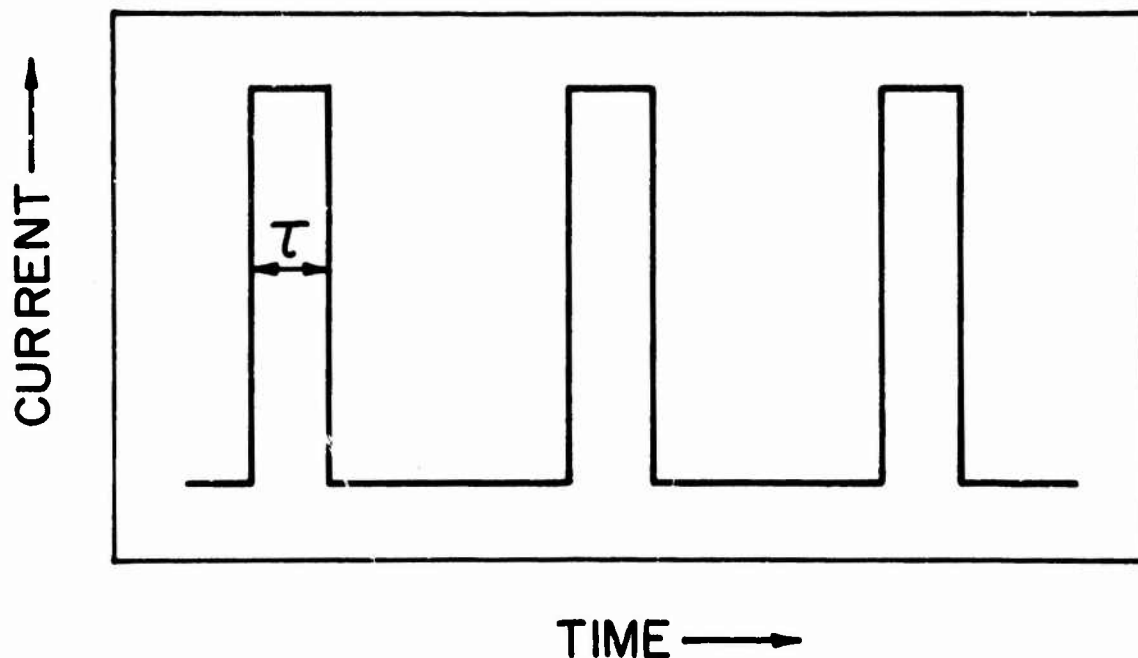
Figure 100:   Parameters of pulse train, from Turnage (1966)

mined the Fourier spectrum of each for a pulse train of the
form indicated in Figure 100.   The modulation of the funda-
mental of this pulse train was calculated for each of the
experimental conditions.   Turnage's results are given in
Figure 101, which shows the high correlation between the CFF
for sine-wave luminance variation and the equivalent modula-
tion CFF of the pulse waveform.

Thus, the Turnage data strongly suggest that one can
decermine the funcamental Fourier component for a pulse
train engergizing a CRT display, and include in that analy-
sis the decay rate of the phosphor, to predict the critical
flicker frequency of a displayed,   bright,   small   item of
information on the phosphor.   In this manner, the concept of
using Fourier analysis for prediction of the temporal char-
acteristics and perception of the display remains valid.
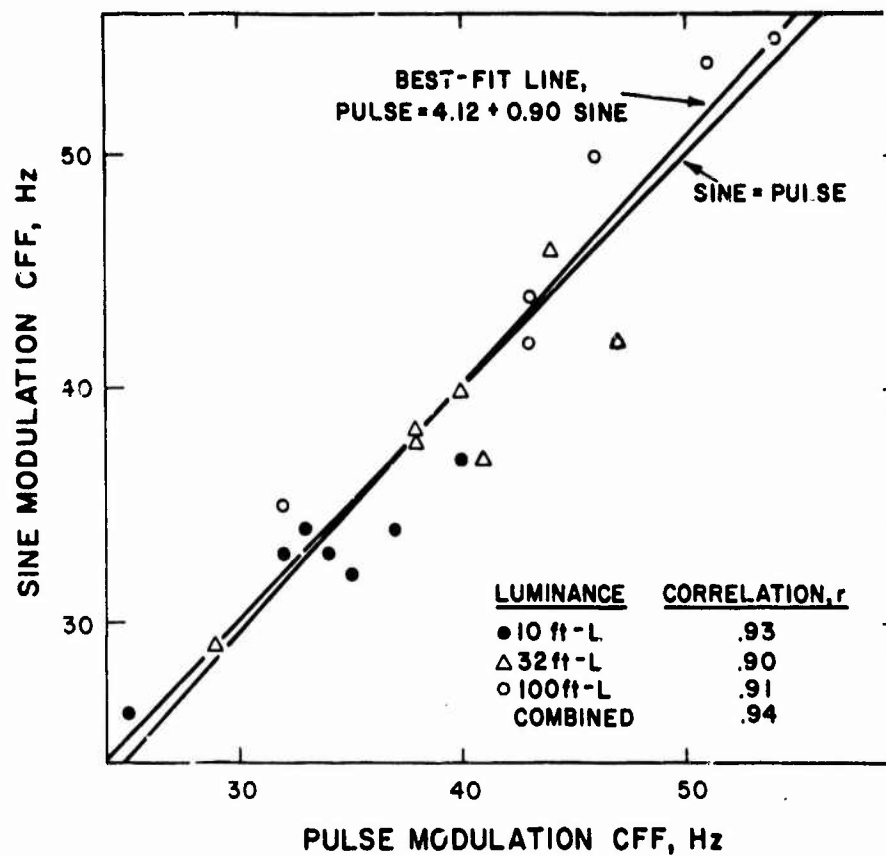
Figure 101: Pulse modulation predicted from sinusoidal thresholds, from Turnage (1966)

5.2.2 Refresh Rate

To estimate the critical flicker frequency for a given phosphor, using a given writing speed, beam current, screen characteristic, etc., one can take the information regarding that phosphor from the standard JEDEC tables of phosphor persistence time, apply to it a Fourier transform, and determine the equivalent sine-wave fundamental modulation of that phosphor at that particular pulse height. Knowing the modulation, as calculated from these parameters and the Fourier transform, one can then use the data from the temporal contrast sensitivity function to predict, with consider-

able accuracy, the refresh rate required to avoid a flicker sensation. Application of this technique has yet to be employed in a practical design situation, although the data of Turnage certainly indicate that this approach is both feasible and valid. The technique has been evaluated in a paper by Krupka and Fukui (1973), who also present a simplified calculational approach that avoids the necessity of Fourier analysis. In either the rigorous approach of Bryden (1966), or the rougher estimate of Krupka and Fukui, there seems little doubt that a direct comparison of the modulation of the Fourier fundamental, compared against temporal contrast sensitivity curves for the proper adapting luminance, will predict with good accuracy the existence of flicker versus fusion.

5.2.3 Noise Integration

The same temporal integration of the visual system that contributes to the determination of the flicker frequency can be of benefit in the visual "smoothing" of temporally fluctuating noise. Displays which are temporally noisy, although seen as not flickering, can produce a subjective impression of constant luminance to the observer even though it is possible to measure time varying luminance in the image. It is therefore important to understand visual thresholds for temporal noise integration as well as those for flicker. Some recent experiments are quite helpful in this regard.

When a TV raster-type display containing noise is viewed
with the left eye unoccluded and the right eye covered by a
neutral density filter, the noise, which looks like uncorre-
lated "snow" in normal vision, moves in an _ordered_ manner:
every noise point seems to follow an elliptical-like trajec-
tory. The plane of the ellipse is perpendicular to the
frontal plane and parallel to the floor. The points move
from right to left "in front" of the display and from left
to right "behind" the display. Changing the neutral density
filter to the other eye changes the direction of the move-
ment; increasing the density of the neutral density filter
increases the apparent depth of the movement. In short, the
perception is similar to that of a classical Pulfrich ster-
eophenomenon.

This phenomenon is probably caused by a combination of
space and time integration which is different for each eye
due to the change in retinal illuminance caused by the neu-
tral density filter. Due to the difference in time integra-
tion (caused by the neutral density filter), the same
points, when presented to the two eyes with small dispari-
ties, create the sensation of depth as in the Pulfrich phe-
nomenon.

This result shows that a temporal delay is equivalent to
the eye-brain system to a change in integration time caused
by a variation in the mean luminance seen by the eye. This
phenomenon can therefore be used as a measuring tool for the

dynamics of the integration time and as partial verification of a visual integration model proposed by Almagor et al. which can describe the whole range of eye-brain performance in the temporal domain. The model, illustrated in Figure 102 and described in greater detail by Almagor et al. (1979), asserts that integration time (Δt) is locally controlled at the retina, and is determined by the mean retinal illuminance level. It then predicts that the contrast sensitivity, to a time-varying stimulus, of observers characterized by a long integration time will be greater than the sensitivity of observers characterized by a short integration time at the same mean luminance. That is, observers with a longer integration time can detect a smaller amplitude variation.

Another prediction of the model is that a large area trapezoidal time-varying luminance stimulus will be perceived as having temporal "bands" similar to the spatial Mach bands, and that the stimulus will have a lower "undershoot" and a higher "overshoot" (Figure 103). The upper band should be shorter and brighter appearing than the lower band, partially due to the hypothesized control of integration time by the time integral of luminance of the stimulus, and partially due to the hypothesized memory cell (Figure 102) working as a differentiator.

At the "upper band" point of Figure 103, the time integral is set by the integral of the incoming stimulus which,
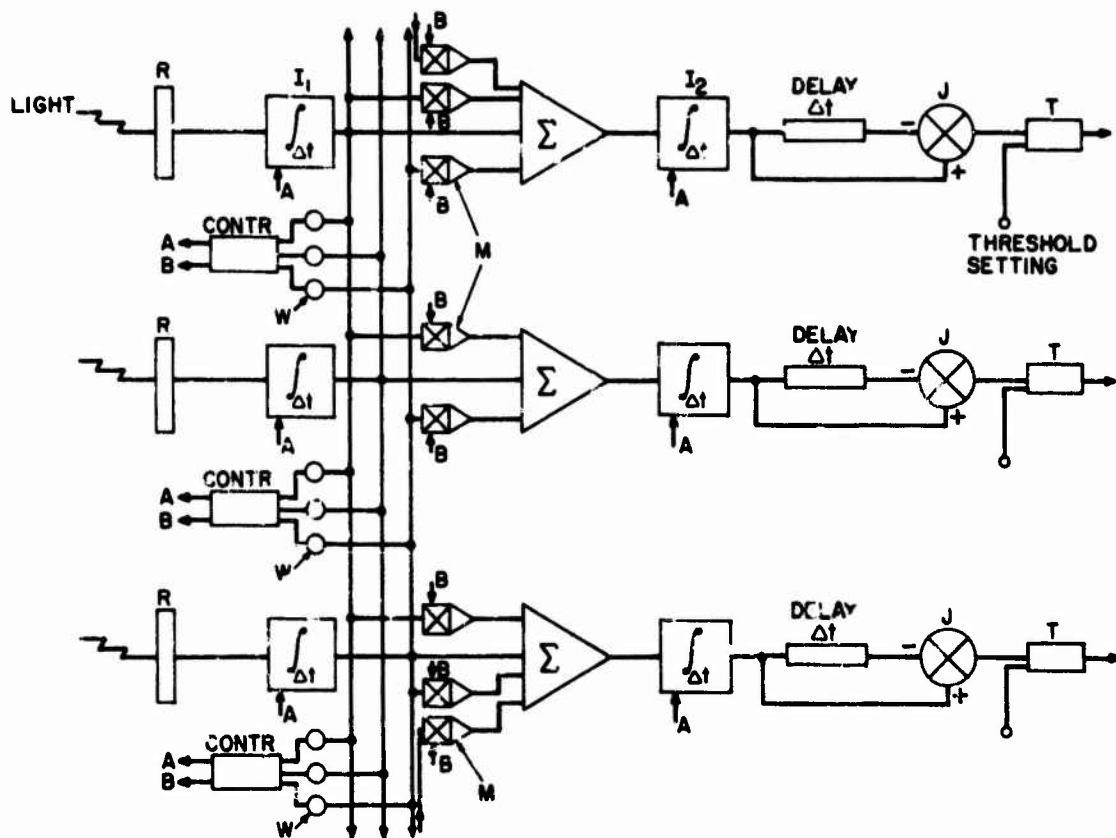
Figure 102: Temporal integration model proposed by Almagor et al. (1979)

at this point, is lower (meaning a longer integration time) than it will be along the subsequent bright constant part of the stimulus, thus creating a brighter sensation. Conversely, at the "lower band" point, the time integral is set by an integrated luminance higher than it will be subsequently during the dark constant part of the stimulus, resulting in a shorter integration time and creating a darker sensation. The temporal bands are accentuated by the second integrator (Figure 102) and by the fact that not only the time integral, but also the spatial summation, varies in the same way.

- 308 -
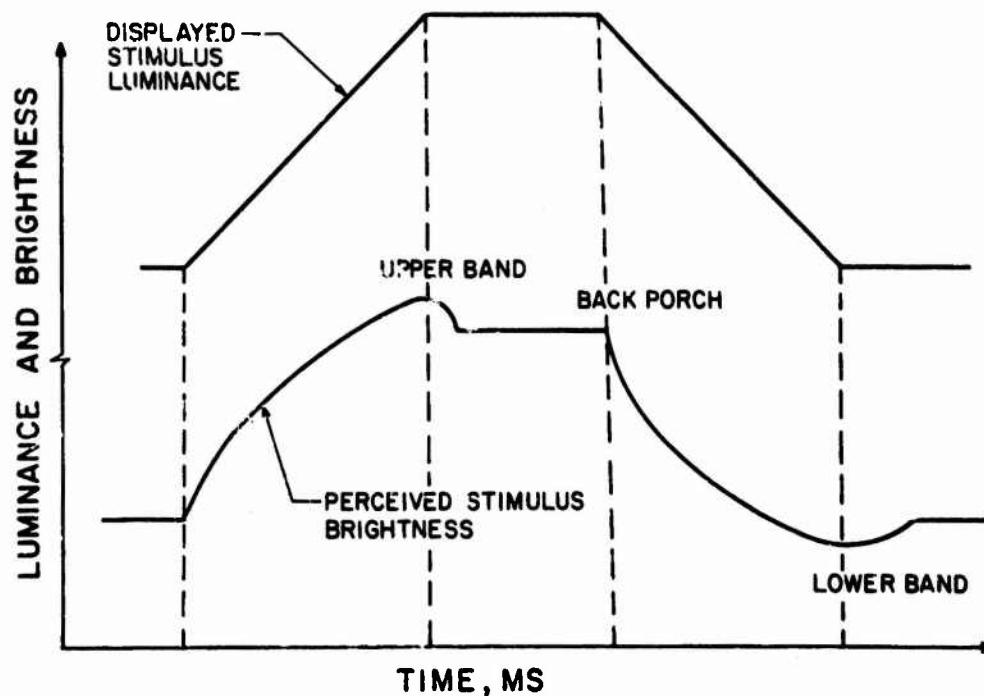
Figure 103: Comparison of physical stimulus and brightness, from Almagor et al. (1979)

Predictions of this model were tested in several experiments (Almagor et al., 1979), all of which essentially confirmed the predictions. The temporal bands were measured for three different luminance slopes, three points on the trapezoidal waveform, and two mean luminance levels. Similarly, flicker modulation sensitivity curves were obtained for two different retinal illuminance levels (77 and 680 trolands) for frequencies ranging from 0.01 to 70 Hz.

The length of the temporal bands decreased with increasing luminance, indicating a negative correlation between integration time and retinal illuminance.

The flicker sensitivity curves show that as the mean luminance increases, the peak sensitivity increases (minimum

modulation decreases) and the curves are shifted towards higher frequencies (e.g., Figure 104). These results agree perfectly with those of Kelly, de Lange, and others. There appears to be a correlation between the peak frequency (integration time) and sensitivity at that peak frequency across subjects for the same experimental condition. That is, subjects having a lower frequency peak (longer integration time) are more sensitive than subjects showing a shorter integration time (higher frequency peak). The peaks of the curves lie in the range of 6 Hz to 18 Hz which, in the proposed model, convert to integration times between approximately 84 ms and 30 ms (e.g., a frequency of 6 Hz has a period of approximately 168 ms; the integration time equals half the period or 84 ms).

A practical application of these results is the optimization of the conditions under which information should be displayed. Images of moving targets, with little or no noise, should be seen in bright light in order to keep the integration time and space at a minimum. For example, dynamic cockpit displays ought to be made as bright as possible.

On the other hand, images embedded in uncorrelated noise (low light level television, for example) should be viewed in a very dim environment to increase the integration time and thereby filter out more noise. Because the eye-brain system trades off integration time for integration space,
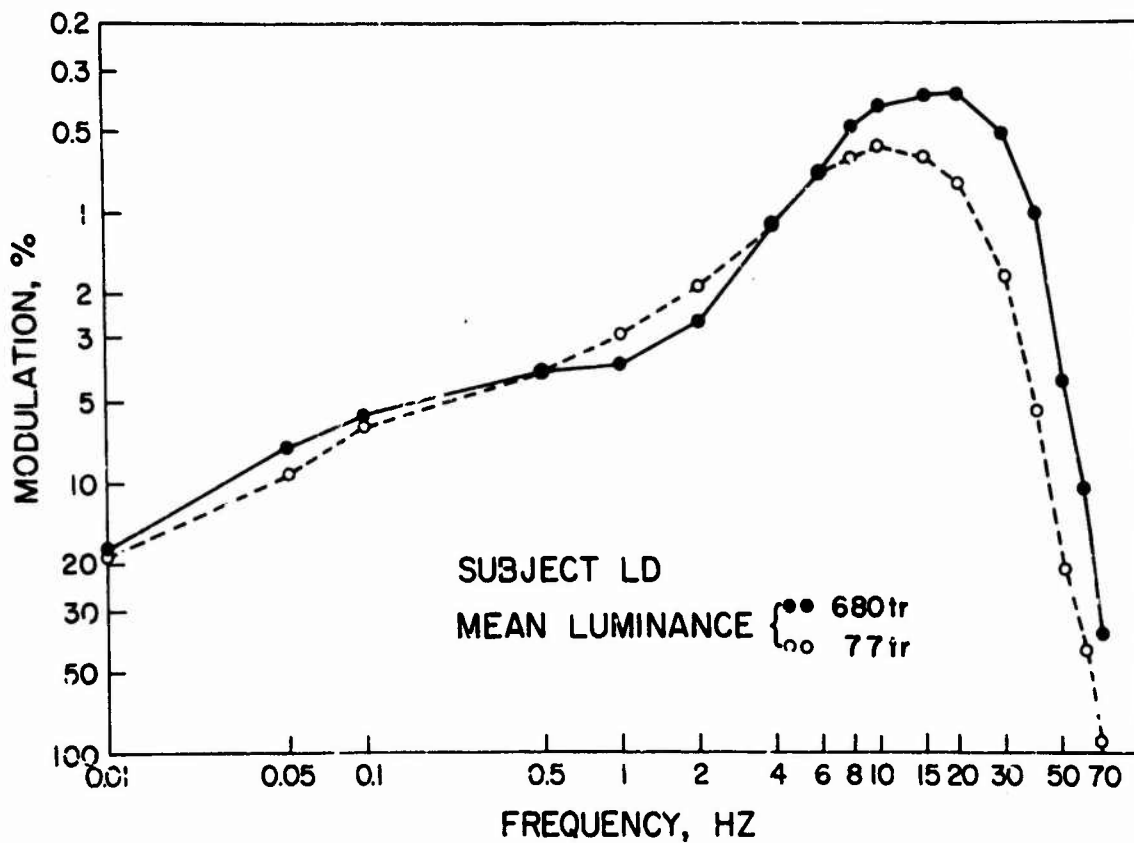
Figure 104: Sinusoidal flicker sensitivity curve for Subject LD, from Almagor et al. (1979)

the acuity remains high while the perceived noise is reduced through increased temporal integration.

Several experiments have demonstrated directly the effect of dynamic display noise on visual performance, the most parametric and complete experiment being that of Keesee (1976). He indicated how the contrast sensitivity function is elevated by dynamic noise for raster scan displays. Further, his results show that low spatial frequency noise produces far more interference than does high spatial frequency noise. These results were discussed previously.

- 311 -

The key issue from this research is, then, how do we use information on visual integration times to evaluate display requirements and noise levels? What is clearly needed is research to relate measured observer integration times to visual performance under known (photometrically measured) displayed noise conditions. That research has not yet been conducted, although several metrics of image quality take displayed noise into account in predicting observer performance. These metrics will be discussed subsequently.

5.3 CHROMATIC PARAMETERS

Chromatic displays are generally used for three reasons. First, many system and display designers, and many users/purchasers, feel that displays of pictorial information should appear "realistic" or "natural." Thus, color television is greatly preferred to achromatic ("black-and-white") television because we generally view our pictorial world in full color, and the color television display more closely approximates that perception. We tend to disregard or be unaware of the poorer image quality in color television more than in achromatic television, and also remain unaware of the limited range of colors produced by the CRT phosphors. This first reason is largely aesthetic, rather than performance related.

The second reason for chromatic displays is that the chosen technology exists only in a given wavelength range.

Thus, we often use monochrome displays (e.g., LED, gas discharge, EL) in place of achromatic displays when no chromatic information discrimination or content is needed or required.

The third, and very important, reason for chromatic displays is to take advantage of color coding. That is, by selectively presenting different types or categories of information in perceptually distinctive colors, search performance of the observer for a specific information category can be enhanced. Much research has been dedicated to this topics, and some general "rules of thumb" have emerged. As we shall see, however, quantitative criteria for color coding and for estimating the efficacy of color coding are essentially nonexistent. Nevertheless, applications of color coding abound in present-day systems, ranging from colored graphics overlays, to multicolor alphanumeric displays to weather radar displays, in which storm intensity is coded in pseudocolors rather than in the more traditional luminance levels.

Limitations to our knowledge of quantitative color coding utility and criteria appear to stem from (1) lack of rigorous quantitative measurement and understanding of the nature of tne chromatic stimuli in most of the published experiments, and (2) some adaptive, nonlinear characteristics of the visual system which lead to situation-specific conclusions. In this section, we shall summarize very briefly the

fundamental requirements for display color coding, and indicate the problems in application of our knowledge of color vision to color display design.

5.3.1 Intrinsic Chromatic Contrast

While one would expect that adequate criteria exist for optimizing, or at least quantitatively specifying, chromatic color coding, such is not the case. As Krebs, Wolf, and Sandvig (1978, p. 1) stated,

> A careful review and analysis of the color litera-
> ture reveals that the issue of color utility is
> not a simple one. The value of color as a coding
> method is entirely dependent on its effective use
> in a specific application. That is, it can be
> beneficial, neutral, or distracting

Following their review, Krebs et al. (1978) proposed 18 steps to be followed in the design of a color display for any application. Pertinent to the present section is Step 5, which states:

> Of the remaining colors available, select up to
> five (maximum). The particular colors chosen
> should be widely spaced in wavelength from one
> another (p. 4).

Blue should be avoided because it leads to poorer legibility (Myers, 1967), and red, yellow, and green should be reserved for "danger," "caution," and "safe," respectively.

At a more detailed level, one can refer to several experiments which evaluated the effects of chromatic characteristics upon observer performance. For example, Pollack (1968) determined the response time to warning lights of

wavelengths ranging from 415 to 657 nm. His results, illustrated in Figure 105, show that response time is shortest to the blue to bluish-green part of the spectrum, and longest to the red end for 2.1-deg lights at luminances below about 0.2 cd/m^2. The dimmer lights get into the mesopic (0.00004 to 0.04 cd/m^2) region, the more critical the light wavelength becomes. Above about 0.2 cd/m^2, and to over 1000 cd/m^2, there are no significant differences among wavelengths. These data apply to the given wavelength signal against a black background. For this size signal, the symbols were not seen as chromatic below approximately 0.1 cd/m^2, the luminance below which wavelength begins to have an effect upon response time.

A somewhat different result was obtained by Tyte, Wharf, and Ellis (1975) under a high ambient illuminance of 10^5 lm/m^2. Their results, illustrated in Figure 106, indicate that greenish-yellow (575 nm) lights yield the longest response times, and that response times are shortest at either end of the spectrum. Under this high ambient illuminance, red warning lights are more effective than green or yellow, especially at lower stimulus luminance levels, as indicated in Figure 107.

Table 18 compares the relative legibility of seven colors for alphanumeric reading. As indicated, there is good agreement between the studies, with blue and green generally poorest, and red and yellow best.

- 315 -

MEDIAN RESPONSE TIME, MS

500

(657 nm)

400

(589 nm)
(530 nm)
(415 nm)
(505 nm)
(462 nm)

300

200

LOW AMBIENT ILLUMINATION
SIGNAL SIZE : 2.1 DEG

0   -3   -2   -1   0   1   2   3

SIGNAL LUMINANCE, log cd/m²
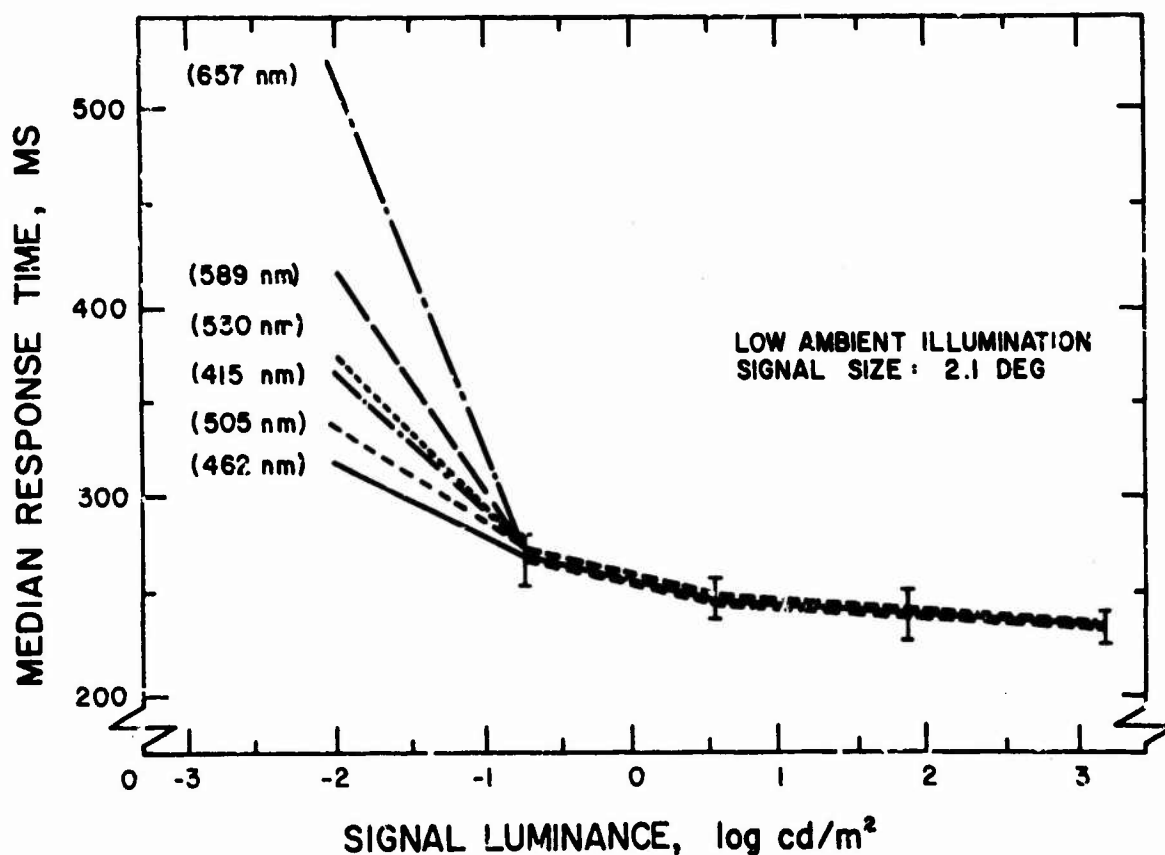
Figure 105:   Median response time as a function of symbol
luminance, from Pollack (1968)

One can cite additional studies of this type with little
additional understanding. It should be noted carefully that
nearly all such studies use (1) a stimulus described by
either a dominant wavelength or a subjective color label,
(2) no measurement of saturation or purity of the stimulus,
and (3) a black or achromatic background. That is, there is
no quantitative radiometric specification of the stimulus
and, most importantly, there are no quantitative data known
to relate performance to the intrinsic chromatic contrast
between one chromatic stimulus and its (different) chromatic
background. Where the potential for such data has existed

- 316 -
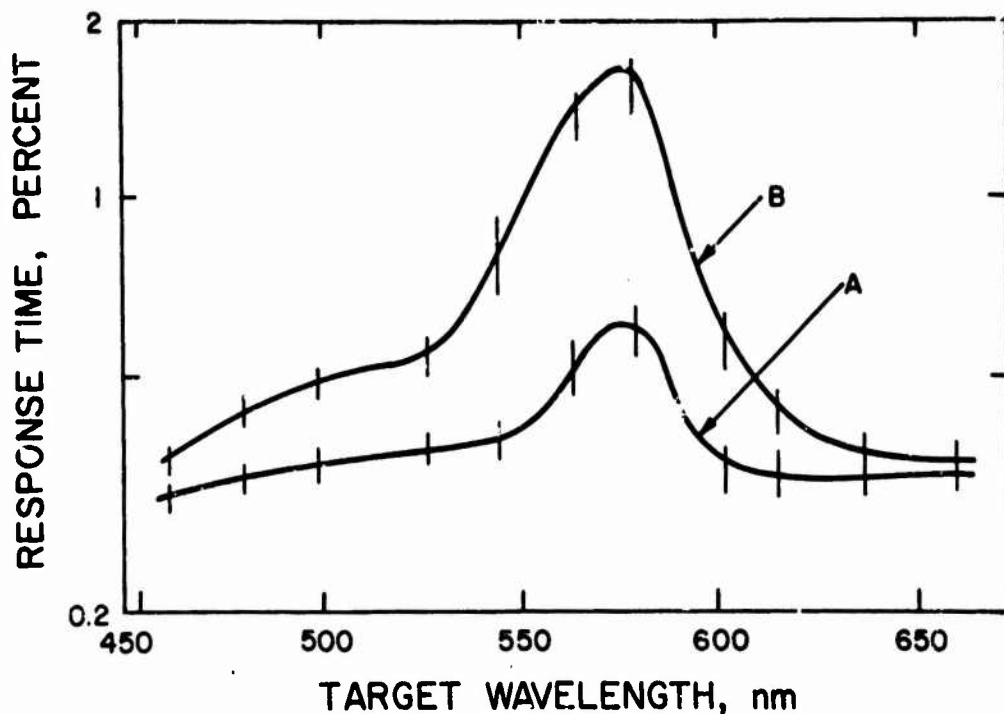
Figure 106: Response times for various wavelengths under high ambient illuminance, (10^5 lm/m^2) for lights of 48 cd/m^2 (A) and 30 cd/m^2 (B), from Tyte et al. (1975)
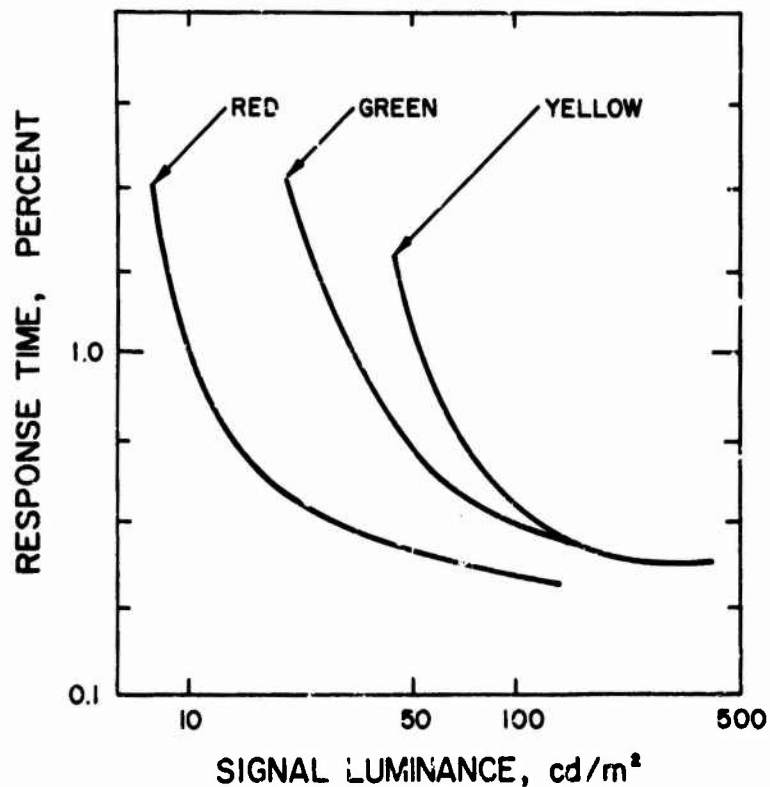
Figure 107: Response times to red, yellow, and green stimuli under high ambient illuminance, from Tyte et al. (1975)

TABLE 18

Relative Ranking of Different Colors in Alphanumeric
Recognition Studies (lowest ranking indicates best
performance)

Color	Meister and Sullivan (1969)	Rizy (1967)
Red	2	1
Yellow	1	2
Magenta	4	3
White	3	4
Cyan	5	5
Blue	7	6
Green	6	7

(e.g., McLean, 1965), the only "contrast" measure made for
the various color stimuli was that of luminance contrast.
Thus, it is not possible to describe the chromatic contrast
in such experiments in any conventional color space system.

Specifications and recommendations exist, however, for
discrete colors independent of their backgrounds. Military
Specification MIL-C-25050A provides CIE chromaticity speci-
fications for aviation white, aviation red, aviation yellow,
aviation green, and aviation blue (Figure 108). Because
these are defined in CIE space, both dominant wavelength and
purity are known. Similarly a recommended set of 10-color
coding dominant wavelengths has been given by Baker and

Grether (1954), as indicated in Table 19. Cook (1974) provided a six-color code including both Munsell and CIE coordinates (Table 20). Cook's X,Y coordinates are plotted in Figure 109, which indicates that the red, yellow, orange, and green are highly saturated colors, but that the purple and blue are only about 40% and 65% saturated, respectively. In fact, the "purple" (x = 0.2884, y = 0.2213) would be considered "white" by many observers.
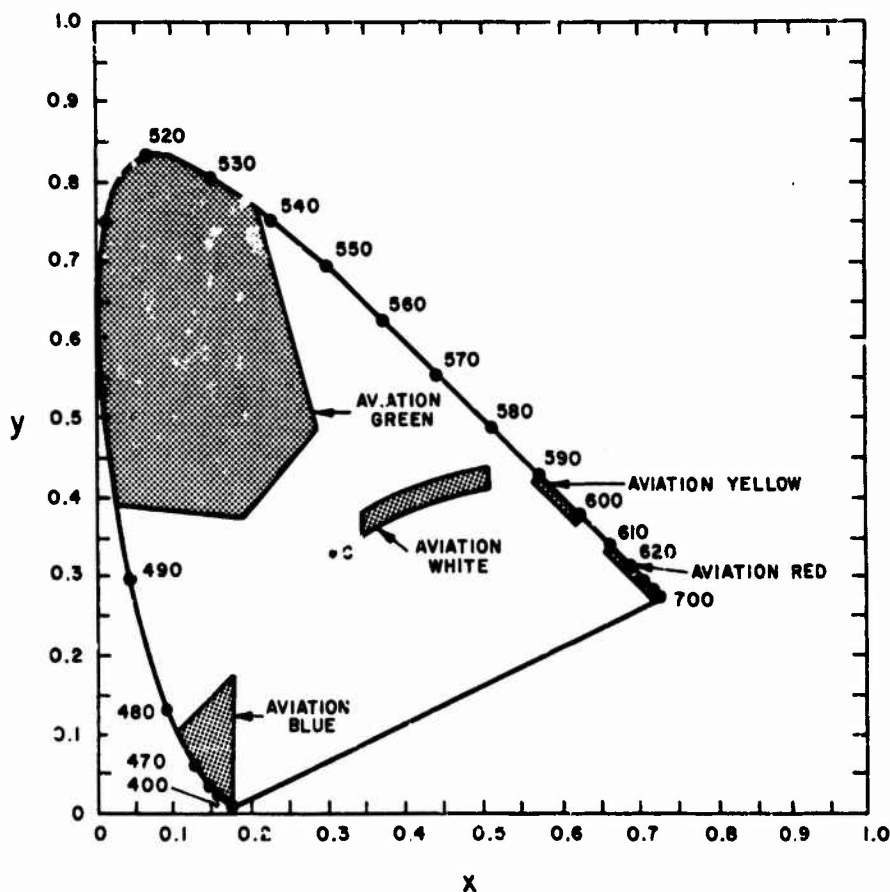
Figure 108: CIE coordinates for aviation white, aviation red, aviation yellow, aviation green, and aviation blue

TABLE 19

Ten Recommended Colors That Can Be Identified Correctly
Nearly 100% of the Time, from Baker and Grether (1954)

Dominant Wavelength (nm)	Color Name
430	Violet
476	Blue
494	Greenish-Blue
504	Bluish-Green
515	Green
556	Yellow-Green
582	Yellow
596	Orange
610	Orange-Red
642	Red

In summary, there are essentially no data to relate operator performance to the _color_ contrast between two chromatic stimuli. While we can reasonably predict reaction time, search time, and the like to different colors against an achromatic background, we cannot estimate performance for one color against a background of another color, even if the two colors are of the same luminance. If they are of different luminances, we have even less understanding of the effective visual contrast, and of how the effective color contrast should be measured.

TABLE 20

Recommended Colors for a Six-Color Code, from Cook (1974)

Color Name	Munsell Notation	CIE Coordinates		Dominant Wavelength(nm)
		$\underline{X}$	$\underline{Y}$	
Purple	1.0 RP 4/19	0.2884	0.2213	430
Blue	2.5 PB 4/10	0.1922	0.1672	476
Green	5.0 G 5/8	0.0389	0.8120	515
Yellow	5.0 Y 8/12	0.5070	0.4613	582
Orange	2.5 YR 6/14	0.6018	0.3860	610
Red	5.0 R 4/14	0.6414	0.3151	642

Finally, we should note that even colors against an achromatic background can vary in purity (or saturation). While it is probably the case that more highly saturated colors in some parts of the spectrum produce better performance, most likely less saturated colors are better in other parts of the spectrum. This can be seen from Table 18, using the logic that a very unsaturated green, as it approaches white, would increase in legibility. Or, as red becomes less saturated, approaching white, it would yield reduced legibility. Because of the lack of radiometric measurement in nearly all pertinent human performance experiments, such generalizations are not safe, but merely indicative of our lack of knowledge.
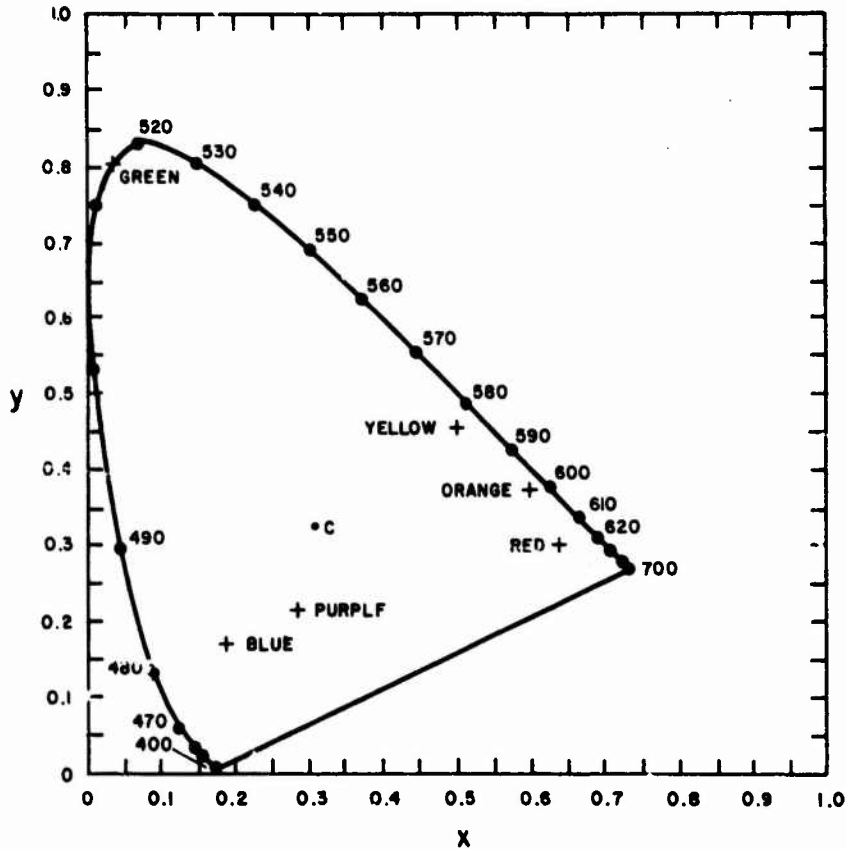
Figure 109: Cook's (1974) coordinates in CIE space

Even after a thorough literature review, Krebs et al.

(1978, pp. 43-44) cautiously concluded only that

> Saturation differences . . . produce many color variations on maps. . . . Hue-saturation combinations can provide a large number of discriminable different values for the color code. Caution should be taken to ensure that the changes in saturation do not produce colors that are difficult to see under some viewing conditions.

5.3.2 Ambient Effects

The effects of the quality and quantity of ambient illuminance can be very critical to display legibility. When ambient illuminance is low, it is relatively easy to achieve

high legibility displays, subject to the constraints discussed previously. However, when the ambient illuminance is high or intensely chromatic, reduced legibility can result. Figure 107 indicated the impact of white ambient illuminance on response times for lights of various wavelengths. In general, when the stimulus drops below about 3 cd/m^2 in luminance, a loss of color coding capability will result under zero ambient illuminance. Under substantial ambient illuminance, the loss of perceived color coding will occur at much higher luminance levels, depending on the chromaticity of the ambient.

The literature is totally inadequate in predicting the effects of combinations of illuminance and chromaticity of the illuminance on the legibility (or any other performance measure) of chromatic displays. However, some subjective estimates of the effect of the chromaticity of the ambient illuminance are given in Table 21 for different display colors. The "effect" is stated only in terms of the perceived color of the display as it is altered by the ambient. Any change in legibility as a result of the ambient is totally unknown. It should be noted, once again, that the chromaticity (hue and saturation) of the ambient and the display are not given, and thus no quantitative generalization would be possible to any precise situation.

Stated simply, we are currently unable to predict the effect of any given combination of illuminance level, illu-

TABLE 21. Perceived Effect of Colored Illuminance on Colored Display Elements, from Semple, Heapy, Conway, and Burnette (1971)

Displayed Color	Ambient Illuminance			
	Red	Blue	Green	Yellow
White	Light pink	Very light blue	Very light green	Very light yellow
Black	Reddish black	Blue black	Greenish black	Orange black
Red	Brilliant red	Dark bluish red	Yellowish red	Bright red
Light blue	Reddish blue	Bright blue	Greenish blue	Light reddish blue
Dark blue	Dark reddish blue	Brilliant blue	Dark greenish blue	Light reddish purple
Green	Olive green	Green blue	Brilliant green	Yellow green
Yellow	Red orange	Light reddish brown	Light greenish blue	Brilliant light orange
Brown	Brown red	Bluish brown	Dark olive green	Brownish yellow

minance hue, and illuminance purity on any type of observer
performance using a display with known luminance and chromi-
nance characteristics. Because of the rapid development of
a variety of chromatic displays, however, such data are
urgently needed.

5.3.3   Chromatic Adaptation

Many of the "color contrast" studies in the literature
have investigated the effects obtained when a chromatic
stimulus is presented with a differently colored surround or
"inducing field" (Yund and Armington, 1975). The results of
such studies show that, for example, a yellow-green stimulus
presented with a yellow surround appears more green than
when the central stimulus is presented alone or against an
achromatic background.

The magnitude of the effect of the surround on the target
increases with increasing surround size (Yund and Armington,
1975), decreases as the spatial separation between the stim-
ulus and the surround increases (Oyama and Hsia, 1966), and
the greatest effects occur with small test fields and large
surrounds (Marsden, 1969). The results of these experiments
are of some interest to display designers: they indicate
that the perceived hue of a color character or symbol is
affected by retinal adaptation to the color of the surround
and the surround size. The problem is that these results
are not related to the perceived contrast between a target

and its background, nor have they been related to typical observer performance measures. For these reasons, they cannot be used directly in the design of displays.

Experimentation is needed on the effects of chromatic adaptation on observer performance with chromatic displays. Not only is there a potentially adverse effect from such chromatic adaptation, but there is also a potentially beneficial effect. For example, Frost (1964) indicated that chromatic adaptation caused by a large surround in an electroluminescent panel could increase the number of distinguishable colors within the panel by creating "Fechner colors."

5.3.4   Chrominance/Luminance Tradeoffs

There has been little research on how chrominance contrast, the contrast between two stimuli of differing hue and/or saturation, and luminance contrast combine into perceived color contrast. Further, there have been very few attempts to determine the form of a metric of color contrast. Chapanis (1949) suggested that chrominance contrast and luminance contrast combine to produce a total contrast, but he failed to offer a definition or means of calculation of chrominance contrast.

Further, very little research is available to contribute to the formation of hypotheses of how chrominance separation contributes to the overall perception of color contrast.

Some research has followed the lead of MacAdam (1949) in determining just noticeable differences among colors. This approach addresses the complement of the visual display color contrast problem; it attempts to specify the amount of chrominance separation permitted before a subject detects a difference between a reference stimulus and a test stimulus. In essence, this research determines the threshold for color contrast achieved by chrominance separation. The results of these studies are of considerable use to the paint and dye industry in specifying quality control criteria, but the results do not enable the display designer to specify particular levels of suprathreshold contrast.

The results obtained by MacAdam and other investigators indicate that the CIE and CIE-UCS color spaces are not uniform with respect to chromaticness or lightness (brightness) (Wysecki and Stiles, 1967). Thus, equal changes in color coordinates do not result in equal changes in perception. Efforts to develop uniform color spaces or to transform existing color spaces into uniform color spaces have met with little success or require complicated transformations (Wysecki and Stiles, 1967, ch. 6) and are often of limited generality.

A further obstacle to the formation of a metric of total perceived color contrast is the lack of correlation between luminance and perceived lightness, subsumed under the label of "luminance additivity." A typical experiment demonstrat-

ing nonadditivity was performed by Guth (1967), who measured the luminance added to a monochromatic test field to make the test field detectable, or just over threshold. If luminance additivity holds, then for monochromatic fields of different wavelengths a constant amount of luminance has to be added to make the test field detectable. The results reported by Guth for monochromatic fields of 10 different wavelengths indicate that luminance addivity does not hold. The results further indicate that the larger the wavelength difference between the subthreshold test field and the added light, the greater the luminance of the added light required to make the compound stimulus detectable.

Experiments investigating luminance nonadditivity have generally been conducted using stimuli at or near threshold. It is important to establish how these results may be generalized to suprathreshold stimuli and the perception of contrast, for suprathreshold contrast is much more pertinent to display design. Studies which would permit such generalization have not been reported.

The nonuniformity of the CIE color space and color spaces derived from the CIE color space in both chromaticness and brightness (or lightness) limits the utility of these systems in predicting perceived color contrast. If these color spaces were uniform then it might be possible to scale, by a constant multiplier, the difference between the coordinates of two stimuli in order to predict the perceived contrast.

The nonuniformity of the CIE and CIE transformed color spaces indicates that equal changes in coordinates do not yield equal changes in perception; thus, no simple function of the difference between coordinates can be expected to correlate highly with perceived contrast.

A rather severe limitation of the psychophysical research on the perception of chromatic stimuli stems from the limited range of stimuli which have been studied. One such limitation is a result of the use of narrowband (monochromatic) stimuli. Although the use of such stimuli simplifies, to some extent, the analysis of results and the formulation of theory, it limits the generality of the results obtained. This is especially true since the emitted spectra of most color displays are quite broad-band.

Another limitation stems from the diversity of experimental paradigms in use and from the hesitancy of investigators to study two or more paradigms in parallel. As a result, there is a great deal of very basic research in the literature, very little of which has been extended beyond a single experimental paradigm. This strategy simplifies the formation of theory and leads to well defined research questions, but it limits the generality of the results. Therefore, much of the existing literature cannot be used to determine the form of a metric of color contrast. Further, it is not clear how the various phenomena being studied might affect the perception of color contrast.

In summary, neither the human factors literature nor the psychological and psychophysical research have provided the data or theory necessary to form a metric of color contrast. Due to very large nonuniformities, neither the CIE nor the CIE-derived color coordinate systems can be simply and efficiently adapted to predict the magnitude or utility of perceived color contrast.

5.4 UNITARY METRICS OF IMAGE QUALITY

There have been literally hundreds of research studies which have attempted to predict the effects of various display parameters upon display quality. Many of these studies have emphasized subjective, or perceived, image quality, while countless others have measured objective forms of visual performance, such as object detection or recognition. While a few studies have attempted to investigate the interaction effects of several display parameters, many have been content to evaluate the effect of only one variable. As Snyder (1973) noted, empirical research designed to investigate all possible and pertinent design variable interactions is an impossibility. Therefore, it is necessary to adopt a research strategy that mixes both empirical data and mathematical modeling approaches to arrive at a useful prediction of the effects of the numerous combinations of design variables.

This section summarizes the pertinent existing models, theories, and data to assess the present status of that research strategy. That is, it summarizes the present level of knowledge on "unitary metrics" of image quality, those metrics which attempt to account for most, if not all, of the display design variables which influence subjective image quality or observer information extraction performance, or both.

Because various types of displays have different geometric contraints, it is convenient to partition the pertinent research, theory, and models into spatially continuous and spatially discrete forms. Spatially continuous displays are those which have continuous sampling in both dimensions, and are not broken by artificial, non-information bearing borders or edges. The most useful and typical example of the spatially continuous display is the photographic image; indeed, much of the research of interest and use has been done on photographic images. Nonraster CRT displays also fit this category.

The spatially discrete display has artificial lines or edges between information-bearing image elements. Examples of the spatially discrete display include all dot matrix displays having separate XY cells.

Prior to discussing spatially discrete display image quality, we shall briefly evaluate image quality concepts for both continuous-image displays and for hybrid displays,

those having one continuous image dimension and one discrete image dimension. Of course, the monochrome television display is a prime example of the hybrid display. While the continuous-image and hybrid-image metrics do not direcly apply to the evaluation of flat-panel discrete displays, they form the basis of nearly all quality metric concepts and therefore will be discussed very briefly.

The last type of display to be discussed in image quality metric terms is the chromatic display, either continuous or discrete.

5.4.1   Spatially Continuous Monochrome Displays

The spatially continuous monochrome display is typified by the photographic image, either in transparent film form or in hard copy paper print. There has been extensive research on photographic image quality and most other image quality metrics derive from this research.

There are many parameters of a photographic imaging system which can be critical to the resulting quality of the image. Taking only the simple photographic case, and excluding for the present time any projection device, it is clear that the exposure of the film bears a critical relationship to the modulation transfer function (MTF); the taking lens and its point spread function also have a bearing on the MTF; the interaction of the lens with the diaphragm or iris permits a tradeoff among depth of field, lens aber-

ration, and limiting diffraction characteristics for the MTF; the film spread function has a direct bearing on the MTF; and any motion of the object or of the image plane relative to the object will also affect the MTF. In addition, of course, one must take into account the overall density versus exposure (D log E) curve of the film, as it is related to the developer used, the development time, the film base, the graininess of the film, etc. Thus, we see a very large number of parameters which can be traded off against one another, with a reasonably large effect upon the overall MTF of the final image. If, of course, one then becomes interested in using the film transparency for either printing into hard copy on paper or film form, or if one is interested in projecting this film image onto a screen, then additional MTFs of the paper, paper developer, projection lens, screen, etc., also become pertinent.

For these reasons, many scientists and engineers have been concerned over the past 40 years with the concept of photographic image quality and its means of measurement. The ultimate desire has been to derive a singular measure of image quality which takes into account all of the important above parameters, and permits the designer to trade one against the other for a final optimized design for any given purpose.

The pertinent literature can be further divided into two fundamental relationships involved in image quality. First

is the bulk of the literature and techniques which are concerned with physical measurement and specification of image quality, while the second category relates these physical measures to the ability of the human to obtain information from the image. The ultimate proof of any measure of image quality is not in whether or not one can measure it, but rather in how it relates to the ability of the observer to gather the required information from the image. If, in fact, this relationship between physically measured image quality and observer performance is very low, then the physical measure of image quality can be considered invalid for the task at hand.

Perhaps the best-stated set of criteria for the inclusion of various elements in a physical measure of image quality, and in the related behavioral problems, has been given by Charman and Olin (1965, p. 385), who proposed that such a measure include at least the following:

1. The general blurring of the image of each object point caused by the cumulative effect of the various stages of the atmosphere-camera emulsion-development-observation process;

2. The "noise" introduced in the perceived image by photographic grain; and

3. The limitations imposed by the physiological and psychological systems of the observer.

- 334 -

Charman and Olin also suggested that any method for
rating the quality of a camera system should take into
account the effects of each stage in the imaging system,
such that they might be combined to produce the overall
quality measure. The modulation transfer function approach
permits exactly this type of sequential analysis. Thus, the
effects of image motion, emulsion type, development, optics,
etc., on the performance of the photographic system could be
studied individually and collectively in order to optimize
overall performance. These considerations have been incor-
porated into the metrics that follow.

5.4.1.1 Modulation transfer function area (MTFA)

Various concepts of photographic image quality have been
proposed over the years, and have been reviewed by Biberman
(1973) and Brock (1964). That one which has the most direct
bearing upon discrete display image quality is the MTFA.

Proposed by Charman and Olin (1965), this measure uses
the MTF curve as well as some assumed detection threshold
curve of the eye. As illustrated in Figure 110, the MTFA is
the area bounded by the system MTF curve and the detection
threshold (or contrast sensitivity) curve, with the integra-
tion performed from 0 spatial frequency to the point of
crossover of these two curves. Charman and Olin, calling
this area measure the threshold quality factor (TQF),
derived an analytical expression for the threshold function,
as indicated in equation (21):

$$M_t(\nu) = 0.034 \, [dD/d(\log_{10}E)]^{-1}$$
$$\times \, [0.033 + \sigma(D)^2 \, \nu^2 \, S^2]^{1/2} \qquad (21)$$

where

$$S = 4.5,$$

$(dD/d(\log_{10}E) = \text{film gamma, } \lambda,$

$\sigma(D) = \text{rms granularity with 24-}\mu \text{ scanning aperture, and}$

$\nu = \text{spatial frequency.}$
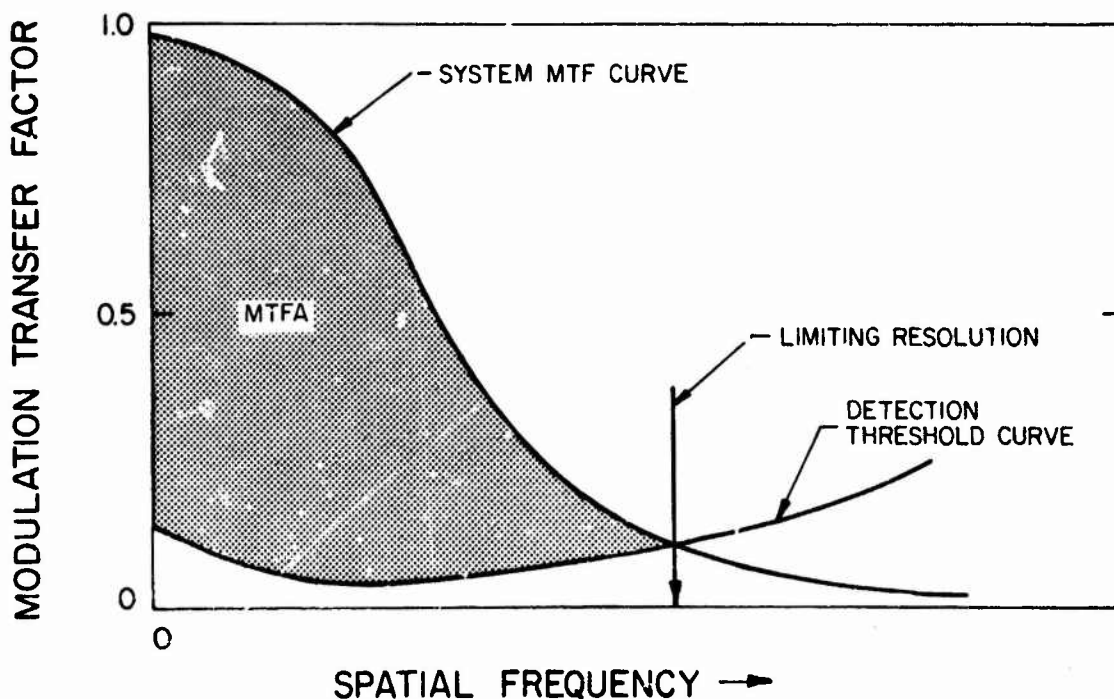
Figure 110: MTFA concept

Their equation makes use of the detection threshold cur e
for tri-bar targets developed by Hufnagel, and includes

- 336 -

by Brock. The threshold curve has a level portion at lower spatial frequencies due to the threshold modulation sensitivity of the eye, assumed to be about 0.04. As the spatial frequency increases, a point is reached where the grain noise causes the modulation threshold to increase.

The detection threshold curve is as shown in equation (21), while the MTFA, plotted on log-log coordinates, becomes:

$$\text{MTFA (log-log)} = \int_{\log \nu_0}^{\log \nu_1} \log\left[\frac{M_0 \, R_0(\nu)}{M_{D,t}(\nu)}\right] d(\log \nu), \qquad (22)$$

where

ν_0 = low spatial frequency limit, in lines/mm,

ν_1 = high spatial frequency crossover of MTF

and $M_t(\nu)$,

$R_0(\nu)$ = MTF value at spatial frequency ν,

M_0 = object target modulation, and

$M_{D,t}(\nu)$ = image target modulation.

When the MTF curve and the detection threshold curve are plotted on linear coordinates, the MTFA is given by equation (23):

$$\text{MTFA(linear)} = \int_0^{\nu_1} \{R_0(\nu) - (M_{D,t}(\nu)/M_0)\} \, d\nu. \qquad (23)$$

The linear form uses no lower frequency cutoff, whereas the log-log formulation (equation (22)) often employs an

- 337 -

the linear form uses no lower frequency cutoff, whereas
the log-log formulation (equation (22)) often employs an
arbitrary cutoff at, say, 10 lines/mm, simply to avoid an
inappropriately large weighting upon the low spatial fre-
quencies on the log-log plot. The nature of the linear pict
avoids the need for such an arbitrary cutoff. For a further
discussion of this concept and its application, see Snyder
(1973).

5.4.1.2 Human observer performance and unitary measures of image quality

Eastman-Kodak Company. Granger and Cupery (1972)
reported the results of a long sequence of experiments
which assessed the appropriateness of an optical merit
function (SQF) which correlated very well with subjective
image judgments. The SQF, or Subjective Qualities Fac-
tor, is based upon the MTF of the system in conjunction
with the contrast sensitivity function. Using the con-
trast sensitivity function of Schade (1964), which shows
the major sensitivity to lie between 10 and 40 lines per
millimeter at the retina, Granger and Cupery defined the
SQF as the integral of the system MTF (including lenses
and films) between the limits of 10 and 40 cycles/mm when
the MTF has been scaled to the retina of the observer by
appropriate considerations of the magnification of the
system. Thus, the SQF is defined mathematically as:

$$SQF = K \int_{10}^{40} |r(f)| \, d(\log f) \tag{24}$$

- 338 -

r(f) = the Optical Transfer Function,

f = spatial frequency, and

K = a normalizing constant.

They further point out that it is inappropriate to limit the considerations to a one-dimensional MTF, because real images contain two-dimensional MTF descriptors, and the person judging image quality obviously takes into account the two-dimensional structure. Further, because the image may not be isotropic, they recommend use of the SQF for two dimensions, as defined in equation (25), which puts the SQF in polar coordinate form:

$$SQF = K \int_{10}^{40} \int_{0}^{2\pi} (f, \theta) \; d(\log f) \; d\theta, \tag{25}$$

where

θ = spatial frequency in cycles/mm along a given azimuth.

They have conducted several experiments to evaluate the effects of various characteristics of the image, all of which relate to one form or another of the modulation transfer function. Examples of the shapes of modulation transfer functions which they have used are illustrated in Figure 111. These results are illustrated in Figure 112, which shows a very strong correlation between the computed SQF and the paired-comparison judgment of print

- 339 -

quality. It should be noted that the criterion of print quality here is simply based upon which of the two images the subject prefers, and no definition of "quality" is given, deliberately.
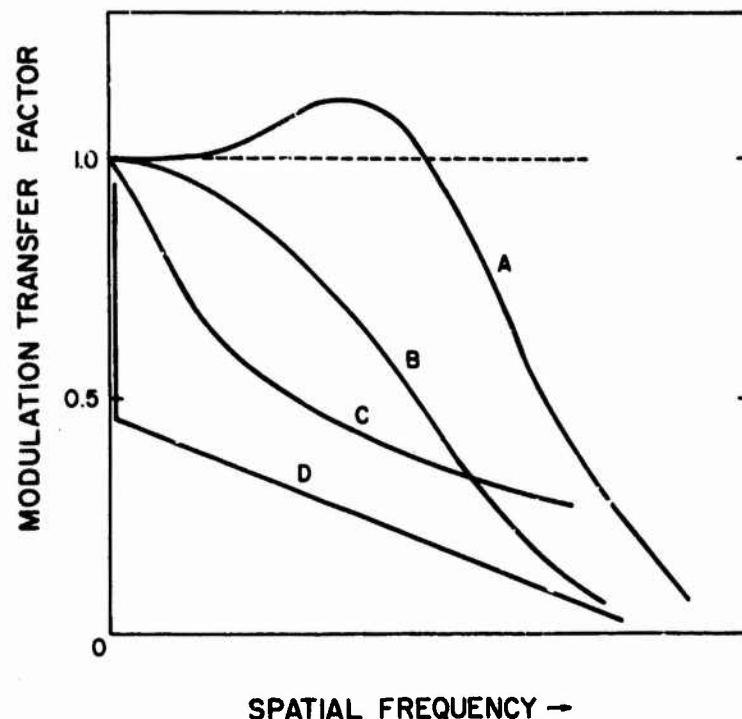
Figure 111: MTF shapes used in Eastman-Kodak studies

Of interest is the fact that granularity and contrast are not included in their work, and they speculate upon the effect of these parameters to some extent. Using the concept of a threshold factor, they suggest that it is necessary to determine the subjective image quality due to quality losses attributable to granularity and contrast. In fact, as illustrated in Figure 113, they recommend a modification of the SQF integral based upon the relationship between the system MTF and this thresh-
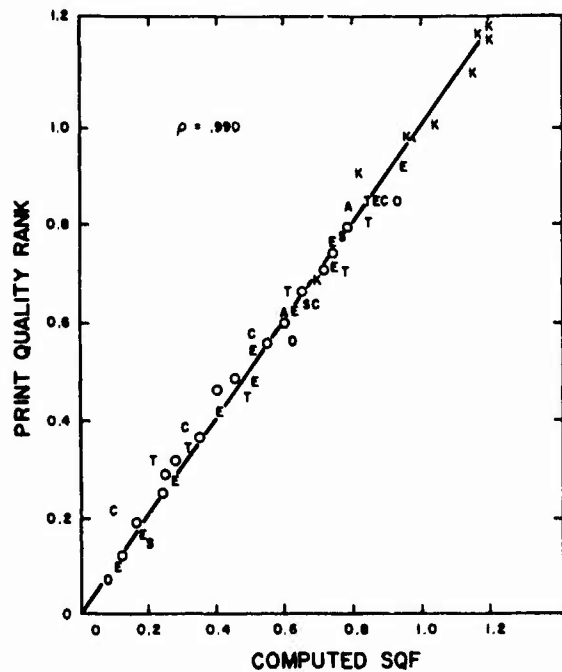
Figure 112: Results of Granger and Cupery (1972)

old curve. Although they do not mention the existence of
the MTFA approach, the lined area between the MTF curve
and the threshold curve, as illustrated in Figure 113 is
essentially the MTFA with the integration performed only
between 10 and 40 lines per millimeter on the retina.

In a separate report from the Eastman Kodak Research
Laboratories, Kriss, Michelson, and Nail (1971) also
related the quality of a photographic image to the con-
trast sensitivity function of the human visual system.
They took color photographs and digitally processed them
to enhance specific spatial frequency bands in relation
to the spatial frequency sensitivity of the human visual
system. In their processed pictures, they varied the
mid-point of the spatial frequency band which was

- 341 -
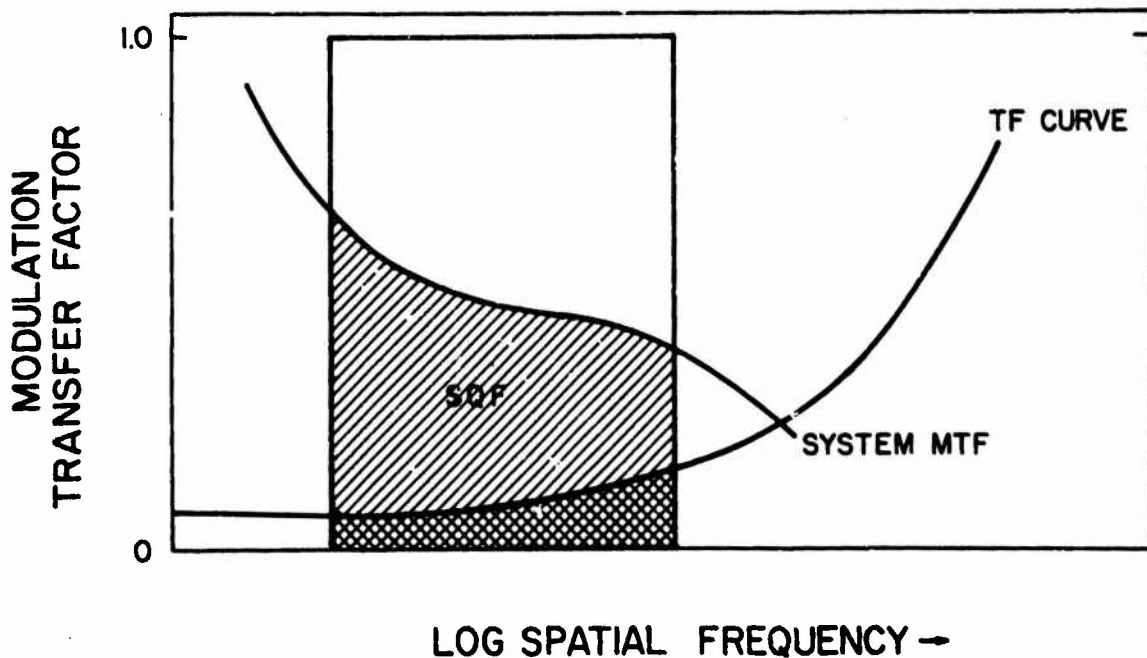
Figure 113:   Definition of SQF as area under the MTF,
              from Granger and Cupery (1972)

enhanced, as well as the slope of the rise and fall of
this enhancement interval.

Kriss et al. then presented pairs of these photo-
graphs to a group of 17 observers, who were asked to
select the picture of the pair which was the most sharp.
Their results, illustrated in Figure 114, indicate that
the maximum sharpness was perceived at a retinal spatial
frequency of about 1 cycle per millimeter, for the view-
ing distance of 30 centimeters. Thus, the peak of this
curve falls approximately on the peak of the response
curve of the eye at this viewing distance, with a signif-
icant decrease in judged sharpness as the midpoint of
this enhancement interval falls to either side of 1 cycle
per millimeter. The authors concluded that

- 342 -

to produce a sharp picture it would be advantageous to increase the contrast of information that falls in the 8 cycles/degree to 15 cycles/degree range of the eye.

Generally, one can conclude that increases in image modulation above the threshold of the eye produce the greatest perception of sharpness, and that for a constant amount of modulation, the greatest sharpness is perceived in the region of spatial frequencies where the eye is most sensitive. Again, it should be noted that these data in no way relate to either the scale of the image or the granularity of the image, simply because the viewing distance was kept constant and the photographs were grain-free.

Boeing Company. The modulation transfer function area (MTFA) concept has been evaluated for photographic imagery in two separate studies. In the firs study (Borough, Fallis, Warnock, and Britt, 1967), the MTFA was related to subjective estimates of image quality obtained from a large number of trained image interpreters. In the second of these experiments, actual information-extraction performance data were obtained, as well as subjective estimates of image quality, and both measures were compared with the MTFA values of the imagery. Nine photographic reconnaissance negatives were used as the basis for laboratory-controlled manipulation of image quality. Each of the scenes was printed in 32 different MTFA variants, determined by the combination of four dif-
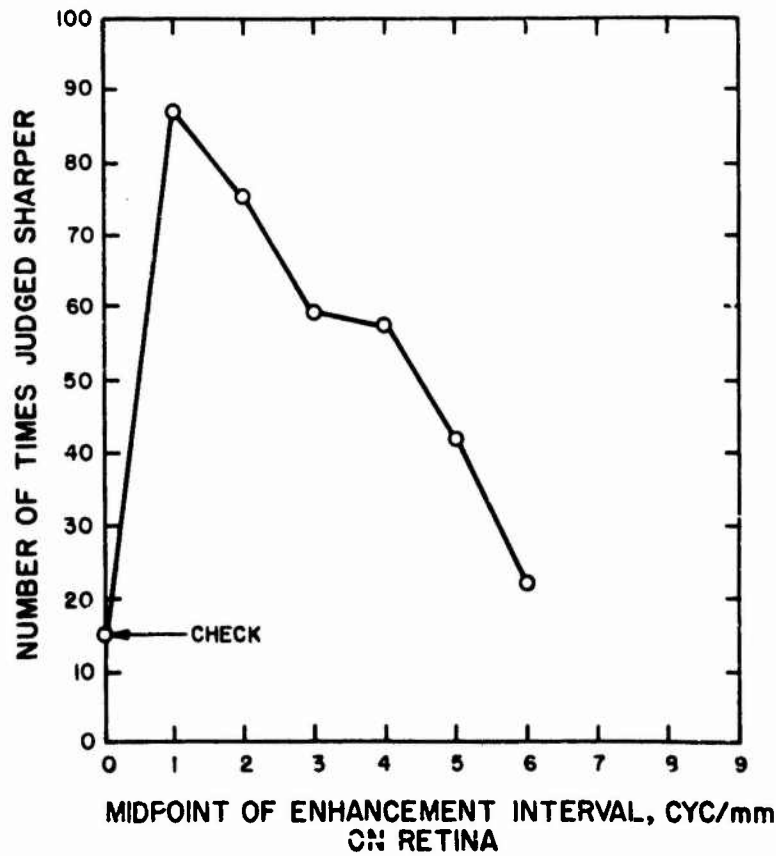
Figure 114: Improvement of subjective sharpness by
selective spatial frequency enhancement,
from Kriss et al. (1971)

ferent MTFs, three levels of granularity, and three
levels of contrast (Figure 115). Four cells of the
matrix were deleted because their MTFA values corre-
sponded to others in the 32-cell matrix.

Correlations were obtained between the subjective
image quality rating (derived from paired comparisons)
for each of the 32 variants and several physical measures
of image quality, including MTFA. Table 22 shows the
results. Most important for our purposes is the mean
correlation of 0.92 between MTFA (linear) and subjective
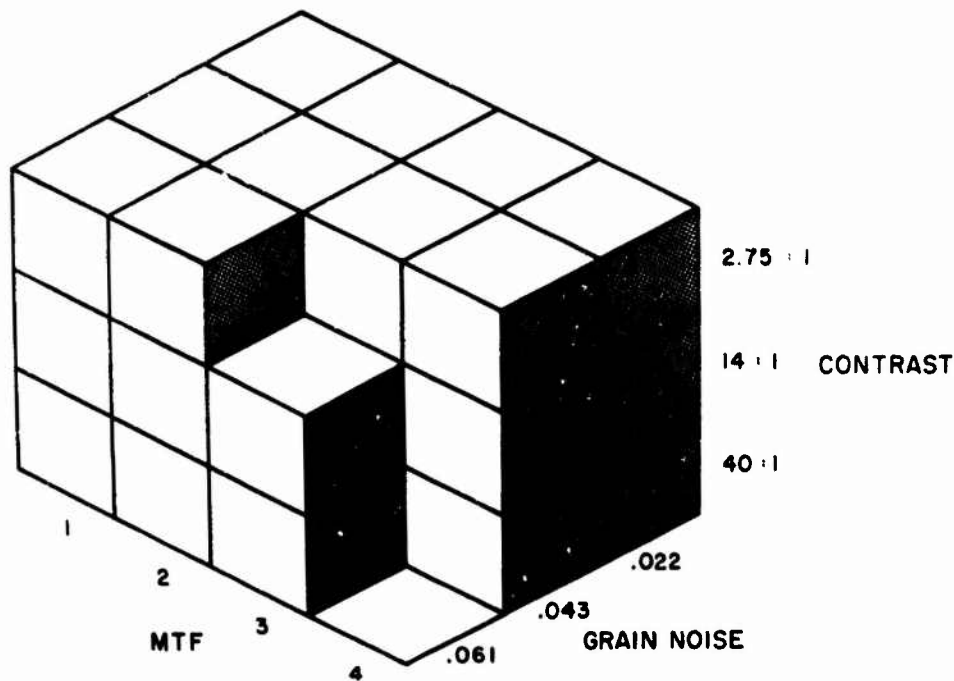
Figure 115: Borough et al. (1967) experimental design

image quality, which indicates that MTFA is strongly related to subjective estimates of image quality.

The second experiment, by Klingberg, Elworth, and Filleau (1970), examined the relationship between objectively measured information-extraction performance and the MTFA values. As a check on the results of the previous experiment, Klingberg et al. also obtained subjective estimates of image quality, so that all three intercorrelations among MTFA, subjective image quality, and information-extraction performance could be obtained. The imagery used for this experiment was the same as that used by Borough et al. (1967).

Figures 116 and 117 show the results of this experiment. The resulting correlation between number of errors

TABLE 22. Correlations between subjective quality and physical measures, from Borough et al. (1967)

Physical variable	Scene Number									Mean* r
	1	2	3	4	5	6	7	8	9	
MTFA (linear)	0.921	0.927	0.900	0.925	0.935	0.919	0.919	0.920	0.913	0.920[†]
Modulation	0.220	0.641	0.511	0.618	0.680	0.699	0.497	0.698	0.632	0.576
MTF	0.698	0.529	0.580	0.660	0.579	0.608	0.697	0.469	0.542	0.601
Granularity	-0.543	-0.632	-0.618	-0.450	-0.516	-0.428	-0.505	-0.589	-0.577	-0.543
MTFA (log-log, 2 cycle)	0.666	0.863	0.866	0.821	0.874	0.890	0.749	0.902	0.876	0.846
MTFA (log-log, 2 cycle)	0.768	0.923	0.923	0.867	0.920	0.921	0.824	0.941	0.920	0.900
Acutance	0.599	0.448	0.526	0.568	0.564	0.599	0.625	0.440	0.602	0.555

* These mean values were determined by transforming the correlations to z values. Such a transformation is necessary when correlations are being combined to obtain a mean correlation.

† This mean value was significantly greater ($p < 0.01$) than all of the other mean correlation values except the value for MTFA (log-log, 2 cycle). This latter value was still significantly less than the MFTA linear value at the 0.05 level of significance.

and mean MTFA is -0.93. (The minus value is due to the use of number of errors as a measure, which is inversely related to MTFA.) Individual correlations among performance, MTFA, and subjective quality are shown in Table 23. It is apparent that the relationship between MTFA and performance is not as high for some scenes as for others, but that the mean correlation across scenes (0.72) is quite high. Further, disregarding scene content and placing all scenes on a common performance continuum, the correlation of -0.93 accounts for over 86% of the variance in information-extraction performance. The 0.96 correlation of MTFA with subjective quality agrees quite well with the correlation of 0.92 obtained by Borough et al. Finally, there was a very high correlation (r = 0.97) between information-extraction performance and subjective scale values of image quality.

The Boeing studies illustrate that a physical measure of image quality based upon the excess of MTF over the threshold detection level, integrated uniformly over all functional spatial frequencies, correlates highly with both the ability of the photointerpreter to obtain critical operational information from the imagery and with his overall subjective estimate of the quality of the imagery. This measure, the MTFA, includes the spatial frequency content of the imagery. the spatial frequency response (CSF) of the observer, and the granularity con-
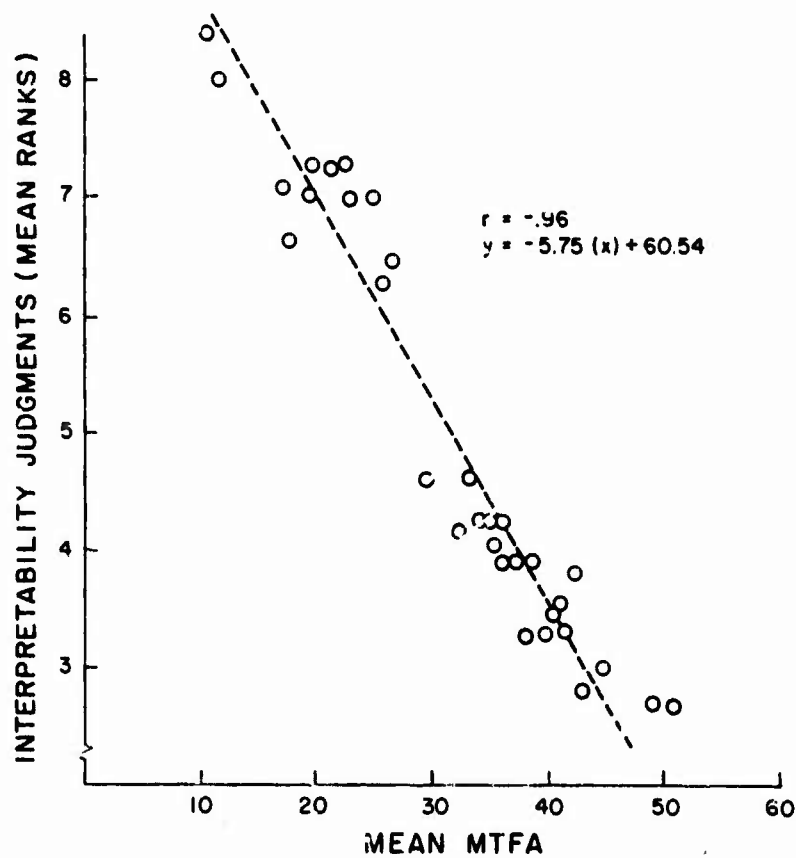
Figure 116: Correlation between MTFA and subjective quality, from Klingberg et al. (1970)

tained in the imagery. Thus, it extends the work performed by the Eastman Kodak researchers to include granularity as a component of image quality, and tends to encompass all pertinent image quality variables, at least for statically presented imagery. This concept has subsequently been applied to electronic displays, as will be discussed in the following sections.
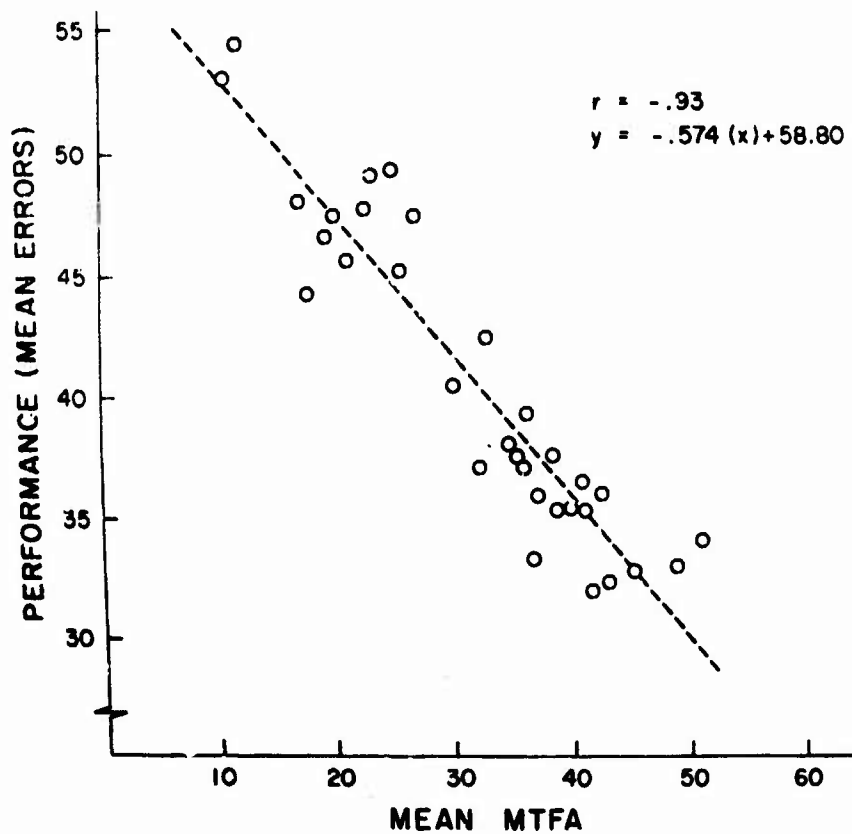
Figure 117: Correlation between MTFA and performance, from Klingberg et al. (1970)

5.4.2 One-Dimensional Spatially Discrete Monochrome Displays

Whereas the photographic film image is continuous in both spatial dimensions, there is a class of display which is continuous in only one dimension and discretely sampled in the other spatial dimension. Two general kinds of systems form such an image or display, the line scan film system and the more familiar television system.

The line scan film system uses one detector or an array of detectors, with suitable imaging optics. to scan across a scene in one direction. The output of these detectors

TABLE 23. Correlations among Image Quality, Interpreter Performance, and Subjective Judgements, from Klingberg et al. (1970)

	Scene									$\bar{r}$	r_m
	1	2	3	4	5	6	7	8	9		
Performance/MTFA	0.69	0.66	0.80	0.65	0.78	0.55	0.84	0.86	0.46	0.72	0.93
Performance/rank	0.71	0.67	0.89	0.60	0.80	0.42	0.78	0.76	0.42	0.70	0.96
MTFA/rank	0.90	0.87	0.90	0.93	0.94	0.87	0.92	0.86	0.83	0.90	0.97

* N = 32 image quality levels (MTFA). $\bar{r}$ is the average of r's using z scores. r_m is the values averaged across scenes before computing scene-by-scene correlations.

(e.g., charge coupled devices, or CCDs) is a continuous voltage proportional to the radiance of the scene. This voltage is then either stored electronically, or is used to modulate the light output of a writing element, such as a CRT beam or an LED. Film recording systems use this light output to expose a synchronized strip across the film plane, producing a film strip proportional in exposure to the radiance strip of the sampled scene. CRT display systems use the detector output to modulate the CRT beam, creating a line on the CRT the luminance of which is proportional to the radiance of the sampled scene strip.

The other dimension of the two-dimensional image is obtained by moving the optical detector (or detector array) to sample a new, parallel and adjacent strip on the scene. This detector movement can be accomplished by translation (e.g., flying an airplane over the scene) or by rotation of the imaging system optical axis.

In line-scanning systems of this type, the resulting image is continuously sampled along the scanning line, but the scene (and therefore) image is discretely sampled in the other direction, from scanning line to scanning line. The output image can, of course, be in either "hard" copy (e.g., film) or "soft" copy (e.g., a CRT display).

The very familiar television display is formed in a similar manner, by continuously scanning the camera's electron beam horizontally, and sampling the image continuously in

that direction, then stepping the beam downward and repeat-
ing the process for the next line. The resulting raster-
scan display is continuously sampled in one direction (i.e.,
horizontally) and discretely sampled in the other (i.e.,
vertical) direction.

5.4.2.1 Display parameters and observer performance

During the past two decades, several hundred laboratory
and analytical studies have been performed to assess the
relationship between variation in many of the line-scan dis-
play image parameters and various measures of observer per-
formance. Conclusions drawn from these studies (e.g., Hair-
field, 1970; Snyder et al., 1967) have indicated that
cross-study comparisons are virtually impossible. The rea-
sons for such inconsistent conclusions across studies is due
to the significant interaction of many display and system
parameters.

Specifically, variation in one or two, or sometimes
three, variables of the system have been examined parametri-
cally, but other variables have been either left constant or
uncontrolled. In the absence of controlled or parametric
combination, any effort to put the data together into some
meaningful, quantitative combination of results is at least
hazardous, if not totally impossible. Parameters which have
been studied, either independently or in concert, include:
mean luminance, size, viewing distance, number of active

raster lines, contrast, scene movement, gamma, signal-to-noise level, aspect ratio, raster direction, video bandwidth, and contrast enhancement techniques. The number of experiments in the literature dealing with these parameters is legion, and there is no need to recap those results here. Rather, effort will be devoted primarily to evaluating those unitary measures of image quality which have attempted to combine many or all of these parameters, and to evaluate the results of such efforts.

The fundamental measures of system image quality have been labeled noise equivalent pas band (N_e), displayed signal-to-noise ratio (SNR_D), and the MTFA, described previously. Variations on these concepts have also been studied empirically (Task, 1979). The concepts of each, along with results obtained to date, will be described very briefly below. Although each of these measures is generally applicable to a raster-scan display, most have been evaluated in terms of tasks such as the detection and recognition of specific man-made objects in the presence of either a cluttered background or a plain background. These evaluation techniques should also be applicable to the individual identification of alphanumeric characters and character chains (e.g., words and paragraphs), although only a limited amount of such research has been performed to date.

$\underline{N_e}$. Given knowledge of this response curve, it has been convenient (Schade, 1953) to consider the quality of the

television image as proportional to the Equivalent Passband, N_e, the passband of an equivalent rectangular noise spectrum with an abrupt cutoff (at spatial frequency N_e) which passes the same _total_ sine-wave energy as the actual spectrum. This concept is illustrated in Figure 118. It should be noted that the sine-wave response is one-dimensional, but that N_e is the two-dimensional aperture response of the system, and therefore is determined from the square of the one-dimensional sine-wave response:

$$N_e = \int_0^\infty [R(N)]^2 \, dN \qquad (26)$$

where

R(N) = the percent response, and

N = the spatial frequency in TV lines/picture height.

This summary measure has been derived and pioneered by Schade, and has been used in the TV industry for some years. For usage in performance prediction of present-day display systems, however, it appears to have one liability, namely, that it does not take into account the varying noise levels which a system might have as, for example, the detector irradiance level changes with changes in scene illumination.

$\underline{SNR_D}$. Using the analyses of Schade (1953) as background, Rosell (1971) developed an approach for analyzing television systems which gets closer to the human observer's visual capability. Rosell's approach is to relate all system parameters to the analytically derived signal-to-noise ratio at the display (SNR_D). Then, assuming the human observer

- 354 -
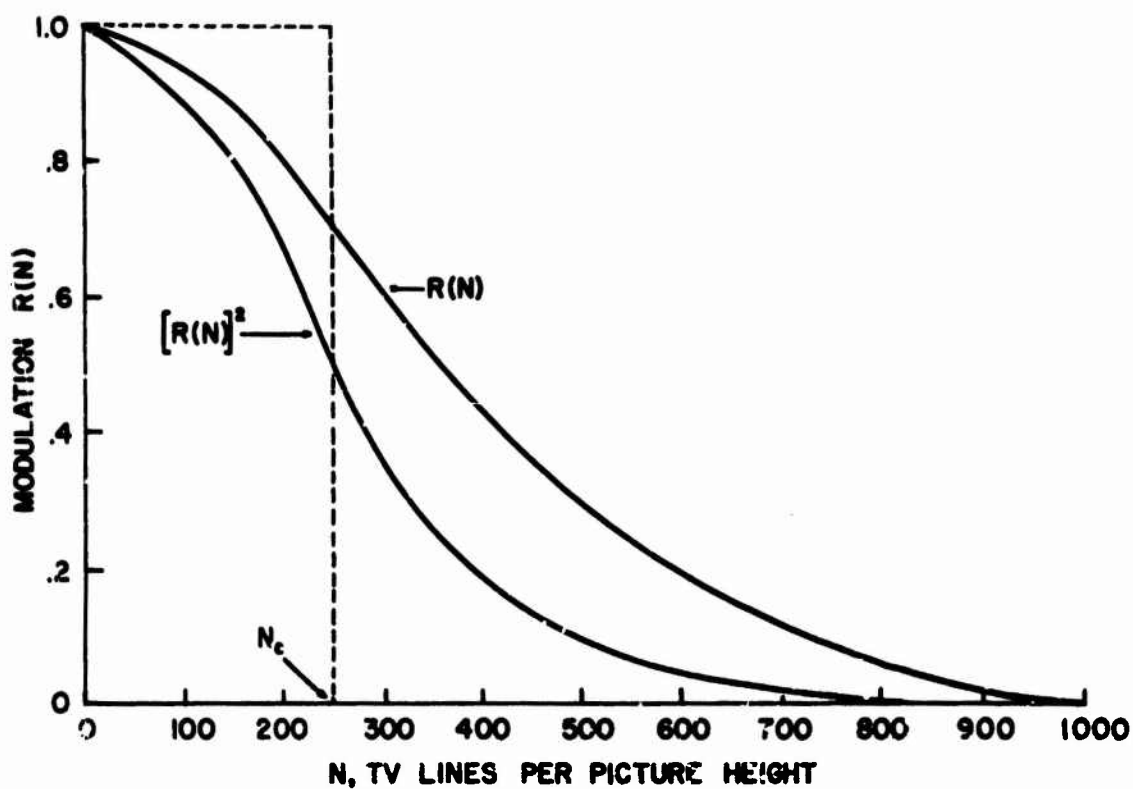
Figure 118: Noise equivalent passband, N_e

requires an SNR_D of approximately 2.8 to have a 50% chance of detecting a target, system tradeoffs are made to achieve this or some other value of SNR_D. Many laboratory studies have been performed to establish this probability of detection for gratings and solid rectangles as a function of SNR_D. Observer confidence levels, task loading, ambient environments, dynamic scenes, target textural characteristics, and other factors have not been considered.

There are many variants of the SNR_D concept, depending upon whether one assumes the limitations in the line-scan system to be, for example, photon limited, preamplifier lim-

ited, display limited, etc. For purposes of discussion, however, an elementary calculational formula is given by Rosell and Willson (1971):

$$SNR_D = (at\Delta f_V/A)^{1/2}$$

$$\times (Ci_{max})/((2-C) \ e \ \Delta f_V \ i_{max})^{1/2} \tag{27}$$

$$= ((a/A) \ t \ \Delta f_V)^{1/2} \ SNR_V \tag{28}$$

where

SNR_D = signal-to-noise ratio at the display,

a = area subtended by target at photosurface,

A = total area of photosurface,

t = integration time of eye, assumed to be constant, between 0.1 and 0.2 s,

Δf_V = video bandwidth, in Hz,

C = target contrast,

i_{max} = maximum photocurrent,

e = charge of an electron, and

SNR_V = signal-to-noise ratio in the video.

The relationship contained in equation (28) relates the signal-to-noise in the video signal to SNR_D, the signal-to-noise in the displayed image. The key to Rosell's approach is in the bracketed term in equation (28). Essentially, this term provides for both spatial and temporal integration of the signal, and reflects the visual system's spatial and temporal integration capabilities. First, the signal is assumed to be integrated over a finite portion of the display (a/A). The larger the proportion of the display sub-

tended by the target, the greater the signal, with signal strength proportional to the square root of the target area, a. Secondly, the signal is integrated over the integration time of the eye, Δt, which is assumed to be constant and about 0.2 s. As demonstrated previously in this report, this integration time is not at all constant. The model also assumes that spatial integration is constant over all areas, a, which has a limiting case, but is certainly reasonable for small target areas. The inclusion of the video bandwidth term in the brackets is only to cancel a like term in the denominator of SNR_V, and has no geometric or temporal meaning. Assuming that the display is isotropic, then the signal strength (or resolution) is certainly proportional to the video bandwidth, and thus the term's inclusion is also logical.

Analysis of equation (28) indicates that SNR_D varies directly with the square root of target area, a, and with SNR_V. Thus, the same SNR_D and hence the same visibility of a target will exist with reciprocal covariation in SNR_V and a. Figures 119 and 120 illustrate this relationship. Rectangles of varying sizes were displayed on a television monitor, and the subject was asked to respond as to whether or not he could discern the target. Variation in the video signal-to-noise ratio resulted in consistent variation in the probability of detection, as illustrated in Figure 119. When the SNR_V was converted to SNR_D by equation (28), the

result was as indicated in Figure 120, which shows that probability of detection is a direct result of SNR_D. Using the data shown in Figure 121, Rosell and Willson (1973) also indicated that the eye is a perfect spatial integrator over targets subtending up to 0.5 deg, but is less efficient in integrating over larger lengths.
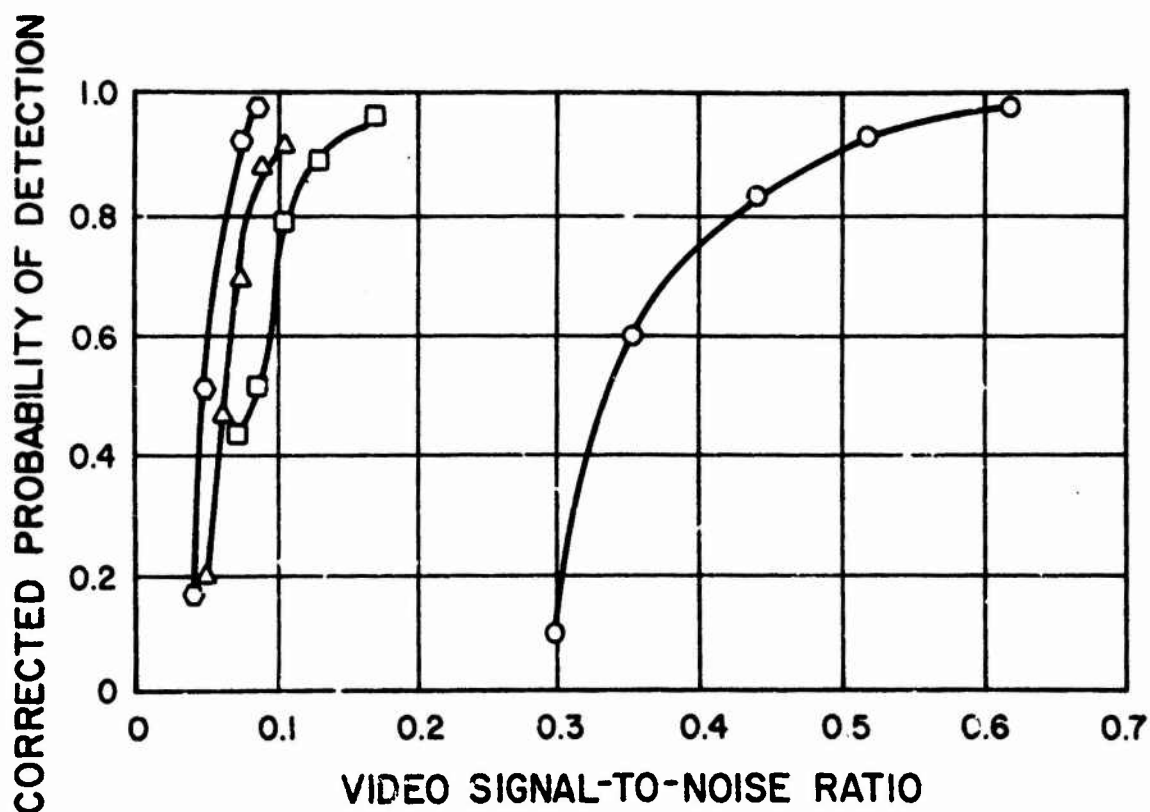
Figure 119: Corrected probability of detection for rectangular images of size (⊙)4x4, (▢)4x64, (△)4x128 and (◇)4x180 scan lines, from Rosell and Willson (1973)

Rosell and Willson (1973) have conducted several experiments, asking the subjects to detect, recognize, and identify various displayed real-world targets in several clut-
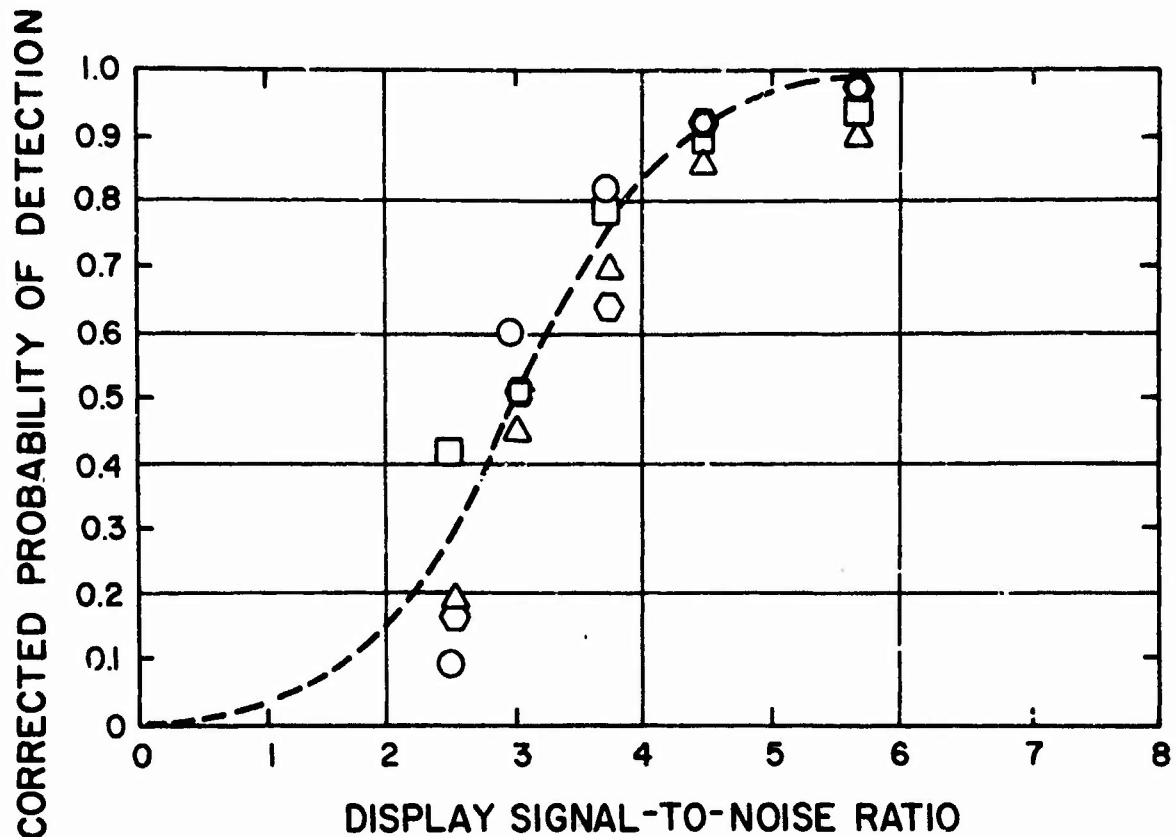
Figure 120: Corrected probability of detection for rectangular image, predicted by SNR_D, from Rosell and Willson (1973)

tered and non-cluttered backgrounds. A summary of their experiments is shown in Table 24, while the results are shown in Table 25. The threshold value of SNR_{DI} should be a constant, independent of spatial frequency, for a given row across Table 25. Such is clearly not the case. The threshold SNR_{DI} for the detection response in a uniform background (totally uncluttered) is 2.8 for all spatial frequencies, as predicted. However, as soon as clutter is added or a more cognitive response is required of the subject, the constancy of SNR_{DI} breaks down. Unfortunately, Rosell and Willson do

ANGULAR WIDTH OF SQUARES, DEG
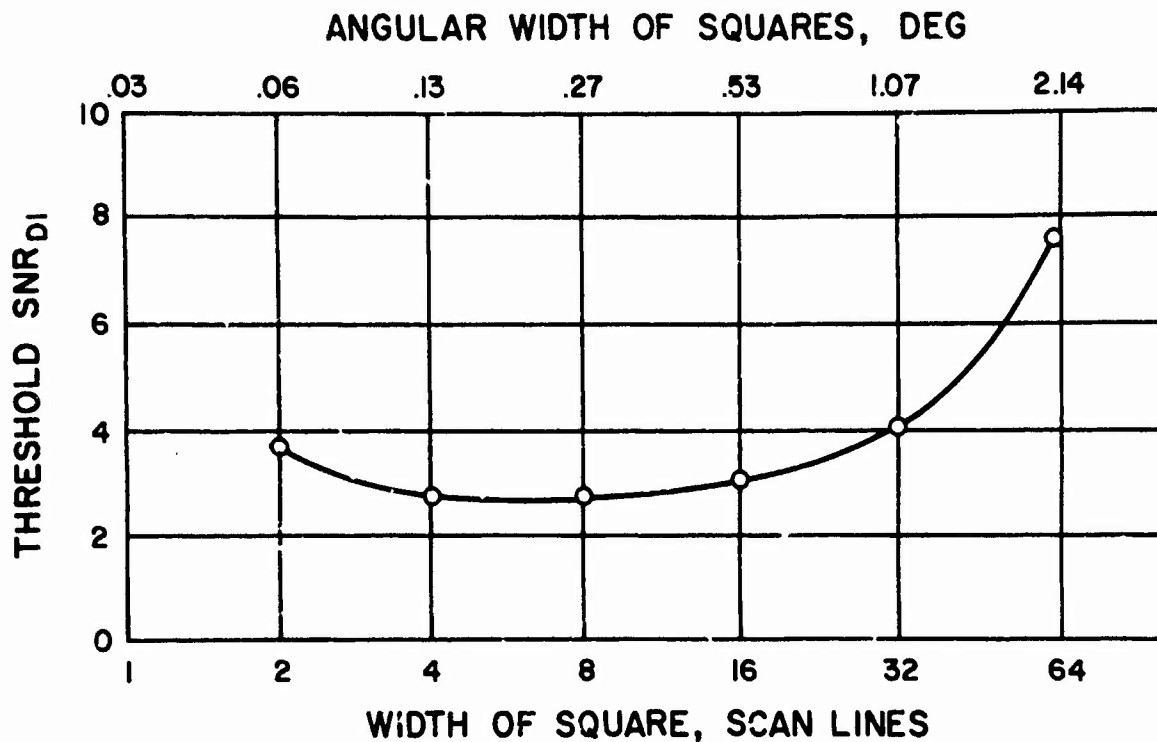
Figure 121:   Threshold SNR_{DI}, from Rosell and Willson (1973)

not present any statistical analyses regarding the signifi-
cance of any differences in their data.   One can probably
safely assume, however,   that the tabled values of SNR_{DI} are
really quite different from one   another within all rows
except the "detection/uniform background" row.

In spite of the lack of a real-word imagery data base and
empirical studies to support the SNR_D concept,   it has great
utility in system tradeoff studies,   simply because the
development of the concept is derived from many first prin-
ciples of the physics of imaging systems.

TABLE 24. Conditions for the Recognition and Identification Experiments, from Rosell and Willson (1973)

Expt. No.	Test Image	Background	D_V/D_H	L_D, ft-L	Δf_v MHz	Number of trials	Number of observers	N*	Frames sec^{-1}
15	Vehicle recognition	Uniform	3.5	1.0	12.5	1000	5	875	25
16	Bar patterns	Uniform	3.5	1.0	12.5	872	5	875	25
17	Vehicle recognition	Road	3.5	1.0	12.5	800	8	875	25
18	Vehicle recognition	Grass	3.5	1.0	12.5	800	8	875	25
19	Vehicle recognition	Grass-trees	3.5	1.0	12.5	800	8	875	25
18	Tank identification	Uniform	3.5	1.0	12.5	1250	5	875	25
19	Tank identification	Uniform	3.5	1.0	12.5	1250	5	875	25

*D_V/D_H is the display viewing distance-to-height ratio; L_D is the average monitor luminance, ft-L; Δf_v is video bandwidth, MHz; N is the number of scan lines per picture height, interlaced 2:1.

<u>MTFA</u>. Because the MTFA was shown to be useful in predicting both the subjective image quality and the objectively determined quality of photographic imagery, Snyder et al. (1974) investigated its applicability to raster-scan displays.

Since extrapolation from theoretically-derived photographic data was inappropriate, they first obtained representative, parametric contrast sensitivity functions for the different line rate/video bandwidth systems of interest. This inappropriate extrapolation derived from (1) the lack of empirically-derived curves, (2) the lack of generalization from photographic static noise (granularity) to dynamic rms noise of a video signal, and (3) the inability to determine the means by which contrast sensitivity curves should be shifted for variations in specific TV parameters, such as line rate, bandwidth, gamma, MTF, noise bandwidth, etc.

For this reason, Keesee (1976) determined contrast sensitivity functions for a variety of video system configurations, varying in line rate (525, 945, 1225 lines/frame), noise passband, and noise amplitude. Keesee's experimental design is illustrated in Figure 122.

While Keesee's results are too lengthy and complex to repeat here, some generalizations are important in an effort to understand the complexity of television image quality. First, variation in the bandwidth (or frequency) of any non-correlated noise has a strong impact on the visibility of

Table 25. Best Estimate of Threshold SNR_{DI} for Detection, Recognition, and Identification of Images (Rosell and Willson, 1973)

Discrimination level	Background	k_d, TV lines per minimum dimension	Threshold SNR_{DI} for a single bar of spatial frequency (in lines/picture height) equal to			
			100	300	500	700
Detection	Uniform*	1	2.8	2.8	2.8	2.8
Detection	Clutter	2	4.8	2.9	2.5	2.5
Recognition	Uniform	8	4.8	2.9	2.5	2.5
Recognition	Clutter	8	6.4	3.9	3.4	3.4
Identification	Uniform	13	5.8	3.6	3.0	3.0

* Treated as an aperiodic object.
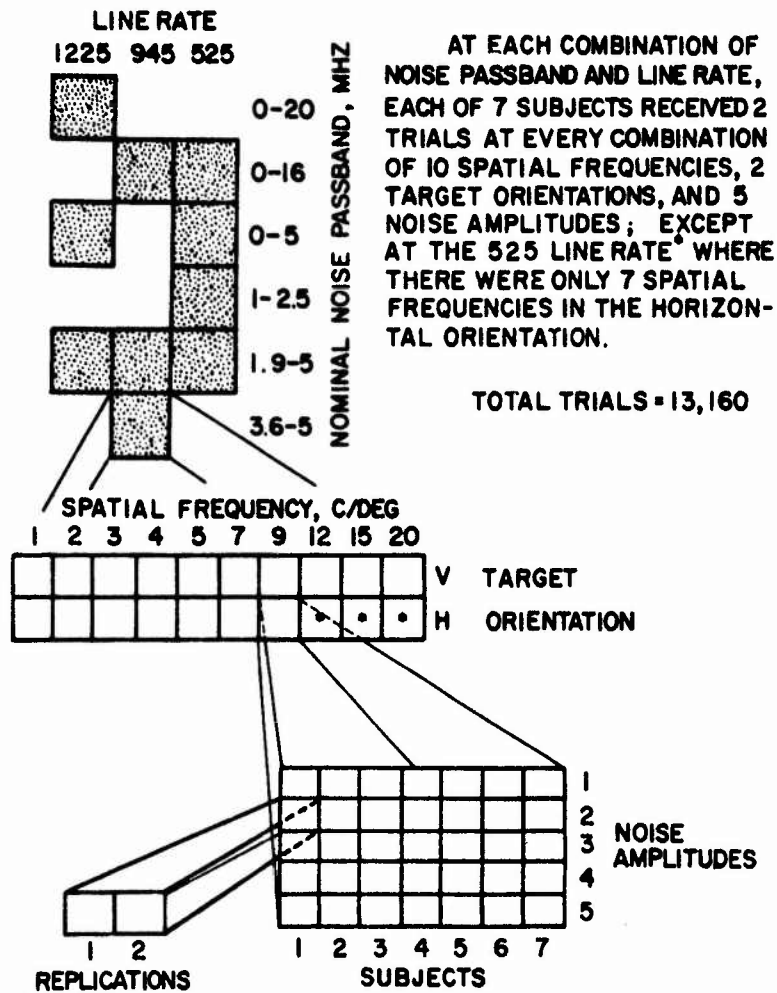
Figure 122: Keesee's (1976) experimental design

image content as indicated by changes in the sine-wave response. Low spatial frequency noise will raise thresholds more than will high-frequency noise. Second, thresholds for gratings parallel to the raster (Figure 123) are higher than thresolds for gratings perpendicular to the raster (Figure 124), whether or not noise is present in the image. Third, the sine-wave thresholds on a raster-scan image are higher than those on a flat-field image (Figures 123 and 124). Last, as the line rate increases, and therefore as the

- 364 -

(black) space between active raster lines decreases, the thresolds are lowered. Stated another way, the presence of a visually prominent raster interferes with grating detection. (This is also critical for dot-matrix displays.)
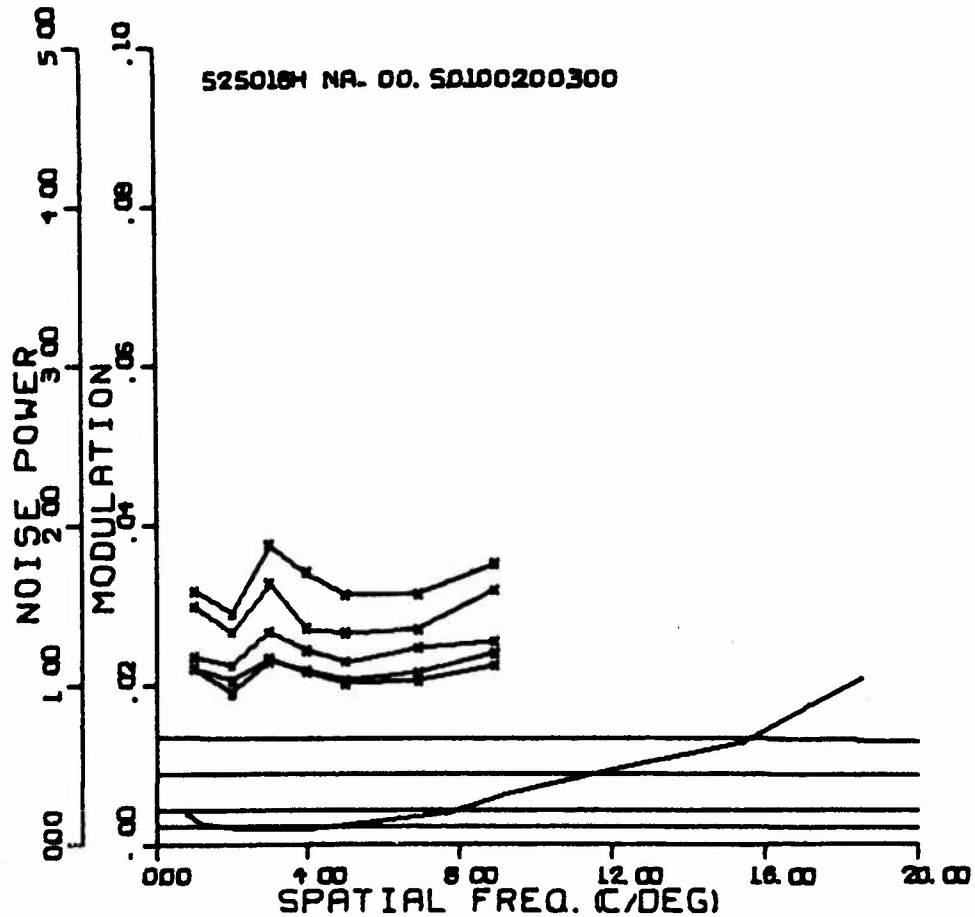
Figure 123: Thresholds for gratings parallel to the raster, Keesee (1976)

Data obtained to date indicate that the MTFA concept is a valid predictor of the overall quality of a raster-scan imaging display. In one experiment, Snyder et al. (1974) obtained target acquisition performance in a dynamic air-to-ground search experiment, using five levels of noise
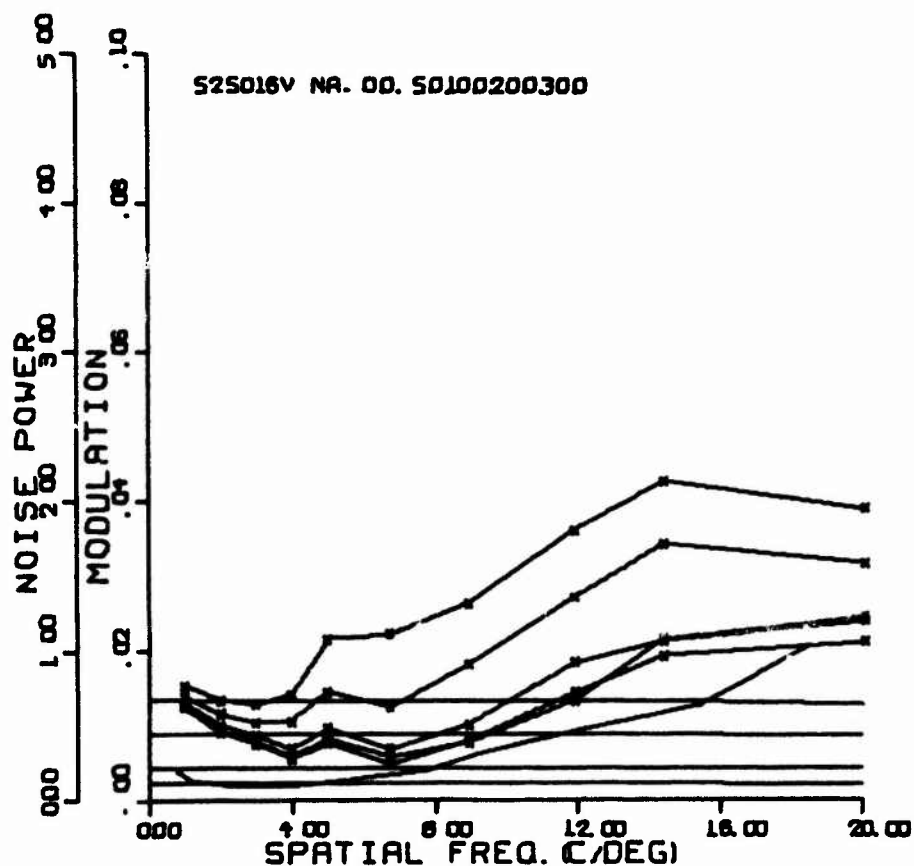
Figure 124: Thresholds for gratings perpendicular to the
raster, Keesee (1976)

on a 945-line, 16 MHz video system. The results indicated
that MTFA correlated 0.965 with the percent correct
response, -.973 with the percent incorrect response, and
0.765 with slant range at the time of recognition. All data
were for real targets in real terrain backgrounds.

Of interest in this context is the extent to which SNR_D
might predict the same performance. Because the targets
were of varying size, and because the responses are averaged
across all targets, it can be shown that the averaged SNR_D
is directly proportional to SNR_V. SNR_V correlates 0.968
with percent correct responses, -.817 with percent incorrect

responses, and 0.514 with slant range. The last two
correlations are good, but not as high as those obtained for
MTFA correlations.

In another test of the MTFA concept, Snyder (1974) meas-
ured the likelihood of correct recognition and the time to
recognize a face (from among 35 faces) on a television dis-
play. Using three video systems, and five noise levels per
system, he found a strong correlation between MTFA and both
measures (Figures 125 and 126).
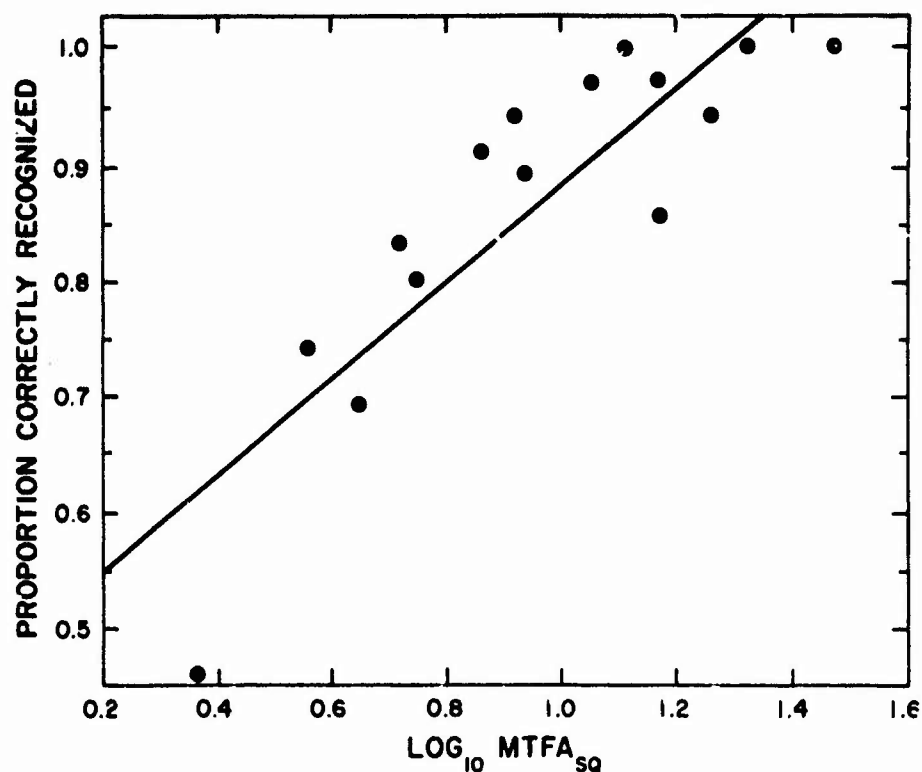
Figure 125: MTFA$_{SQ}$ and likelihood of facial recognition
(Snyder, 1974)

The reason for the log and log-log transformations in
Figures 125 and 126 is to obtain a best linear fit. As Sny-
der (1974) points out,
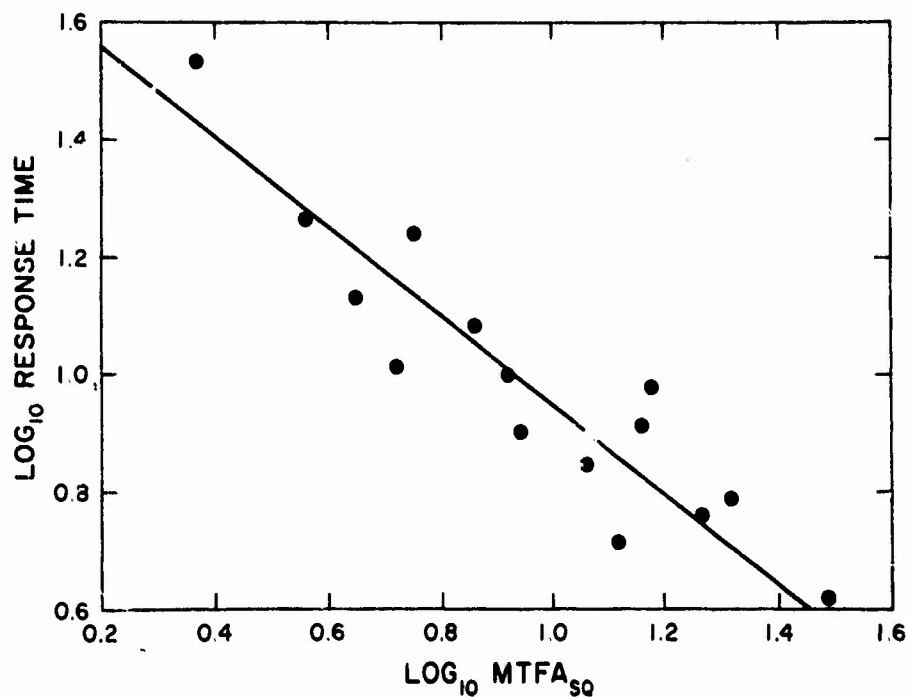
Figure 126: MTFA$_{SQ}$ and response time for facial recognition
(Snyder, 1974)

> the nonlinearity of the best-fit expression . . .
> is probably due to the large images presented to
> the subjects (relatively low spatial frequencies)
> and the simplicity of the task. That is, as the
> MTFA$_{SQ}$ value becomes even moderately large, facial
> recognition performance reaches a ceiling.

The ceiling is based upon the minimum time the subject

requires to look at the TV display, turn to the photographic

images of the 35 faces, and identify the number of the indi-

cated face. Similarly, the percent correct measure must

necessarily asymptote at 100% and, if this value is reached

with a less-than-maximum MTFA, the relationship must, of

necessity, become nonlinear.

Subsequent studies in this series have investigated the

validity and utility of the MTFA concept by systematically

studying the effects of design and system variables which

change the MTFA. Simultaneously, methodological advances were made in developing techniques to measure photometrically (1) the response of various components in a television chain (Snyder, 1976), and (2) the dynamic noise on a television display (Snyder, Almagor, and Shedivy, 1979).

One MTFA evaluation experiment determined the effect of the shape of the MTF curve (high pass, low pass, attenuated) on recognition of alphanumeric characters (Gutmann, Snyder, Farley, and Evans, 1979). Another evaluated the effect of MTF shape on dynamic, air-to-ground search with low-pass, attenuated, and normal video systems (Gutmann et al., 1979). While the search/recognition performance and eye movement measures of visual search were predictably affected by these MTF shape variations, the correlations of performance with MTFA were not very high. For the alphanumeric search experiment, $r = 0.349$ ($p > .05$) between number of targets correctly recognized and MTFA, and $r = -.703$ ($p < .05$) between search time and MTFA. For the dynamic imagery experiment, the correlations between MTFA and performance were $r = 0.73$ ($p < .05$) with range to target and $r = 0.593$ ($p > .05$) with number correct. Thus, the correlations were only fair, causing the authors to conclude that these

> correlation(s) (are) not as great as might be desired for careful system design tradeoffs. The remaining unpredicted variation in most (MTFA) experiments is substantial (p. 68).

Task (1979) evaluated MTFA, as well as other image quality metrics, in an experiment comparing nine different MTFs,

but with no noise added to the video display. He obtained
very high correlations between performance and MTFA (as well
as with other metrics), as shown in Table 26. It may well
be the case that MTFA is a good predictor of complex task
performance only with noise-free displays.

TABLE 26

Task's (1979) Correlations

FOM	Target Recognition (TV Study)	Target Detection POL 1000 POL 2000 (TV Study)		Target Recognition (Film Study)
log BLMTFA	-0.948	0.931	0.878	-0.977
JNDA-log (1/2 cpd)	-0.937	0.928	0.853	-0.983
log S.T. Res	-0.909	0.923	0.864	-0.969
log JNDA	-0.906	0.902	0.838	-0.984
JNDA-log (2 cpd)	-0.896	0.902	0.835	-0.983
S.T. Res	-0.888	0.900	0.834	-0.973
log MTFA	-0.878	0.866	0.777	-0.950
JNDA	-0.876	0.880	0.802	-0.983
log SSMTFA	-0.869	0.857	0.766	-0.978
GFP-log	-0.847	0.869	0.760	-0.858
log Info Dens	-0.820	0.880	0.842	-0.971
ICS	-0.818	0.837	0.724	-0.978
MTFA	-0.811	0.829	0.717	-0.912
Info Dens	-0.795	0.824	0.766	-0.974
log Lim Res	-0.783	0.795	0.709	-0.885
SQF	-0.781	0.803	0.702	-0.979
GFP	-0.781	0.798	0.670	---
Lim Res	-0.764	0.778	0.695	-0.851
N_e	-0.726	0.761	0.618	-0.729

One of the results that is more pertinent to dot-matrix displays is the effect of raster suppression on observer performance. Keesee (1976) showed that the existence of a visually prominent raster interferes with sinusoidal grating detection. It follows, then, that reduction of the raster prominence by spreading the scanning spot verticaly to fill in the between-line spaces should improve observer performance, and should also increase the measured MTFA. In two experiments using the spot-wobble technique of raster suppression (Beamon and Snyder, 1975; Snyder, Beamon, Gutmann, and Dunsker, 1980), it was found that (1) raster suppression did in fact improve operator performance with a noise-free display, and (2) with a noisy display, raster suppression had little or no benefit. Thus, it again appears that the impact of significant random noise on a video display overshadows the prediction accuracy of the MTFA metric. Nonetheless, the importance of suppressing the raster (or sampling space) in a noise-free display must be emphasized because of its direct bearing upon the image quality of two-dimensional discrete sampled displays.

These results, then, support with some limitations the appropriateness of the MTFA concept for evaluating the quality of television, or raster-scan, images. Again, the correlation between MTFA and objectively measured observer performance is generally high, but there is ample room for improvement of prediction.

5.4.3 Two-Dimensional Spatially Discrete Monochrome Displays

In the previous discussion, it was noted that a visibly discrete raster sampling had measurably interfering effects on operator performance. Because most matrix-addressed, flat-panel displays have visible spaces (discrete sampling) between elements, these concerns were addressed in a three-year research program (Snyder and Maddox, 1978). Part of that program led to the sequential development of a quality metric of performance prediction for dot-matrix displays. Because that metric development is the only research dealing with overall image quality of dot-matrix displays, it is described in considerable detail here.

The objective of the metric-oriented research was to utilize multiple stepwise regression techniques to derive equations which reliably predict observer performance on representative tasks with dot-matrix displays. The display parameters used in these predictive equations were obtained by Fourier analysis of microphotometric scans of each combination of experimental variables, as was the case for previous research with photographic and television displays.

In addition to developing a quantitative relationship between dot-matrix parameters and observer performance, this reearch also refined the methodology of scanning microphotometry and Fourier analysis of sampled intensity distributions.

The first study was done in two contiguous segments. The first experimental phase yielded observer performance and photometric data from which predictive metrics for three separate tasks were obtained. The second phase of this research attempted to provide some degree of predictive validity for the metrics obtaind in the previous segment.

A rather wide range of within-character parameters was used in the first experimental phase. A completely factorial design in which three shapes, three center-to-center dot spacings, three dot sizes, and two levels of luminance contrast were combined was employed for this stage of the research.

For the subsequent verification phase of the research, three displays built by commercial manufacturers were simulated in as much detail as possible. The three displays chosen were the Burroughs SELF-SCAN II, the Owens-Illinois DIGIVUE, and the prototype Westinghouse TFT EL.

All the characters used in the initial phase of the research were comprised of a dot matrix which was 5 dots wide and 7 dots high. In the verification phase, matrix size was the other independent variable. The three sizes used were 5 x 7, 7 x 9, and 9 x 11. The font was constant throughout the experiment and was based on a study of 5 x 7 dot matrix fonts, as reported in Snyder and Maddox (1978). For the larger matrix sizes, the 5 x 7 font was scaled up as necessary due to the availability of more dots per character, but the general font "style" remained unchanged.

The performance measures were (1) differential reading speed, (2) time to locate the target in a structured search, and (3) time to locate the target in a random, unstructured search. These three measures (or tasks) are representative of the types of activities engaged in by actual users of computer-generated displays. The shapes, sizes, and spacings of the simulated display elements for the metric verification study are indicated in Figure 127.

DIGIVUE
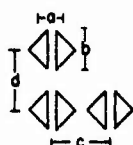
a=.813mm
b=1.240mm
c=1.697mm
d=1.697 mm
VIEWING DISTANCE=
203.2cm

SELF SCAN
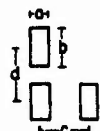
a=.610mm
b=1.067mm
c=1.554mm
d=1.554mm
VIEWING DISTANCE=
101.6cm

TFT
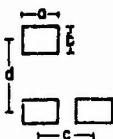
a=.813 mm
b=1.118mm
c=1.600mm
d=2.083mm
VIEWING DISTANCE=
101.6cm

Figure 127: Simulated dot parameters, from Snyder and Maddox, (1978)

The method used to acquire photometric data was identical for both phases of this research and is described in detail in Snyder and Maddox (1978). Further, the approach has been proven useful for a variety of display evaluation applica-

tions, and is described briefly here for that reason. The system is indicated schematically in Figure 128. When a "go" signal was sent to the computer, the scanning drive relay was closed for exactly 60 s, causing the scanning slit to traverse 10 mm during which the analog front end of the computer was sampling the output of the photometer (through the operational amplifier) at a frequency of 100 Hz. At the end of a scan, a file containing 6000 data points taken at equally spaced intervals in time and space existed on magnetic tape.

OPTICAL
CONNECTION

LPS
II

PM TUBE

PHOTOMETER

AMPLIFIER

SCANNING
EYEPIECE

POWER
SUPPLY

'GO' SWITCH
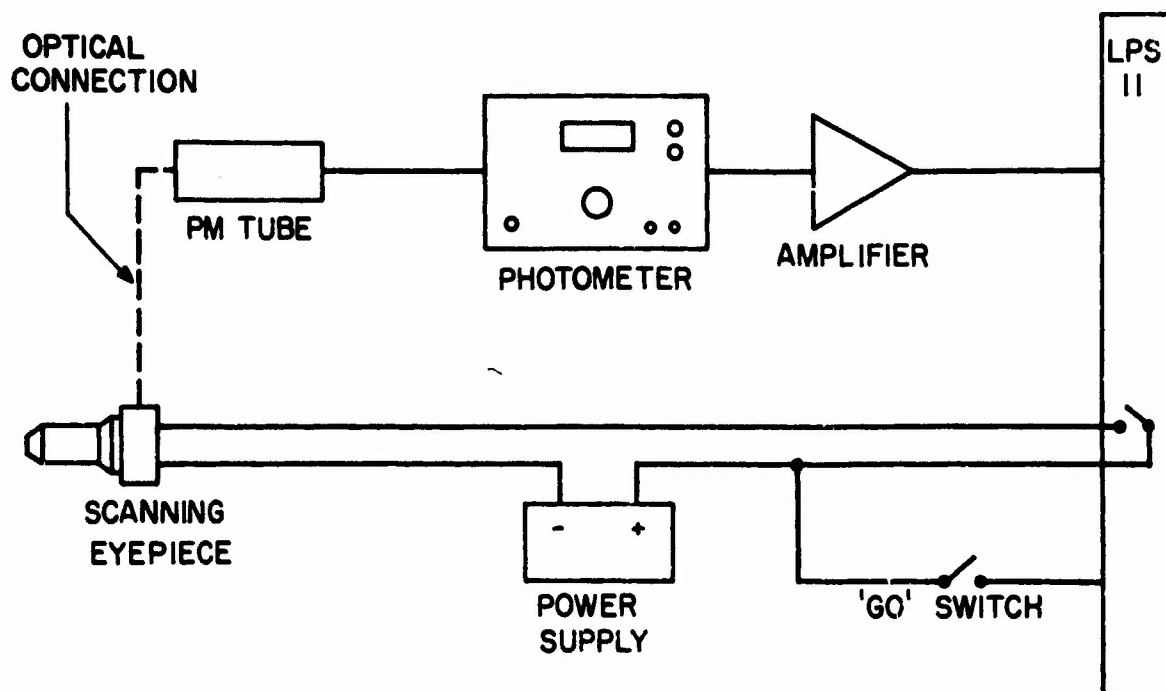
Figure 128: Schematic of photometer interconnection, from Snyder and Maddox, (1978)

Such a scan was taken both vertically and horizontally for every size, shape, spacing, and illuminance combination

- 376 -

used in both phases of this research. These scans were then evaluated by numerical Fourier analysis, which resulted in calculation of the fundamental spatial frequency and the modulation associated with it and with each of its first 19 harmonics. These data were then used to calculate other quantities of interest. Using the contrast sensitivity function of De Palma and Lowry (1962), a pseudo-MTFA was calculated for each scan. This calculation was begun by determining the crossover frequency relative to the CSF. Then, the area between the display modulation and CSF was calculated using numerical straight-line integration. The resulting area is referred to as a "pseudo" MTFA because the display modulation curve is not a transfer function, but rather the modulation of frequencies derived by a specific mathematical technique.

In addition to the pseudo-MTFA, several other scan-related constants were produced by this routine. Two such quantities were used in later analyses, namely the crossover frequency and the frequency range, i.e., the range in cycles/degree from the fundamental to the crossover.

After the photometric values were obtained, scatter plots of these variables with the dependent observer performance variable for each task were generated. This was done to visualize what effect any transformations of variables would have. Since there is no ideal method for choosing proper predictor variables for empirical curve fitting, variables

were selected which, on the basis of past and present research, should account for reasonable proportions of measured variance. In addition to research-based variables, variables transformed to fit observed data patterns were used.

Using both approaches, a total of 20 variables was eventually used as a pool of regression predictor variables. These 20 variables were divided between vertical and horizontal terms. All predictor variables used are listed and defined in Table 27.

Using stepwise multiple regression (SMR) procedures, three metrics (equations) were derived to predict the observer performance data obtained in the first experimental phase. These metrics are presented in Table 28, along with the proportion of observed variance accounted for by that particular model (R^2), the maximum R^2 if all variables are entered into the model, and the correlation coefficient of that model with the observed data (R).

Utilizing the metrics described above from the first phase and photometric data from the phase two simulated display types, predicted performance means were calculated for each task. These were calculated from 5 x 7 matrices only, since the actual photometric values do not change with different matrix sizes. The predicted and observed performance measures are shown in Table 29.

TABLE 27

Pool of Predictor Variables, from Snyder and Maddox (1978)

Vertical	Horizontal	Description
VFREQ	HFREQ	Fundamental spatial frequency (cyc/deg)
VFLOG	HFLOG	Base 10 log of fundamental spatial frequency
VSQR	HSQR	Square of (fundamental spatial frequency minus 14.0)
VMOD	HMOD	Modulation of fundamental spatial frequency
VDIV	HDIV	Fundamental spatial frequency divided by modulation
VLOG	HLOG	Base 10 log of VDIV and HDIV
VMTFA	HMTFA	Pseudo-modulation transfer function area
VMLOG	HMLOG	Base 10 log of VMTFA and HMTFA
MCROS	HCROS	Spatial frequency at which modulation curve crosses the threshold curve
VRANG	HRANG	Crossover frequency minus fundamental frequency

Several things can be seen from the table of predicted versus actual data. The most obvious discrepancy between measured and predicted values occurs for the DIGIVUE simulation in all tasks. The probable explanation for this mismatch is that the fundamental spatial frequency for the DIGIVUE is approximately 22 cyc/deg. The range of spatial

- 379 -

TABLE 28

Predictive Equations, from Snyder and Maddox (1978)

Tinker SOR:

Adjusted Reading Time (s) = 1.43 + 0.023(VSQR)
+ 0.364(HMTFA) + 0.221(VMTFA) - 4.825(HMLOG)

Correlation Coefficient R = 0.76

$R^2 = 0.573$

Asymptotic $R^2 = 0.70$

Menu Search:

Search Time (s) = 0.78 + 0.024(VSQR) + 2.72(HLOG)
+ 0.193(VMTFA)

Correlation Coefficient R = 0.69

$R^2 = 0.471$

Asymptotic $R^2 = 0.59$

Random Search:

Search Time (s) = -48.50 - 138.49(HFLOG)
+ 192.89(VFLOG) - 0.642(HMTFA) - 0.734(HSQR)
+ 0.982(VSQR) - 0.043(HDIV)

Correlation Coefficient R = 0.71

$R^2 = 0.499$

Asymptotic $R^2 = 0.60$

---

frequencies used in the regression procedure which yielded
the predictive metrics was 4-17 cyc/deg. The DIGIVUE ele-
ments are clearly out of this range and regression equations
are often unpredictable in such extrapolated regions. The
correlation between predicted and actual means for all three

TABLE 29

Predicted and Measured Performance Data, from Snyder and Maddox (1978)

Task	Type	DIGIVUE	SELF-SCAN	TFT
Tinker SOR	Predicted	3.08	1.40	1.71
	Measured	1.65	1.67	1.53
Menu Search	Predicted	8.15	4.53	5.89
	Measured	5.09	4.97	4.90
Random Search	Predicted	35.91	3.58	-11.78
	Measured	5.45	5.82	5.93

tasks and all three display types is 0.16, while the Spearman rank-order correlation between predicted and actual means, excluding the random search data, is 0.73.

The other anomaly in the predicted data is the negative value predicted for the TFT random search measure. This type of task historically produces quite variable data, since human performance is so dependent on individual factors such as search strategy and set. (The original performance data used to derive the metric for random search had greater variance than the data from either of the other two tasks.)

The verification study revealed that the original metrics were poor predictors of subject performance on the DIGIVUE

display due to the restricted range upon which the metrics were based. In an attempt to eliminate this shortcoming, an extended prediction model was derived for Tinker speed of reading (SOR) and menu search performance.

To extend the range of the original prediction equations, it was necessary to include photometric and performance data from higher spatial frequency displays in the pool of regression variables. The improved equation generation was accomplished exactly as it was for the original metrics. The same pool of predictor variables was subjected to step-wise linear multiple regresion analysis. This time, however, the data submitted for analysis originated from the following studies of Snyder and Maddox (1978):

1. The original dot-matrix study upon which the first metrics were based,

2. The simulation study which was used as a pre-dictive validation of study (1), and

3. A later verification study which used real dis-plays instead of simulation.

Sources (2) and (3) contained data from displays having fun-damental spatial frequencies in the 20-30 cyc/deg range.

The data pool allowed the generation of extended metrics for the Tinker SOR task and the menu search task. It was not possible to generate an equation for the random search

task, since no performance data were taken for this particular task in the verification study. The resultant prediction equations and their R^2 values are shown in Table 30. The meaning of each variable name is the same as for the original equations.

TABLE 30

Extended Predictive Equations, from Snyder and Maddox (1978)

Task	Metric and Related Information

Tinker SOR:

Adjusted Reading Time (s) = 5.74 + 0.3111(HFREQ)
+ 2.479(HMOD) + 4.365(HLOG)
- 14.973(HFLOG) + 1.112(VMLOG)

Correlation Coefficient R = 0.72

R^2 = 0.525

Asymptotic R^2 = 0.637

Menu Search:

Search Time (s) = 7.27 + 0.027(HDIV) + 2.159(HLOG)
+ 5.916(VFLOG) - 0.339(VMTFA)
- 0.054(VRANG) + 5.487(VMLOG)

Correlation Coefficient R = 0.71

R^2 = 0.500

Asymptotic R^2 = 0.575

Several features of the extended prediction equations should be noted. First, and perhaps most important, these equations apply to dot-matrix displays in which the fundamental spatial frequency of the repetitive dot pattern falls between 4-30 cyc/deg. This virtually doubles the usable frequency range of the equations.

The next notable thing about the extended equations is that the range doubling was accomplished with very little loss of correlation between observed performance and performance predicted by the regression equations. The extended equations correlate 0.72 and 0.71 with observed performance. These compare closely with the 0.76 and 0.69 correlations, respectively, obtained in the original equations.

These extended predictive metrics are believed to be very good predictors of relative observer performance using a wide variety of dot-matrix displays. As such, they represent the best available empirically derived measures of dot-matrix image quality. They should not, and cannot, be applied to non-dot matrix characters, nor to dot matrices the fundamental spatial frequency/modulation of which is below visual threshold. Thus, for example, a double-electrode prototype Owens-Illinois DIGIVUE, with 23.6 (double) dots per centimeter cannot be analyzed by these equations simply because the fundamental spatial frequency is 59 cyc/deg (47 dots/cm at 71 cm viewing distance), which is

below visual threshold at the displayed modulation of those dots.

Informal verification of these metrics has been attempted against Tinker SOR data taken with other displays at a different laboratory. The predictions, from photometric scans of those displays, predicted the subjects' mean performance within 10%.

Clearly, these predictive equations are quite useful. At the same time, it sould be noted that they are empirically derived and have no basis in visual theory such as the models proposed for continous and semi-continuous displays, except for CSF concepts contained in the variables. More work is thus needed to relate current visual theory to quality metrics of two-dimensional discrete displays. In the meantime, we appear to have a pair of valid prediction equations for two generic operator tasks.

5.4.4 Chromatic Displays

In spite of the plethora of research and modeling devoted to monochromatic displays, there appears to be no proposed metric of image quality devoted to chromatic displays. One might expect that chromatic quality metrics might have been derived from such metrics as the least-squared deviations from "true" color, as is used in monochrome image processing concepts, or even from an MTFA-type approach for each of the three color primaries plus luminance for a color TV system. Such is not the case.

There have been studies of the influence of individual chromatic display parameters on both subjective quality estimates and upon some simple observer performance tasks, as noted above. But apparently no effort has been made to develop an all-inclusive model of all the chrominance/luminance variables in a complex display. Thus, there does not seem to exist a valid metric of color contrast, let alone a metric of color image quality. Such a metric needs to be developed, although the task currently appears ominous.

5.5   INFORMATION TRANSMISSION

The previous section stressed composite or unitary image quality metrics which apply to pictorial images, such as television or photographs, and to alphanumeric images formed on either raster scanned displays or discrete dot matrix displays. These metrics deal with inherent photometric and geometric characteristics of the display and its ability to present an image commanded to the display. They do not deal with the selection of various properties of the commanded, or input, image, at least to any significant extent. Thus, the metrics of image quality largely disregard such input variables as alphanumeric font, matrix size, symbol encoding, stroke segmentation, etc. They also would be insensitive to input variables of graphical displays, such as shading levels, rate of image movement, hidden line requirements, and edge "scalloping."

In this section we shall briefly touch upon some of these information input variables, particularly as they might be pertinent to the design and selection of flat panel displays. This discussion is not intended to be a complete dissertation on display encoding, or principles of visual esthetics, for entire books have been written on those subjects. Rather, the limited discussion should acquaint the reader with a few of the problems in information presentation which are unique to flat-panel two-dimensional discrete displays.

5.5.1 Alphanumeric Displays

If one assumes that a flat-panel display is fixed in its contrast range and in its element geometry, then the number of decisions needed to display alphanumeric information has been significantly reduced. Unfortunately, generalization from existing design guidelines, which apply to and were developed for continuous image displays, is not straightforward. As Maddox, Burnette, and Gutmann (1977, p. 89) indicated:

> The "dot" matrix characters used in the computer output systems are quite different in appearance from their conventional "stroke" counterparts inthat stroke characters are composed of continuous line segments. It has been recognized for some time that certain characteristics of stroke alphanumeric characters affect their relative legibility. Much research has been undertaken to ascertain which stroke font is the most legible under certain conditions. However, it has not been satisfactorily demonstrated that the conclusions from stroke font research are directly transferable to dot-matrix fonts.

Typically accepted design criteria for stroke-written alphanumerics are the following:

Strokewidth-to-height ratio: 1:6 to 1:8 for black on white, and 1:8 to 1:10 for white on black,

Font (character style): NAMEL (or AMEL) for poor visibility conditions; Lincoln/Mitre also acceptable,

Width-height ratio: 3:5 generally, but up to 1:1 for upper case letters is acceptable,

Character height: 12 arcmin for noncritical characters and 24 arcmin for critical characters or under adverse reading conditions, and

Contrast: 80% modulation; reduction in contrast requires an increase in size.

If one attempts to apply these criteria to dot-matrix displays, however, the extrapolations are substantial and unwarranted. Specifically, "strokewidths" will be limited to integer multiples of element or dot sizes; fonts can only be constructed in accordance with the discrete element locations; width-height ratios will be constrained by element sizes and element requirements for font presentation; character heights will be driven by element sizes and software addressing tradeoffs; and contrast will be governed by the contrast available at the display under the given ambient illuminance.

To illustrate but one example, Figures 129 through 132 comapare a stroke-written Lincoln/Mitre alphabet with a dot-matrix constrained "Lincoln/Mitre" alphabet consisting of 5 x 7, 7 x 9, and 9 x 11 element matrices.

A B C D E F G H I J K L M N O P Q R

S T U V W X Y Z 1 2 3 4 5 6 7 8 9 0

Figure 129: Lincoln/Mitre stroke written font

To provide some design criteria for dot-matrix addressed alphanumeric displays, Snyder and Maddox (1978) compared four different alphanumeric fonts at five different sizes for single character legibility. The four fonts were the Lincoln/Mitre, Maximum Angle, Maximum Dot, and Huddleston. The Lincoln/Mitre is illustrated in 5 x 7, 7 x 9, and 9 x 11 matrix sizes in Figures 130 through 132, while the other three fonts are illustrated in Figures 133 through 141. The

Figure 130: Lincoln/Mitre 5 x 7 dot matrix font
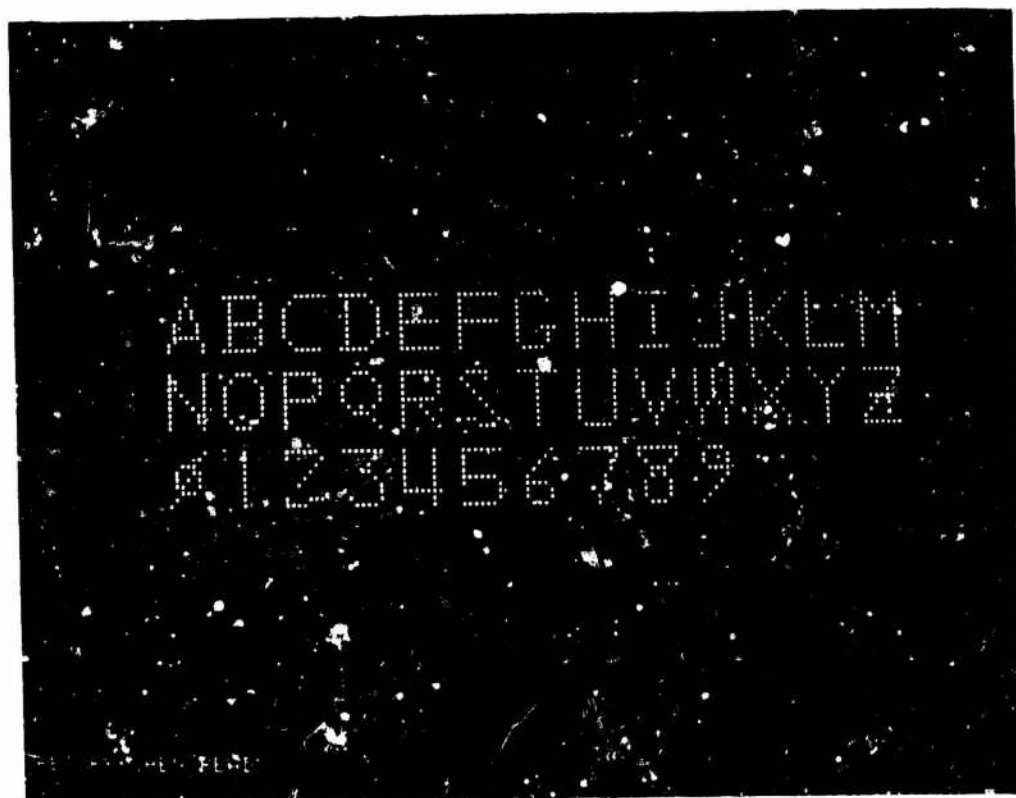
Figure 131: Lincoln/Mitre 7 x 9 dot matrix font
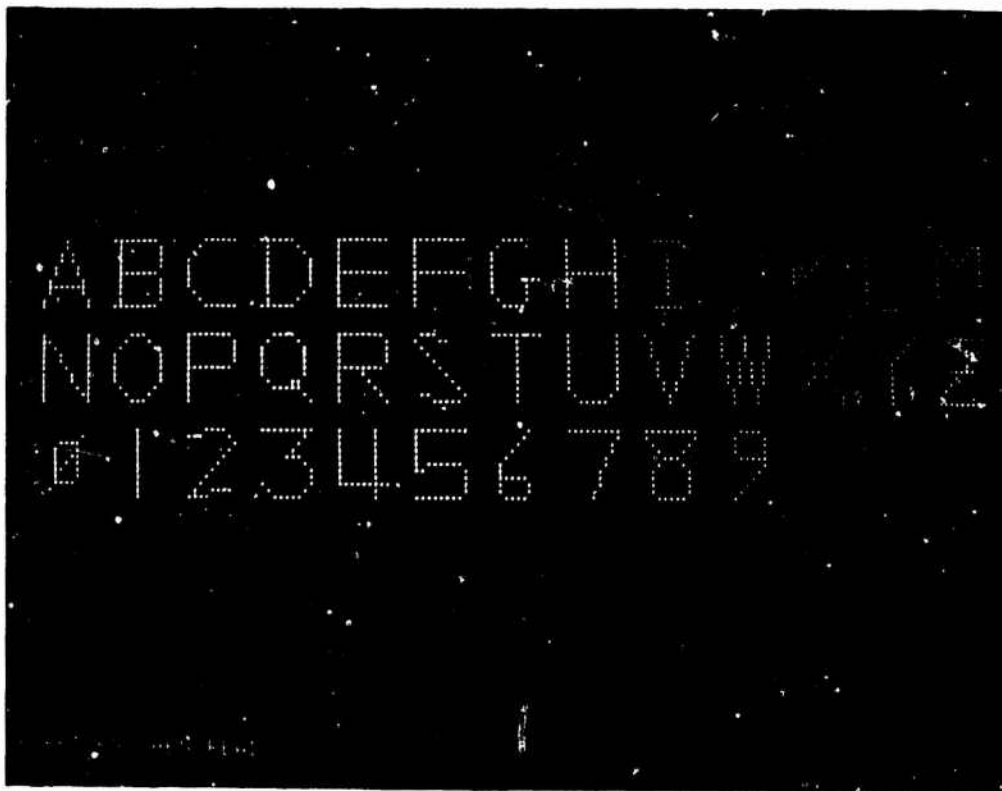
Figure 132: Lincoln/Mitre 9 x 11 dot matrix font

Huddleston font was designed to maximize legibility under high ambient illuminance, while the Maximum Dot and Maximum Angle fonts were designed in accordance with theoretical vision perception principles (Maddox et al., 1977). Snyder and Maddox (1978) experimentally separated matrix size (5 x 7, 7 x 9, 9 x 11) from character size by also evaluating 7 x 9 and 9 x 11 matrix sizes which were equal in height to the 5 x 7 matrix size (Figure 142).

The results of their experiment are shown in Figures 143 through 145. Averaged across all matrix and character sizes, the Lincoln/Mitre and Huddleston fonts, while not significantly different from one another, were both superior to the Maximum Dot and Maximum Angle fonts, as seen in Figure 143.

Figure 133: Maximum angle 5 x 7 dot matrix font

Figure 134: Maximum angle 7 x 9 dot matrix font

Figure 135: Maximum angle 9 x 11 dot matrix font

Figure 136: Maximum dot 5 x 7 dot matrix font

Figure 137: Maximum dot 7 x 9 dot matrix font

Figure 138: Maximum dot 9 x 11 dot matrix font

Figure 139:   Huddleston 5 x 7 dot matrix font

Figure 140:   Huddleston 7 x 9 dot matrix font

Figure 141: Huddleston 9 x 11 dot matrix font
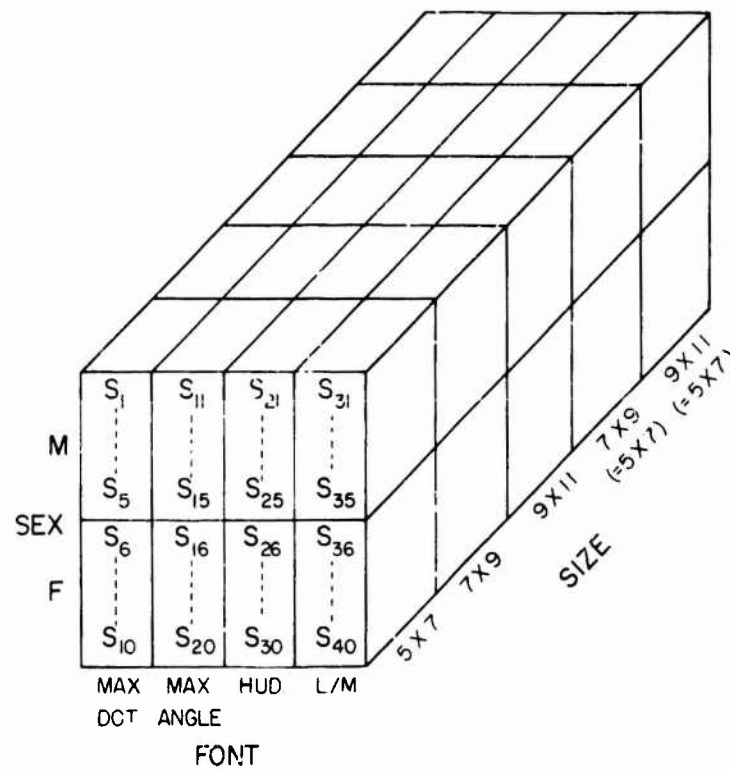
Figure 142: Experimental design for Snyder and Maddox
(1978) font study
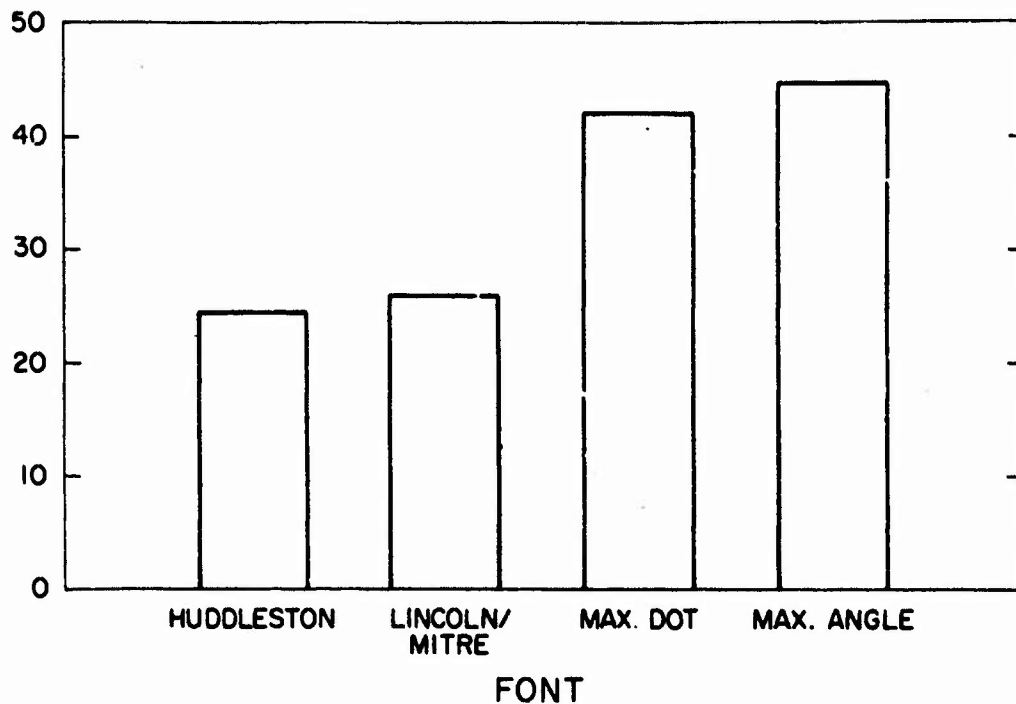
Figure 143: Effect of font on single character legibility, from Snyder and Maddox (1978)

Figure 144 shows that, averaged across all fonts, the 5 x 7 matrix size produced more errors than any of the other sizes. As the matrix size was increased to 7 x 9 and to 9 x 11, significantly fewer errors occurred.

Of considerable interest is the fact that the smaller 7 x 9 matrix size led to significantly fewer errors than the larger 7 x 9 matrix. A similar result obtained for the two 9 x 11 matrix sizes. (The angular subtenses of the characters were 48.5 arcmin for the 5 x 7, 7 x 9 (= 5 x 7), and 9 x 11 (= 5 x 7); 63.0 arcmin for the 7 x 9; and 77.2 arcmin for the 9 x 11.) In general, the results indicate that (1) the larger the number of dots in the matrix, the better the legibility, and (2) the smaller the character, within these
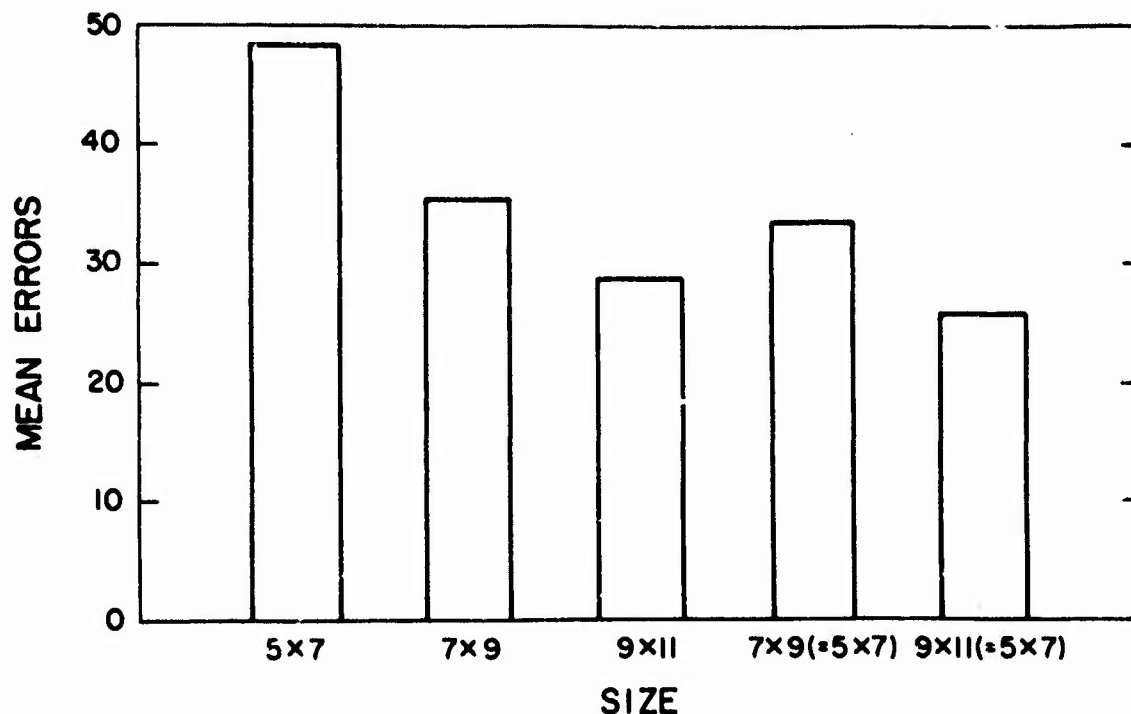
Figure 144: Effect of character/matrix size on single character legibility, from Snyder and Maddox (1978)

better the legibility. Even the smallest character size was well above the 24 arcminutes recommended for adverse viewing conditions. Thus, the main benefit of reducing the character size was to proportionally reduce the space between the elements, an advantage discussed previously in Section 5.1.3.

The most important result from this experiment, however, is shown in Figure 145. At the 5 x 7 matrix size, the Huddleston font produced fewer errors than did any of the other fonts, while for the other four sizes there were no significant differences between the Huddleston and the Lincoln/Mitre fonts. Thus, for the 7 x 9 and 9 x 11 matrix
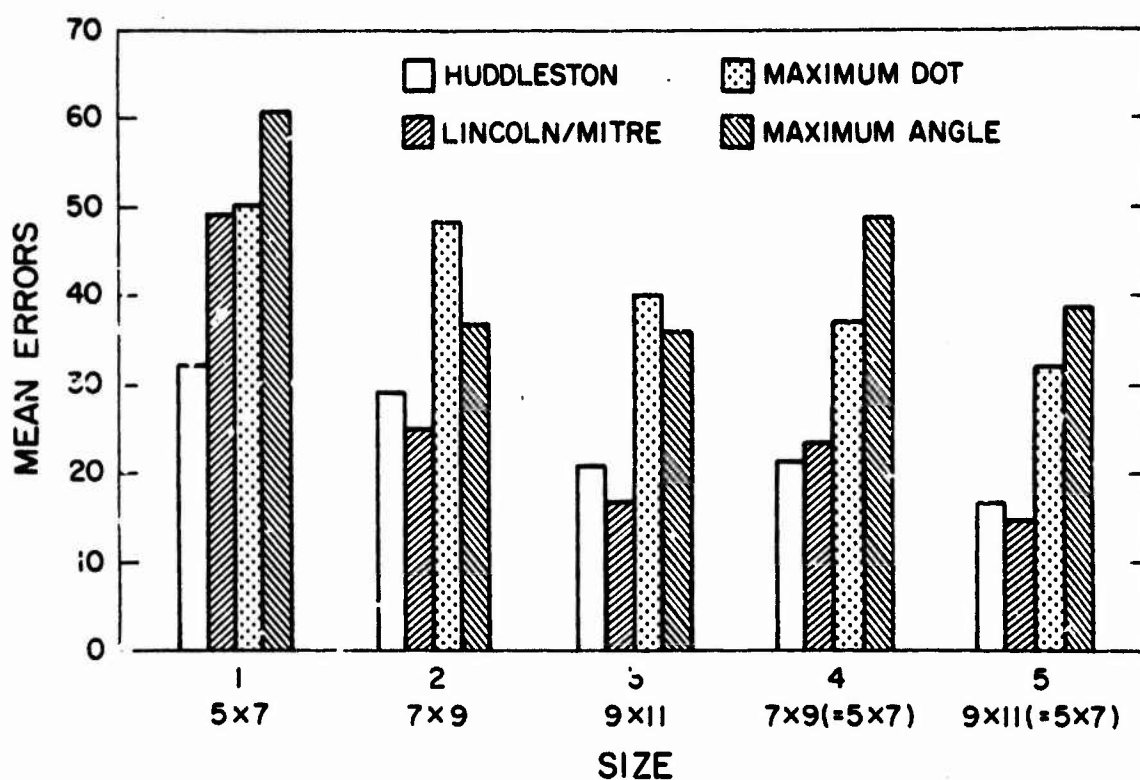
Figure 145: Interaction of font and character/matrix size
in determining single character legibility
(Snyder & Maddox, 1978)

sizes, the Lincoln/Mitre and Huddleston fonts are essentially equivalent.

It should be noted that this research examined only single character presentation legibility, and did not include lower case letters or symbols. In view of the nonintuitive results, further research on lower case letters, symbols, and superscripts and subscripts is needed. Further, tasks other than single character recognition should be studied.

The prediction of character legibility, to avoid lengthy empirical investigations of the type described above, has

been attempted by Suen and Shiau (1980) and Maddox (1980).
Suen and Shiau developed an iterative, multi-rule process
for maximizing character differences, but presented no human
performance data to support their conclusions. Maddox
(1980), using the results presented above, attempted a post
hoc prediction using both a two-dimensional Fourier analysis
and a digital phi coefficient. The phi coefficient led to
better prediction than did the two-dimensional Fourier, and
for logical reasons. The mean correlation (r) between legi-
bility errors and the Fourier spectrum was 0.14, while the
mean correlation for the digital phi prediction was 0.58.

5.5.2 Graphical Displays

The concept that "a picture is worth a thousand words"
has led in part to a rapid proliferation of computer-gener-
ated graphical displays. Indeed, an entire "computer graph-
ics" industry has been born and reached adolescence during
the past 15 years. As one might predict, the current capa-
bilities in computer graphics are constrained only by the
imaginations of the programmers, and by the hardware and
software device limitations, which change weekly. Rarely in
this dynamically growing industry does one see any signifi-
cant concern with the visual requirements and capabilities
of the user, a conclusion substantiated by only two opera-
tor-relevant printed papers out of three days of parallel
sessions at the 1979 SigGraph meeting. Further, little

empirical research has been dedicated to investigations of the effects of display parameters upon the legibility, perceived quality, workload, etc., of the user.

In this section, we touch briefly on this problem, mostly to define it, indicate an example or two of how visual theory can be applied to computer graphics, and to identify existing gaps in our knowledge.

The nature of the computer graphics/visual system interface problem was summarized well by Montalvo (1979, p. 121), who indicated that

> The very reason for the rise in the use of graphics in the first place is deeply imbedded in the way humans perceive. Graphs and pictures convey more information more quickly than pages of numbers. We know all this intuitively. Few of us, however, stop to consider why.

5.5.2.1 Line and edge problems

The geometric constraint of fixed element sizes and positions that affects alphanumeric legibility also has an effect on graphics displays. As indicated in Figure 146, a diagonal line one or two elements wide cannot be placed accurately and with smooth edges on a discrete element display unless the line is perpendicular or parallel to the element rows. In the nonorthogonal case, elements turned "on" to best charcterize the line's location will cause a jaggedness to the edge of the line. This jaggedness, or stairstepping, is most noticeable as the angle of the line approaches either the vertical or the horizontal. That is,

the jaggedness is nonexistent at 0 deg or 90 deg to the element row, is minimal at 45 deg, and is maximal <u>near</u> 0 deg or 90 deg.. The larger the display elements or the greater their separation, the more noticeable is the jaggedness. In computer graphics language, "aliasing" is the error caused by a discrete element approximation to a smooth image bound- ary, as in Figure 146. Similarly, "staircasing" is the error caused by a discrete element approximation to a curved surface (Montalvo, 1979). The visual system is particularly sensitive to these geometric distortions, for reasons well predicted by visual theory.
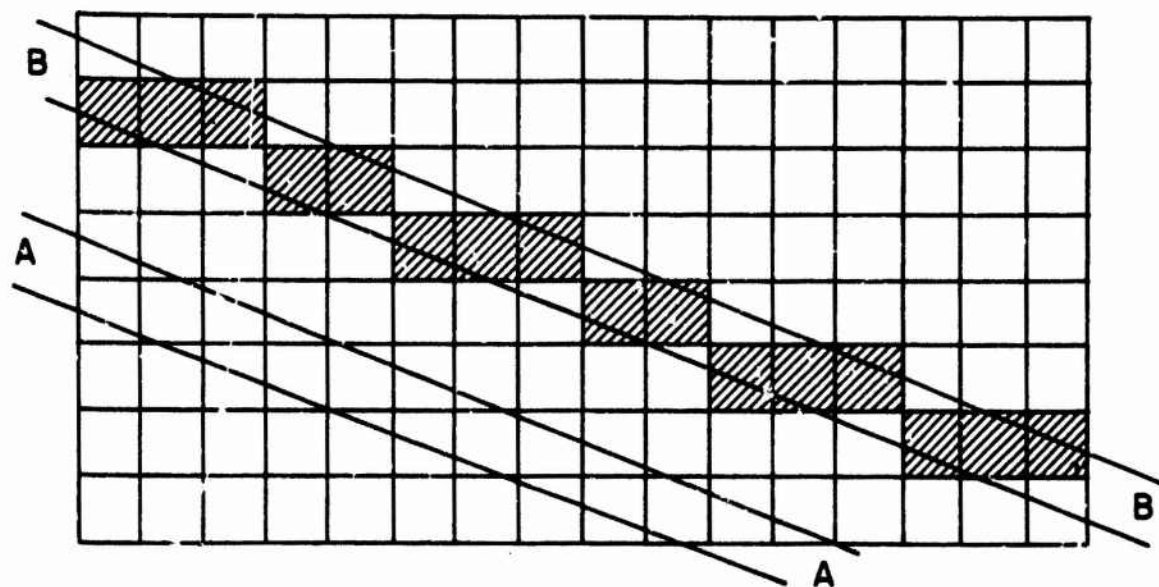
Figure 146: Line A is true line position, while shaded elements associated with parallel line B indicate irregularities

- 402 -

The aliasing problem has been known to physicists for many years. Following Nyquist sampling theory, it occurs when the highest spatial frequency in the image being sampled exceeds one-half of the sampling spatial frequency. In this case, the sampling frequency is simply the element density at the display. Since we are usually "sampling" sharp-edged lines in computer graphics, lines which ideally have a cross-sectional spatial frequency spectrum with higher harmonics, the sampling frequency is rarely greater than twice the fundamental frequency of the line, and certainly never twice the spatial frequency of its higher order harmonics. Thus, aliasing occurs physically at the display surface.

Aliasing is a particularly annoying or distracting consequence, although no behavioral research is known to relate the extent of line aliasing to operator performance in even the simplest of tasks.

Increasing display element density can reduce the perceived aliasing, but it may not necessarily eliminate the aliasing problem. Aliasing appears more pronounced, of course, as the edge gradient of the display element becomes steeper. One attractive solution to aliasing is to let each display element represent an average of a corresponding area in the input image. Thus, elements which are only partially intersected by the input line would be turned "on" partially, proportional to the weighted intensity average of

the corresponding area in the input. Similar weighting can be done with in the color domain, by reducing saturation to correspond to weighted area averaging (Crow, 1977).

Thus, a diagonal line as in Figure 146 can cut through two or three adjacent elements. If the center element, or two or three, are turned on, jaggedness results. If, however, the two or three elements are set at the proper shades of gray, the jagged effect is greatly reduced or disappears.

The implementation of such antialiasing algorithms can be costly. The intensity levels of many more elements must be calculated and then commanded to the display. Several times as much computer memory is typically needed.

In addition, some technologies lack adeqate gray scale capability and are therefore not compatible with the above area averaging approach. In such cases, temporal modulation can be used, in which each element is turned on and off rapidly and repeatedly such that the commanded intensity controls the "on" duty cycle of the element.

Another antialiasing technique, which is also used to generate shades of gray on binary ("on"-"off") displays is ordered dither, a means by which a small group of elements (e.g., 4 or 9) is assigned different proportions of "on" or "off" levels, the visually resulting intensity being proportional to the number of "on" elements in the group. Judice (1976) points out that this technique, like half-tone newspaper photography, takes advantage of the low-pass spatial

filter characteristic of the visual system, in essence caus-
ing the eye to average intensity levels which change at a
spatial frequency higher than the upper cutoff of the CSF.

The author knows of no published research to predict
quantitatively the perception of jaggedness of graphics
lines by discrete element sampling. However, one brief
experiment, conducted in the VPI laboratory in 1979, demon-
strates precisely a possible technique for such prediction.
In this experiment, circular black dots were partially over-
lapped to print a straight line (Figure 147). The degree of
overlap varied from 10% to 60% of the dot diameter. Less
overlap produced more edge scalloping.

10% OVERLAP
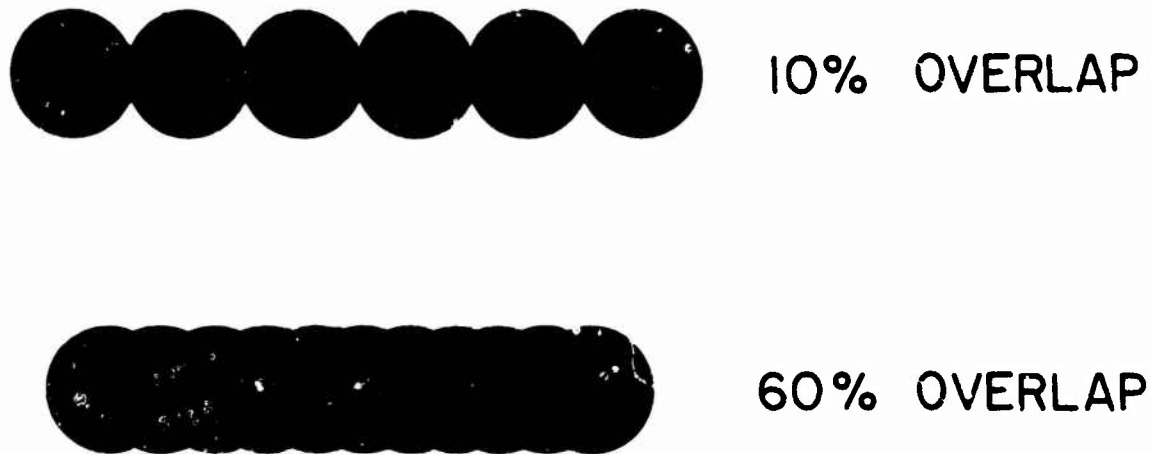
60% OVERLAP

Figure 147: Schematic of dot overlap patterns

Subjects were asked, in a series of approaching and receding trials, to determine the distance at which they could barely detect the edge irregularity. The lines were then scanned with a microphotometer slit, the length of which was exactly equal to the dot diameter and the width of which was 1/100 times the length. Smaller dot overlaps of course produced greater modulation in the photometer output.

The analysis technique of the photometer tracings was similar to that described in Section 5.4.3. These tracings were Fourier analyzed and the modulation at the fundamental spatial frequency of each overlap level was determined. The spatial frequency fundamentals were calculated in angular units for the threshold distances obtained in the psychophysical portion of the experiment. Figure 148 shows these modulation/spatial frequency points for all six dot overlap levels, plotted with the De Palma and Lowry (1962) contrast sensitivity function. The figure also indicates the calculated modulation/spatial frequency for each dot pattern if the dot pattern, in this case for a hard copy line printer, were presented at a "normal" reading distance.

As indicated by Figure 148, the threshold modulation for each dot overlap is above the CSF, probably a result of the threshold criteria used by the subjects, that is, barely detectable. The "normal" viewing distance points appropriately lie below the CSF; in fact, the edge irregularities were not detectable at this viewing distance for any of the
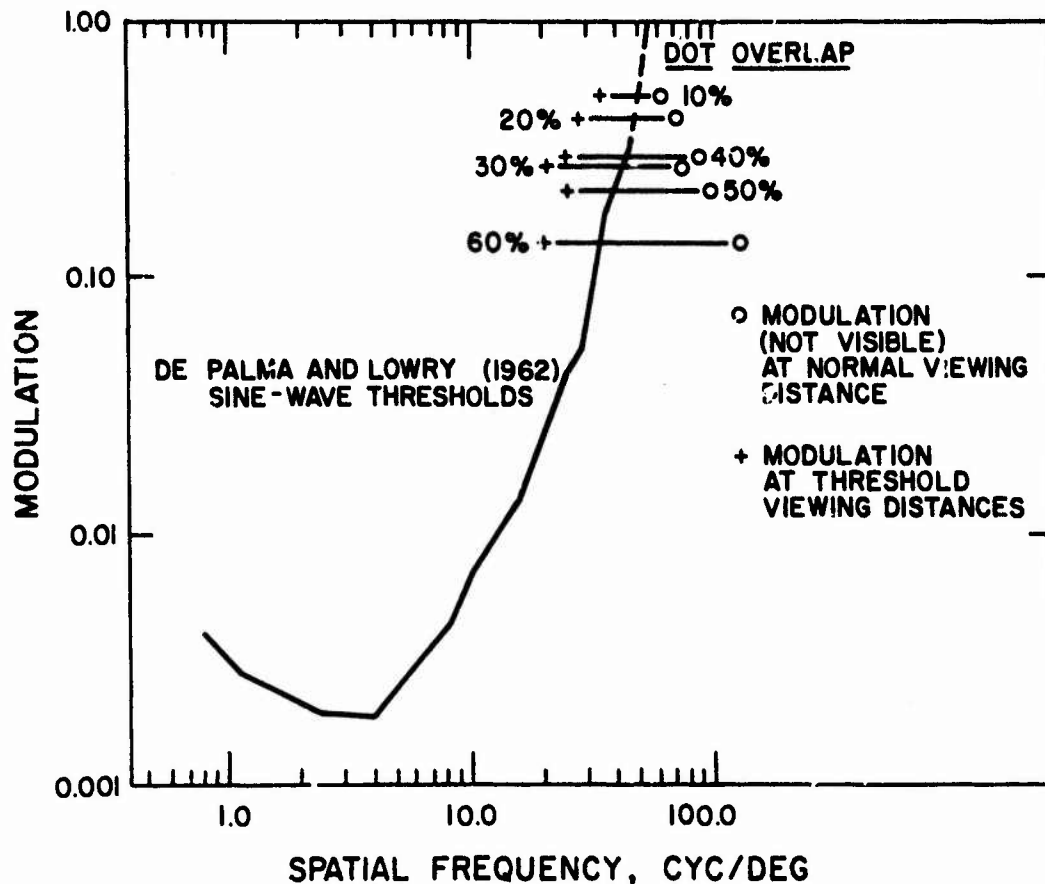
Figure 148: Prediction of dot-overlap pattern thresholds from photometric scanning and contrast sensitivity function

overlap patterns. It seems quite reasonable to conclude that photometric scanning techniques, followed by Fourier analysis to determine the amplitude and frequency of the Fourier fundamental, can be used to predict the existence of visual "jaggedness" of lines on discrete element displays. Thus, the concepts described in earlier sections seemingly apply to computer graphics concepts as well. Further research is severely needed, but these very preliminary results are encouraging.

5.5.2.2 Color graphics

Color graphic displays most commonly use the red, green, and blue P22 phosphors on a CRT, addressing anywhere from 8 colors (two levels per color gun in all combinations) to 10 intensity levels per color, or $2^{30} = 1.07 \times 10^9$ combinations. The former number of levels is below maximum observer absolute judgment capabilities, while the latter number is well beyond human absolute discriminability, and most likely beyond relative discriminability within the color space available to the three phosphors.

Of course, as has been noted in Section 5.3, chromatic constrast is not quantitatively defined and threshold separations among colors vary throughout the visible color space. Discriminati n of hue differences, for constant brightness, is more acute near yellow than near blue, perhaps because of the sharp drop in sensitivity of red- and green-sensitive cones. Thus, color graphics must take into account color sensitivities in the encoding and display process. Precise equal-sensitivity mapping of the available color space for color computer graphics must await completion and validation of a color contrast metric.

Some color graphics guidance is immediately available, however. Knowing that the CSF for color is characterized by a lower frequency response than for luminance information (Section 4.4), we know that aliasing will not be as pronounced for chromatic edges varying in hue but not in lumi-

nance. Further, edges varying in hue or saturation, but not in brightness, will result in even less apparant aliasing. Thus, chromatic edges need not be antialiased as much as edges varying only in luminance. However, empirical research to define quantitatively these antialiasing requirements must still be conducted.

5.5.2.3 Shading, hidden lines, rotation, and other graphics techniques

As stated above, the image concepts on the computer graphics display are limited largely by the ingenuity of the programmer. Algorithms abound for generating complex images or scenes, for rotating images, for calculating and suppressing "hidden" lines and faces, for shading, and for "moving" the assumed illumination source. Unfortunately, virtually no research has attempted to relate, systematically and quantitatively, the perceived results of these processing algorithms to the perceptual needs of the human observer. Such research should be conducted soon and extensively to provide guidance for technology advances in computer graphics and to reduce the amount of fruitless software development.

Section 6

SUMMARY AND CONCLUSIONS

6.1   SUMMARY

The application of linear systems analysis to research in
visual psychophysics has permitted the development of a
means by which complex displays can be evaluated against
visual performance criteria. Spatial, temporal, and chro-
matic contrast sensitivity functions can be used to assess
the detectability of displayed information as a function of
spatial, temporal, and chromatic characteristics of the dis-
play device.

Furthermore, existing research data indicate how the CSF
shifts due to changes in ambient illuminance, adapting lumi-
nance, viewing time, display motion, and other related vari-
ables. Thus, the CSF selected for display evaluation may be
chosen to be that appropriate for the task and .environmental
conditions.

Finally, the CSF and other predictor variables have been
used in combined, unitary measures of image quality which
attempt to predict observer performance as it is influenced
by a wide variety of display design characteristics. Some
of these unitary metrics are reasonably successful, while
others have high predictive validity only under limited cir-
cumstances.

While the unitary metrics may not have reached the desired level of utility, and the data base for CSF application to display evaluation is only partially complete, there exists nonetheless a very strong data base by which the human factors/systems engineer can evaluate or select flat panel display designs. These data were sampled in Section 5 and combined with the flat panel technology survey of Section 3 to demonstrate a design evaluation approach, as summarized for the two examples of Section 2. While this design evaluation approach is necessarily presented only briefly in this report, it is hoped the reader can modify the approach and select the supporting data/theory which applies to his particular problems.

Throughout this report, data gaps and research needs are indicated. For convenience of the reader, the most urgent research needs are summarized here.

6.2   DATA GAPS AND NEEDS

6.2.1   Uniformity Data Needs

As indicated previously, there is very little in the literature to advise us of minimum requirements for the uniformity of visual displays. No studies are known to provide either thresholds of detection for, or tolerance limits for, large-area nonuniformities. In general, we simply do not know how much large-area nonuniformity is a reasonable design goal.

- 411 -

The case for small-area nonuniformity is similar. Unless one applies the basic sine-wave sensitivity data to a given form of small-area nonuniformity distribution, and attempts to predict the detectability of nonuniformity, there is not even a currently suggested means for evaluation.

Finally, except for the initial study by Riley and Barbato (1978), we have little knowledge of the effects of line errors (on or off) or of element errors (on or off) on display legibility and utility. For these reasons, research efforts to fill the data void are needed.

6.2.1.1 Large area uniformity research

Thresholds need to be determined for large-area nonuniformity, and the effect must be determined of various levels of large area nonuniformity on representative observer tasks, such as character legibility, reading, search, and image interpretation. Variables which should be investigated include display angular subtense (to the observer), mean display luminance, degree of nonuniformity (maximum drop in luminance), and shape of the luminance gradient. The drop in uniformity across the display should be both continuous, as would occur in a CRT, and discrete, as in a matrix-addressed, discrete element display.

6.2.1.2   Small area uniformity research

In a related or separate set of experiments, studies are needed to determine the thresholds for small-area nonuniformities, such as would occur with dimming or brightening of discrete elements. In addition to threshold determination, such studies should determine the impact of small-area nonuniformities on representative observer tasks. Variables to be investigated should include at least the following: increase (or decrease) in luminance of individual elements from neighboring elements, overall mean display luminance, number of elements having an increase or decrease in luminance, distribution of aberrent elements in the display, and number of aberrent elements in the display.

6.2.1.3   Line, cell loss (on and off)

Some technologies use line-at-a-time addressing, and are therefore likely to miss an entire line of the display. In such cases, an entire line could be set at an "on" level or at an "off" level. In addition, single (or multiple) lines can, in some technologies, be set at intermediate luminance levels, giving rise to gray level stripes in the display. Edge-butting lines are particularly likely, as matrix sizes grow by joining smaller matrices into larger arrays. Such edge-butting lines are currently noticeable in several LCD arrays.

The effect of line errors of these types should be studied empirically. Again, both threshold and performance measures should be taken, and related to at least the following design variables: number of lines in error; direction of lines in error; luminance error, positive or negative; display size; and mean display luminance.

6.2.2 Chrominance/Luminance Contrast Tradeoffs

There is a wealth of literature on luminance contrast thresholds and of the effect of luminance contrast on representative observer visual tasks, such as reading, legibility, tracking, search, etc. There is also a large amount of data on color discrimination, using various color systems. Most of the threshold data are not easily extrapolated to suprathreshold tasks, and extrapolation to high performance levels, such as 99% legible, is extremely unsafe from the existing data. Most importantly, there is little information, of a _quantitative_ nature, that relates observer performance in suprathreshold tasks, to the physically measured colored display. That is, most such studies compare "red" with "green," but do not measure, spectrophotometrically, what is meant by red or green. Thus, application of the data to narrow wavelength dispersion displays, or even to wide dispersion emissions, is inexact or inappropriate. Finally, there are no empirical laboratory data which attempt to relate, quantitatively, both chrominance and

luminance contrasts. Thus, we cannot evaluate the relative
merit of displays which differ in both chrominance and lumi-
nance, even though there are many technologies which cur-
rently offer that option (e.g., CRT, LED, LCD, EC, EPID).
For some applications, especially under high ambient illumi-
nance, this combined contrast measure may be of utmost
importance to cost, operating life, and operator perform-
ance.

6.2.2.1   Metric of color contrast

A fundamental requirement of a research program in this
area is to develop a meaningful measure of color contrast,
one that takes into account both the luminance contrast in a
visual task and the chrominance contrast. Because the con-
ventional CIE color system is arbitrary, and certainly nonu-
niform in wavelength discrimination, one must attempt to map
the chrominance/luminance hyperspace in an interval scale,
using either successive, just-noticeable differences or mag-
nitude scaling as a psychophysical technique. Emphasis
should be placed on suprathreshold scaled differences in
developing the perceptual space, for the application is to
be made to high levels of accuracy in observer discrimina-
bility, not to simple difference thresholds. This objective
is being met by a separate task under the ONR contract which
supported preparation of this report.

6.2.2.2   Effect on legibility

Once a metric of color contrast is defined, and is proven
to be of an interval nature on all combined dimensions, then
the sensitivity and predictability of the color contrast
metric to operator performance must be determined. That is,
equal performance increments for representative tasks should
be scaled to the color contrast space.   The product of this
activity would be a set of design criteria which would
relate the color contrast metric to performance such as per-
cent legibility errors, search time, reading time, etc.
Careful spectrophotometric quantification of the display is
the key to the success of this task.

6.2.2.3   Wavelength distribution versus dominant wavelength

While the visual system "sees" a colored stimulus in rel-
ation to its dominant wavelength, it is also clear that some
displays are narrow band emitters (e.g., gas discharge),
while others are broad-band emitters (e.g., green CRT phos-
phors). It is necessary to determine if the wavelength dis-
tribution of the display emission (or modulation)   has any
impact on the above established metric of color contrast.
If so, then the metric should be revised to account for the
distribution of the emission spectrum as well as the domi-
nant wavelength.

This research should include parametric variation in the
wavelength distribution as well as the dominant wavelength,

and should determine the effects of these two variables on both color contrast thresholds and representative observer task performance.

6.2.3   Font, Matrix Requirements

It was pointed out earlier that optimal font design for dot-matrix displays is very different than that for continuous stroke alphanumeric characters. Considerable progress has been made in the design of optimum fonts and matrix sizes for dot-matrix displays. However, much more is needed. The research indicated below addresses critical variables and design criteria.

6.2.3.1   Upper and lower case alphabets

Optimal font design for dot-matrix characters, to date, has been limited to upper-case (all capitals) letters. It is also necessary to optimize both lower case letters and the mix of upper and lower case. Because the confusion among characters increases with the size of the character set, there is no evidence to suggest that font/matrix combinations currently considered to be optimal for upper-case letters will remain so for combinations of upper and lower case. This is particularly important because (1) reading speed is faster for combinations of upper- and lower-case letters, and (2) lower-case letters have descenders (parts of the letter "below the line") which can require larger

matrix sizes for maximum legibility. Techniques for these investigations might be modeled after the recent dot-matrix font optimization research of Snyder and Maddox (1978).

6.2.3.2 Symbols

Very little font research has been performed on stroke symbols, such as the usual mathematical symbols, Greek letters, and the remainder of the ASCII keyboard. Research has been conducted to determine requirements for symbol presentation in a dot-matrix addressed display, even though several commercially available displays have significant amounts or symbology programmed into the character set (e.g., PLATO).

6.2.3.3 Contextual effects

It has been demonstrated that alphanumeric characters in context (e.g., words, paragraphs, equations) benefit from the contextual redundancy, and require less stringent design criteria for the same level of legibility. Such information is only available for upper-case letters. This work needs to be replicated (only one study has been performed) and extended to other matrix sizes and fonts, as well as to upper- and lower-case displays.

6.2.3.4 Symbol rotation

Only one set of experiments has investigated the effects of symbol rotation on single character legibility (Vanderkolk et al., 1974). That study, conducted with partial factorial combinations of other variables, used a limited alphabet, a single font, and limited levels of other critical variables. Further, it involved only a single-character recognition task. While the data are interesting and useful, they must be extended to a more representative variety of viewing conditions.

6.2.3.5 Vectorgraphic antialiasing

Only very recently has attention been given to the improvement of line graphic displays in a manner to reduce the "stair-stepping" or irregular effect of the fixed dot pattern on obliquely oriented lines. At the present time, coarse dot-matrix displays produce an annoying amount of this irregularity, and several techniques have been suggested to reduce it. These methods, of course, include element-ordered jitter and half-brightness modulation of adjacent picture elements to "soften" the line edges. While some of these examples presented to date are subjectively impressive, there are no data on the improvement of performance with various types of antialiasing, or, for that matter, data on the threshold detection of differences. Since such antialiasing is expensive, in terms of software devel-

opment and display cycle time, these evaluations are needed
before the techniques are implemented in numerous applica-
tions.

6.2.4   Resolution Requirements

Previous sections of this report have indicated perform-
ance improvements with increases in element density, and
also performance improvements with decreases in the element
space/element size ratio, which is inversely related to the
measure of percent active area. It was shown, for simple
legibility tasks, as percent active area drops below about
45%, performance deteriorates under difficult viewing condi-
tions. Similarly, improvements in performance for simple
tasks can be demonstrated for element densities up to about
2.4/mm.

However, many technologies are capable of producing much
higher densities and much greater percent active areas, but
at a substantial cost. For this reason, it is critical to
determine under what conditions, if any, further improve-
ments in such measures of resolution might cause increases
in observer performance.

6.2.4.1   Tradeoff with percent active area

Because electrode thickness will essentially be constant,
as the element density increases greatly, the percent active
area will decrease. Since this relationship has been found

to be a critical determinant of legibility and reading speed, further studies to explore the higher resolution cases are needed. Studies to date have been limited to about 15 or 20 elements per degree angular subtense. The effects must be studied down to the limits of visual resolution, or about 50 to 60 elements per degree. The tradeoff between element density and percent active area must be determined and related to performance on meaningful and typical tasks.

6.2.4.2   Performance/cost/addressing tradeoffs

As resolution is increased, through either element density or percent active area, increases in either fabrication cost of the display or in addressing circuit costs typically result. Thus, the increased performance that may result from resolution increases must be evaluated against the increased costs associated with the increased resolution. This analysis can be done following completion of the resolution studies described above.

6.2.5   Predictive Model Development

Human factors engineering has moved beyond empirical data collection as its fundamental source of information. Increasing use is being made of simple mathematical models to predict human performance in a variety of situations and for a variety of tasks. These models range from the very

global multi-operator type to more molecular models of manual control to simple perceptual models. While the validity of such models should always be empirically established, the extension of these models to other, related applications can be economical and reasonably safe. It is, of course, highly desirable to have a valid mathematical model to predict operator performance in a given task as a function of pertinent display variables. Section 5.4 described the existing models for both spatially discrete and spatially continuous displays. From that discussion, it was readily realized that the existing models are quite limited in the variables they include, as well as in the ranges of the variables. Table 31 summarizes the by-variable coverage of these models or concepts, as well as the data gaps requiring future research.

It is clear from this summary table that there is only fair coverage of the variables critical to a video display. Although the most promising models (SNR_D, MTFA) were developed for CRT displays, they should be adaptable for video information presented on flat-panel displays. Such adaptation has not been attempted yet. These video-developed models are essentially worthless for graphic or alphanumeric display evaluation.

On the other hand, the two models which have been specifically developed for application to flat-panel, matrix addressed displays are limited to their explicit purposes.

Table 31. Range of Display Variables Covered by Extant Models of Display Quality

Display Variable	SNR_D	N_e	MTFA	SQF	Power Density Spectrum	Maddox Contingency Model	Dot-Matrix Model
Video Displays:							
Line Rate	complete	no	complete	doubtful	possible	no	≤ 50 cyc/deg
Bandwidth	complete	complete	complete	doubtful	possible	no	doubtful
Target	simplified	no	simplified	no	complete	no	no
Noise Amplitude	complete	complete	complete	doubtful	possible	no	no
Noise Passband	no	no	complete	doubtful	doubtful	no	no
MTF	complete	complete	complete	complete	complete	no	doubtful
Display Size	simplified	no	simplified	no	no	no	no
Color	no	no	no	no	no	no	no
Predictability	fair-good	poor	fair-good	poor	fair	none	none
Graphic Displays:							
Stairstepping	possible	no	possible	no	no	no	doubtful
Line Width	possible	no	possible	no	doubtful	no	≤ 50 cyc/deg
Symbology	no	no	no	no	doubtful	yes	no
Predictability	none	none	none	none	none	limited	none
Alphanumerics:							
Element Size	possible	no	possible	possible	possible	no	≤ 50 cyc/deg
Element Shape	possible	no	doubtful	doubtful	possible	no	complete
Element Spacing	possible	no	possible	doubtful	possible	no	≤ 50 cyc/deg
Modulation	possible	no	probable	probable	probable	no	complete
Font	no	no	no	no	no	upper case	no
Matrix Size	no	no	no	no	no	5x7,7x9,9x11	no
Task Type	no	no	no	no	no	no	three types
Predictability	none	none	none	none	none	limited	fair-good

Maddox's (1979) contingency model permits some prediction of legibility of alphanumerics, but is currently limited to upper-case characters.

The Snyder and Maddox (1978) dot-matrix model takes into account many of the alphanumeric display geometric and photometric variables, plus reading and search type tasks. However, it appears of limited validity beyond spatial frequencies of 50 cyc/deg, and cannot be applied to video or graphics displays.

Thus, the existing models are all limited in their range of applicability to the three generic types of display, as well as to the range of variables they can accept within any display type. The following research activities appear warranted.

6.2.5.1   Performance prediction

A more complete prediction model is needed to relate representative observer performance on typical tasks to display design variables that are specific to matrix addressed displays, or to discrete element alphanumeric displays. A meaningful starting point might be the dot-matrix prediction model of Snyder and Maddox (1978), enlarged as necessary to include specific design variables. It is likely that such a model, after its initial formulation, would require empirical validation studies to refine or revise the model. In view of the rapid development of a wide variety of dot-ma-

- 424 -

trix displays, both monochrome and multi-color, this effort will be of greater importance in future years.

6.2.5.2 Readability, legibility, search

Previous research has shown that these three types of tasks require, for best observer performance, different display design criteria. There are undoubtedly other tasks which also require different design optimization. As a result, logical development of any performance prediction models should include, as a major variable, the task requirements of the observer. Heuristic or empirical development may be required, but empirical validation is considered mandatory.

6.2.5.3 Parametric design tradeoff

The prediction models requirements noted above should directly address design variables of the various display technologies. That is, observer performance should be directly related to specific design variables, much as the SNR_D image quality metric is directly related to specific design parameters of the electrooptical imaging system and the imaged scene. In this manner, the resulting prediction models can be used by design engineers in display research in addition to the obvious usage in display selection.

REFERENCES

Allan, R. LEDs: Digital domination. IEEE Spectrum, 1975,
12.

Almagor, M., Farley, W. W., and Snyder, H. L. Spatio-
temporal integration in the visual system. Wright-
Patterson AFB, OH: Aerospace Medical Research Laboratory
Technical Report, AMRL-TR-78-126, February 1979.

Anderson, B. C. and Fowler, V. J. AC plasma panel TV
display with 64 discrete intensity levels. SID Symposium
Digest of Technical Papers, 1974, V, 28-29.

Baker, C. A. and Grether, W. F. Visual presentation of
information. Wright Air Development Center, OH:
Technical Report WADC-TR-54160, 1954. (AD 43 064)

Beamon, W. S. and Snyder, H. L. An experimental evaluation
of the spot wobble method of suppressing raster structure
visibility. Wright-Patterson AFB, OH: Technical Report
AMRL-75-63, 1975.

Bhargava, R. N. Recent advances in visible LEDs.
Proceedings of the S.I.D., 1975, 16, 103-113.

Biberman, L. M. Perception of displayed information. New
York: Plenum, 1973.

Blackwell, H. R. Contrast thresholds of the human eye.
Journal of the Optical Society of America, 1946, 36,
624-643.

Blondel, A. and Rey, J. Sur la perception des lumieres
breves a la limite de leur portee. Journale de
Physiologie, 1911, 1, 530-550.

Borough, H. C., Failis, R. F., Warnock, R. H., and Britt, J.
H. Quantitative determination of image quality. Boeing
Company Report D2-114058-1, 1967.

Brock, G. C. Image evaluation for reconnaissance. Applied
Optics, 1964, 3, 11.

Brody, T. P. Comparative evaluation of solid state
displays. Personal communication, 1979.

Brown, C. R. and Forsyth, D. M. Fusion contour for intermittent photic stimuli of alternating duration. Science, 1959, 129, 390-391.

Brown, J. L. Flicker and intermittent stimulation. In C. H. Graham (ed.), Vision and visual perception. New York: Wiley, 1965.

Bryden, J. E. Some notes on measuring performance of phosphors used in CRT displays. S.I.D. Symposium, 1966, Boston, 83-103.

Bryngdahl, O. Characteristics of the visual system: Psychophysical measurements of the response to spatial sine-wave stimuli in the photopic region. Journal of the Optical Society of America, 1966, 56, 811-821.

Burnette, J. T. Optimal element size-shape-spacing combination for a 5 x 7 dot matrix visual display under high and low ambient illuminance. Unpublished M.S. thesis, VPI&SU, Blacksburg, VA, 1977.

Burnette, K. T. and Melnick, W. Multi-mode matrix (MMM) flat-panel LED varactor-graphic concept demonstrator display. Proceedings of the S.I.D., 1980, 21, 113-126.

Campbell, F. W. and Kulikowski, J. J. Orientational selectivity of the human visual system. Journal of Physiology, 1966, 187, 437-444.

Campbell, F. W., Kulikowski, J. J., and Levinson, J. The effect of orientation on the visual modulation of gratings. Journal of Physiology, 1966, 187, 427-436.

Campbell, F. W. and Robson, J. G. Application of Fourier analysis to the visibility of gratings. Journal of Physiology, 1968, 197, 552-568.

Carel, W. L., Herman, J. A., and Hershberger, M. L. Research studies for the development of design criteria for sensor display systems. Hughes Aircraft Company Technical Report, March 1976.

Chang, I. F. Electrochromic and electrochemichromic materials and phenomena. In A. R. Kmetz and F.K. von Willison (ed.), Nonemissive electrooptic displays. New York: Plenum, 1976.

Chapanis, A. How we see: A summary of basic principles. In Committee on Undersea Warfare, National Research Council, A Summary Report on Human Factors in Undersea Warfare. Washington: National Research Council, 1949.

Charman, W. N. and Olin, A. Tutorial: Image quality criteria for serial camera systems. Photographic Science and Engineering, 1965, 9, 385-397.

Chodil, G. Gas discharge displays for flat-panel. Proceedings of the S.I.D., 1976, 17, 14-22.

Cook, T. C. Color coding--A review of the literature. Aberdeen Proving Ground, MD: Technical Note HEL 9074, November 1974.

Cornsweet, T. N. Visual perception. New York: Academic Press, 1970.

Crisp, M. D., Hinson, D. C., and Siegel, J. I. Luminous efficiency of a Digivue display/memory panel. Proceedings of 1974 Conference on Display Devices and Systems, 9-10 October 1974.

Crow, F. C. The aliasing problem in computer-generated shaded images. CACM, 1977, 20, 799-805.

Dalisa, A. L. Electrophoretic display technology. Proceedings of the S.I.D., 1977, 18, 43-50.

de Lange, H. Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light: I. Attenuation characteristics with white and colored light. Journal of the Optical Society of America, 1958, 48, 777-784.

de Lange, H. Eye's response at flicker fusion to square-wave modulation of a test field surrounded by a large study field of equal mean luminance. Journal of the Optical Society of America, 1961, 51, 415-421.

De Palma, J. J. and Lowry, E. M. Sine-wave response of the visual system. II. Sine-wave and square-wave contrast sensitivity. Journal of the Optical Society of America, 1962, 52, 328-335.

Erickson, R. A., Linton, P. M., and Hemingway, J. C. Human factors experiments with television. Naval Weapons Center, China Lake, CA. Technical Report NWC TP 4573, October 1968.

Ernstoff, M. N. Liquid crystal pictorial display. Paper presented at S.I.D. Technical meeting, Culver City, CA, 6 November 1976.

Ernstoff, M. N. A head-up display for the future. Proceedings of the S.I.D., 1978, 19, 169-179.

Farrell, R. J. and Booth, J. M. Design handbook for imagery interpretation equipment. Boeing Technical Report D180-19063-1, December, 1975.

Forsyth, D. M. Use of a Fourier model in describing the fusion of complex visual stimuli. Journal of the Optical Society of America, 1960, 50, 3347-341.

Frescura, B. L. and Luechinger, H. Large GaAsP monolithic XY addressable LED arrays for continuous matrix displays. SID Digest, 1975, 196-198.

Frost, J. Generation of subjective colors in an electroluminescent display. Autonetics Technical Report TM-64-342-04-101, 1964.

Fugate, K. O. High display viewability provided by thin-film EL, black layer, and TFT drive. Proceedings of the S.I.D., 1977, 18, 125-133.

Fukushima, M., Murayame, S., Kaji, T., and Mikoshiba, S. A flat gas-discharge panel TV with good color saturation. Proceedings of the S.I.D., 1975, 16, 69-74.

Gaskill, J. D. Linear systems, Fourier transforms, and optics. New York: Wiley, 1978.

Goede, W. F. A digitally-addressed flat panel CRT review. IEEE Intercon, 1973, 5, 33/2.

Goede, W. F. Flat-panel displays: Six major problems and some solutions. Stanford/S.I.D. Seminar, San Francisco, 17 April 1978.

Goodman, L. A. Passive liquid displays: Liquid crystals, electrophoretics, and electrochromics. IEEE Transactions on Consumer Electronics, 1975, CE-21, 247-259.

Goodman, L. A. The relative merits of LEDs and LCDs. Proceedings of the S.I.D., 1975, 16, 8-19.

Graham, C. H. Vision and visual perception. New York: Wiley, 1965.

Granger, E. M. and Cupery, K. N. An optical merit function (SQF), which correlates with subjective image judgments. Photographic Science and Engineering, 1972, 16,

Granger, E. M. and Heurtley, J. C. Visual chromaticity-modulation transfer function. Journal of the Optical Society of America, 1973, 63, 1173-1174.

Guth, S. L. Nonadditivity and inhibition among chromatic luminances at threshold. Vision Research, 1967, 7, 319-328.

Gutmann, J. C., Snyder, H. L., Farley, W. W., and Evans, J. E., III. An experimental determination of the effect of image quality on eye movements and search for static and dynamic targets. Wright-Patterson AFB, OH: USAF Technical Report AMRL-TR-79-51, August 1979.

Hairfield, H. W. Night and all-weather target acquisition: State-of-the-art review. Part III: Television and low-light-level television systems. Boeing Company Report D162-10116-3, 1970.

Herold, E. W. History and development of the color picture tube. Proceedings of the S.I.D., 1974, 14, 141-149.

Hitchcock, L. Preliminary listing of cockpit display and control requirements. Naval Air Development Center, Warminster, PA: Technical Memorandum 73-009-2, 5 March 1973.

Humes, J. M. and Bauerschmidt, D. K. Low light level TV viewfinder simulation program. Phase B: The effects of television system characteristics upon operator target recognition performance. Wright-Patterson AFB, OH: USAF Avionics Laboratory Technical Report AFAL-TR-68-271, November 1968.

Inoguchi, T., Takeda, M., Kakihara, Y., Nakata, Y., and Yoshida, M. Stable high-brightness thin-film electroluminescent panels. S.I.D. Digest, 1974, 84-85.

Ivey, H. F. Electroluminescence and related effects. In L. Martin (ed.), Advances in electronics and electronic physics. New York: Academic Press, 1963.

Judice, C. N. Introduction to special issue on processing of images for bilevel displays and the generation of pseudo-gray scale. Proceedings of the S.I.D., 1976, 17, 62.

Kaelin, G. R., Nakahara, R. H., Piatt, D. M., Price, J. G., and Riggs, A. L. LED X-Y matrix display. Wright-Patterson AFB, OH: AFFDL-TR-75-49, May 1975.

Kaneko, E. Personal communication, visit to Hitachi Research Laboratory, 14 November 1978.

Kaneko, E., Kawakami, H., and Hanmura, H. Liquid crystal television display. Proceedings of the S.I.D., 1978, 19, 49-54.

Kawarada, H. and Ohshima, N. EC EL materials and techniques for flat-panel TV display. Proceedings of the IEEE, 1973, 61, 907-915.

Kazan, B. Electroluminescent displays. Proceedings of the S.I.D., 1976, 17, 23-29.

Keesee, R. L. Prediction of modulation detectability thresholds for line-scan displays. Wright-Patterson AFB, OH: USAF Technical Report AMRL-TR-76-36, 1976.

Kelly, D. H. Effects of sharp edges in a flickering field. Journal of the Optical Society of America, 1959, 49, 730-732.

Kelly, D. H. Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements. Journal of the Optical Society of America, 1961, 51, 422-429.

Kelly, D. H. Frequency doubling in visual responses. Journal of the Optical Society of America, 1966, 56, 1628-1633.

Kerr, J. L. Visual resolution in the periphery. Perception and Psychophysics, 1971, 9, 375-387.

Kling, J. W. and Riggs, L. A. Experimental psychology. New York: Holt, Rinehart, and Winston, 1971.

Klingberg, C. S., Elworth, C.S., and Filleau, C. R. Image quality and detection performance of military photointerpreters. Boeing Company Report D162-10323-1, 1970.

Krebs, M. J., Wolf, J. D., and Sandvig, J. H. Color display design guide. Minneapolis, MN: Honeywell Report ONR-CR213-136-2F, October 1978.

Kriss, M., Michelson, M., and Nail, N. Image structure and visual sharpness. Paper presented at 1971 Optical Society of America Meeting, Ottawa, Canada.

Krupka, D. C. and Fukui, H. The determination of relative critical flicker frequencies of raster-scanned CRT displays by analysis of phosphor persistence characteristics. Proceedings of the S.I.D., 1973, 14, 87-93.

Levinson, J. Flicker fusion phenomena. Science, 1960, 160, 21-28.

Lewis, J. C. Electrophoretic displays. In Kmetz, A. R. and von Willisen, F. K. (ed.), Nonemissive electrooptic displays. New York: Plenum, 1976.

Ludvigh, E. and Miller, J. W. Study of visual acuity during the ocular pursuit of moving test objects. I. Introduction. Journal of the Optical Society of America, 1958, 48, 799-802.

Luxenberg, H. R. and Kuehn, R. M. Display system engineering. New York: McGraw-Hill, 1968.

Maddox, M. E. Two-dimensional spatial frequency content and confusions among dot-matrix characters. Proceedings of the S.I.D., 1980, 21, 31-40.

Maddox, M. E., Burnette, J. T., and Gutmann, J. C. Font comparisons for 5 x 7 dot-matrix characters. Human Factors, 1977, 19, 89-93.

Marsden, A. M. An elemental theory of induction. Vision Research, 1969, 9, 653-664.

MacAdam, D. L. Color discrimination and the influence of color contrast on visual acuity. Optique Physiologique, Coleurs, 1949, 28, 161-173.

McLean, M. V. Brightness contrast, color contrast, and legibility. Human Factors, 1965, 7, 521-526.

Meister, D. and Sullivan, D. J. Guide to human engineering design for visual displays. Washington: ONR Report under Contract N00014-68-C-0278, 1969. (AD 693 237)

Mills, G. S., Grayson, M. A., Loy, S. L., and Jauer, R. A., Jr. Research on visual display integration for advanced fighter aircraft. Wright-Patterson AFB, OH: USAF Technical Report AMRL-TR-78-97, 1978.

Mito, S., Suzuki, C., Kanatami, Y., and Ise, M. TV imaging system using electroluminescent panels. S.I.D. Digest, 1974, 86-67.

Montalvo, F. S. Human vision and computer graphics. Proceedings of SigGraph, 1979, 121-125.

Myers, W. S. Accommodation effects in multicolor displays. Wright-Patterson AFB, OH: Technical Report AFFDL-TR-67-161, December 1967 (AD 826 134).

NAVSHIPS, U.S. Navy. Operational stations book for amphibious command ships (LCC 19 Class). Report NAVSHIPS 0909-003-1010 (Revised), Department of the Navy, Naval Sea Systems Command, Washington, D. C., 1976.

Ohishi, I., Kojima, T., Ikeda, H., Toyonaga, R., Murakami, H., Kioke, J., and Tajima, T. An experimental real-time color-TV display with a DC gas-discharge panel. Proceedings of the S.I.D., 1975, 16, 62-68.

Ota, I., Ohnishi, J., and Yoshiyama, M. Electrophoretic display panel--EPID. Proceedings of the IEEE, 1973, 61, 832.

Ota, I., Sato, T., Tanaka, S., Yamagami, T., and Takeda, H. Paper presented at Laser 75 Seminar, Munich, June 1975.

Oyama, T. and Hsia, Y. Compensatory hue shift in simultaneous color contrast as a function of separation between inducing and test fields. Journal of Experimental Psychology, 1966, 71, 405-413.

Patel, A. S. Spatial resolution by the human visual system. The effect of mean retinal illuminance. Journal of the Optical Society of America, 1966, 56, 689-694.

Piper, W. W. and Williams, F. E. Electroluminescence. In F. Seitz and D. Turnbull (ed.), Solid state physics, Vol. 5. New York: Academic Press, 1957.

Pollack, J. D. Reaction time to different wavelengths at various luminances. Perception and Psychophysics, 1968, 3, 17-24.

Radio Corporation of America. Electro-optics handbook. Camden, NJ: RCA, 1974.

Riley, T. Multiple imaging in LED displays. Human Factors, 1977, 19, 79-81.

Riley, T. M. and Barbato, G. J. Dot-matrix alphanumerics viewed under discrete element degradation. Human Factors, 1978, 20, 473-479.

Robson, J. G. Spatial and temporal contrast-sensitivity functions of the visual system. Journal of the Optical Society of America, 1966, 56, 1141-1142.

Rosell, F. A. Analysis of electro-optical imaging sensors. Technical Report ADTM No. 105, Westinghouse Electric Corp., 1971.

Rosell, F. A. and Willson, R. H. Recent psychophysical experiments and the display signal-to-noise ratio concept. In L. M. Biberman (Ed.), Perception of displayed information. New York: Plenum, 1973.

Sasaki, A. and Takagi, T. Display-device research and development in Japan. IEEE Transactions on Electron Devices, 1973, ED-20, 925-933.

Scanlan, L. A. and Carel, W. L. Human performance evaluation of matrix displays: Literature and technology review. Wright-Patterson AFB, OH: AMRL-TR-76-39, June 1976.

Schade, O. H. Image gradation, graininess, and sharpness in television and motion-picture systems. Part III: The grain structure of television images. Journal of the Society of Motion Picture and Television Engineers, 1953, 61, 97-164.

Scheffer, T. J. Liquid crystal color displays. In A. R. Kmetz and F. K. von Willison (Ed.), Nonemissive electrooptic displays. New York: Plenum, 1976.

Schlam, E. Electroluminescent phosphors. Proceedings of the IEEE, 1973, 61, 894-901.

Schober, H. A. W. and Hilz, R. Contrast sensitivity of the human eye for square wave gratings. Journal of the Optical Society of America, 1965, 55, 1086-1091.

Semple, C. A., Heapy, R. J., Conway, E. J., and Burnette, K. T. Analysis of human factors data for electronic flight display systems. Wright-Patterson AFB, OH: Technical Report AFFDL-TR-70-174, 1971.

Sherr, S. Electronic displays. New York: Wiley, 1979.

Silman, L. O. The effect of spatial-temporal image properties on observer identification of displayed alpha code symbols. Unpublished Master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1978.

Slocum, G. K. Airborne sensor display requirements and approaches. Culver City, CA: Hughes Aircraft Company Technical Report TM-888, September, 1967.

Snyder, H. L. Image quality and observer performance. In L. M. Biberman (ed.), Perception of displayed information. New York: Plenum, 1973.

Snyder, H. L. Visual search and image quality: Final report. Wright-Patterson AFB, OH: USAF Report AMRL-TR-76-89, December 1976.

Snyder, H. L. et al. Low-light-level TV viewfinder
    simulation program. Phase A: State-of-the-art review.
    Wright-Patterson AFB, OH: Technical Report AFAL-
    TR-67-293, 1967.

Snyder, H. L., Almagor, M., and Shedivy, D. I. Raster-scan
    display photometric noise measurement. Wright-Patterson
    AFB, OH: USAF Report AMRL-TR-79-13, 1979.

Snyder, H. L., Beamon, W. S., Gutmann, J. C., and Dunsker,
    E. D. An evaluation of the effect of spot wobble upon
    observer performance with raster scan displays. Wright-
    Patterson AFB, OH: Technical Report AMRL-TR-79-91, 1980.

Snyder, H. L. and Greening, C. P. The effects of direction
    and velocity of relative motion upon dynamic visual
    acuity. Las Vegas, NV: Aerospace Medical Association
    Meeting, April, 1966.

Snyder, H. L., Keesee, R. L., Beamon, W. S., and Aschenbach,
    J. R. Visual search and image quality. Wright-Patterson
    AFB, OH: Aerospace Medical Research Laboratory Technical
    Report AMRL-TR-73-114, 1974.

Snyder, H. L. and Maddox, M. E. Information transfer from
    computer-generated, dot-matrix displays. Blacksburg, VA:
    VPI&SU Technical Report HFL-78-3/ARO-78-1, October 1978.

Stein, I. H. The effect of active area on the legibility of
    dot-matrix displays. Proceedings of the S.I.D., 1980,
    21, 17-20.

Stiles, W. S. Investigations of the scotopic and
    trichromatic mechanisms of vision by the two-colour
    threshold techhnique. Review of Ophthalmology, 1949, 28,
    215-237.

Suen, C. Y. and Shiau, C. An iterative technique of
    selecting an optimal 5 x 7 matrix character set for
    display in computer output systems. Proceedings of the
    S.I.D., 1980, 21, 9-18.

Tannas, L.E., Jr. Flat panel displays in perspective.
    Proceedings of the S.I.D., 1978, 19, 193-198.

Tannas, L.E., Jr. Flat-panel displays: Six major problems
    and some solutions. Stanford/S.I.D. Seminar, San
    Francisco, 17 April 1978.

Task, H. L. An evaluation and comparison of several
    measures of image quality for television displays.
    Wright-Patterson AFB, OH: USAF Technical Report AMRL-
    TR-79-7, January 1979.

Turnage, R. E. The perception of flicker in cathode ray tube displays. Information Display, 1966, 3, 38.

Tyte, R., Wharf, J., and Ellis, B. Visual response times in high ambient illumination. SID Digest, 1975, 98-99.

Umeda, S., Murase, K., Ishizaki, H., and Juraheshi, K. Picture display with gray scale in the plasma panel. S.I.D. Symposium Digest, 1973, 70-71.

van der Horst, G. J. C. Fourier analysis and color discrimination. Journal of the Optical Society of America, 1969, 59, 1670-1676.

van der Horst, G. J. C. and Bouman, M. A. Spatio-temporal, chromaticity discrimination. Journal of the Optical Society of America, 1969, 59, 1482-1488.

van der Horst, G. J. C., de Weert, C., and Bouman, M. A. Transfer of spatial chromaticity contrast at threshold in the human eye. Journal of the Optical Society of America, 1967, 57, 1260-1266.

Vanderkolk, R. J., Herman, J. A., and Hershberger, M. L. Dot matrix display symbology study. Wright-Patterson AFB, OH: USAF Report AFFDL-TR-75-75, 1975.

Van Meeteren, A., Vos, J. J., and Bongaard, J. Contrast sensitivity in instrumental vision. Institute for Perception Report Number IZF 1968-9, Soesterberg, The Netherlands. 1968.

Van Nes, F. L. and Bouman, M. A. Spatial modulation transfer in the human eye. Journal of the Optical Society of America, 1967, 57, 401-406.

Van Raalte, J. A. Matrix TV displays: Systems and circuit problems. SID Proceedings, 1976, 17, 8-13.

Vecht, A., Werring, N. J., Ellis, R., and Smith, P. J. F. Direct-current electroluminescence in zinc sulfide: State of the art. Proceedings of the IEEE, 1973, 61, 902-907.

Walsh, J. W. T. Photometry. New York: Dover, 1965.

Watanabe, A., Mori, T., Nagata, S., and Hiwatashi, K. Spatial sine-wave responses of the human visual system. Vision Research, 1968, 8, 1245-1263.

Weibel, G. E. Short communication: Mechanism of electrochromism in WO_3. In A. R. Kmetz and F. K. von Willison (ed.), Nonemissive electrooptic displays. New York: Plenum, 1976.

Weston, G. F. Plasma panel displays. *Journal of Physics E: Scientific Instruments*, December 1975, *8*, 981-991.

Williams, L. G. and Erickson, J. M. Contrast sensitivity as a function of spatial and temporal frequency, luminance, and stimulus position on the retina. Report F0259-IR1, Honeywell, Inc., Minneapolis, 1974.

Wysecki, G. and Stiles, W. S. *Color science: Concepts and methods, quantitative data and formulas*. New York: Wiley, 1967.

Wysocki, J. J., Becker, J. H., Dir, G. A., Madrid, R., Adams, J. E., Hass, W. E., Leder, L. B., Mechlowitz, B., and Saeva, F. D. Cholesteric-nematic phase transition displays. *Proceedings of the S.I.D.*, 1972, *13*, 114-120.

Yund, E. W. and Armington, J. C. Color and brightness contrast effects as a function of spatial variables. *Vision Research*, 1975, *15*, 917-929.

Zeller, H. R. Principles of electrochromism as related to display applications. In A. R. Kmetz and F. K. von Willison (ed.), *Nonemissive electrooptic displays*. New York: Plenum, 1976.

TECHNICAL REPORTS DISTRIBUTION LIST

OFFICE OF NAVAL RESEARCH

CODE 455

CDR Paul R. Chatelier
Office of the Deputy Under Secretary
    of Defense
OUSDRE (E&LS)
The Pentagon, Room 3D129
Washington, D. C. 20301

Director, Undersea Technology
Code 220, Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Communication & Computer Technology
Code 240, Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director, Physiology Program
Code 441, Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Eastern/Central Regional Office
ATTN: Dr. J. Lester
Bldg. 114, Sec. D, 666 Summer Street
Boston, MA 02210

Commanding Officer
ONR Branch Office
ATTN: Dr. C. Davis
536 South Clark Street
Chicago, IL 60605

Commanding Officer
ONR Western Regional Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Dr. Robert G. Smith
Office of the Chief of Naval
    Operations, OP987H
Personnel Logistics Plans
Washington, D. C. 20350

Director, Engr. Psychology Programs
Code 455, Office of Naval Research
800 North Quincy Street
Arlington, VA 22217 (5 cys)

Director, Aviation & Aero. Tech.
Code 210, Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Electronics & Electromagnetics
    Technology, Code 250
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Information Systems Program
Code 437, Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Special Assistant for Marine Corps
    Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Western Regional Office
ATTN: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D. C. 20375 (6 cys)

Dr. W. Mehuron
Office of the Chief of Naval
    Operations,
OP 987
Washington, D. C. 20350

Dr. Andreas B. Rechnitzer
Office of the Chief of Naval
  Operations, OP 952F
Naval Oceanography Division
Washington, D. C. 20350

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, FL 32813

Mr. John Quirk
Naval Coastal Systems Laboratory
Code 712
Panama City, FL 32401

Mr. Merlin Malehorn
Office of the Chief of Naval
  Operations, OP 102
Washington, D. C. 20350

Director, Organizations and Systems
  Research Laboratory
US Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

US Army Aeromedical Research Lab
ATTN: CPT Gerald P. Krueger
Ft. Rucker, AL 36362

Dr. Donald A. Topmiller, Chief
Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

CDR Robert Biersner
Naval Medical R&D Command
Code 44, Naval Medical Center
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Dr. Jerry C. Lamb
Submarine Sonar Department
Code 3293
Naval Underwater Systems Center
New London, CT 06320

Dr. Sam Schiflett
Human Factors Section
Systems Engineering Test
  Directorate
US Naval Air Test Center
Patuxent River, MD 20670

LCDR W. Moroney
Code 55MP
Naval Postgraduate School
Monterey, CA 93940

Dr. Joseph Zeidner, Tech. Dir.
US Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
US Army Human Engineering Labs
Aberdeen Proving Ground, MD 2.J05

US Air Force Office of Scientific
  Research
Life Sciences Directorate, NL
Bolling AFB, Washington, D.C.20332

CDR R. Gibson
Bureau of Medicine & Surgery
Aerospace Psychology Branch
Code 513
Washington, D. C. 20372

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

Head
Aerospace Psychology Department
Code L5, Naval Aerospace Medical
  Research Lab
Pensacola, FL 32508

Dr. A. L. Slafkosky, Scientific Adv.
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

Chief, C^3 Division
Development Center
MCDEC
Quantico, VA 22134

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 340F
Washington, D. C. 20361

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 4701
Washington, D. C. 20360

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology
  Establishment
Teddington, Middlesex TW11 0LN
ENGLAND

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, VA 22202

Dr. Richard R. Rosinski
Department of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260

Dr. Arthur I. Siegel
Applied Psychological Services, Inc.
404 East Lancaster Street
Wayne, PA 19087

Dr. W. S. Vaughan
Oceanautics, Inc.
422 6th Street
Annapolis, MD 21403

Commanding Officer
MCTSSA
Marine Corps Base
Camp Pendleton, CA 92055

Mr. Arnold Rubinstein
Naval Material Command
NAVMAT 08D22
Washington, D. C. 20360

Commander
Naval Air Systems Comman
Crew Station Design
NAVAIR 5313
Washington, D. C. 20361

Mr. Phillip Andrews
Naval Sea Systems Command
NAVSEA 0341
Washington, D. C. 20362

Director, Human Factors Wing
Defence & Civil Institute of
  Environmental Medicine
P. O. Box 2000
Downsview, Ontario M3M 3B9
CANADA

Defense Technical Information Ctr.
Cameron Station, Bldg. 5
Alexandria, VA 22314
(12 cys)

Dr. Robert G. Pachella
Human Performance Center
University of Michigan
330 Packard Road
Ann Arbor, MI 48104

Dr. T. B. Sheridan
Dept. of Mechanical Engineering
Massachusetts Inst. of Technology
Cambridge, MA 02139

Dr. Amos Freedy
Perceptronics Inc.
6271 Variel Avenue
Woodland HIlls, CA 91364

Dr. Robert Fox
Department of Psychology
Vanderbilt University
Nashville, TN 37240

- 4 -

Dr. Alphonse Chapanis
Department of Psychology
The Johns Hopkins University
Baltimore, MD 21218

Dr. James H. Howard, Jr.
Department of Psychology
The Catholic University of America
Washington, D. C. 20064
    Menlo Park, CA 94025

Dr. Christopher Wickens
University of Illinois
Department of Psychology
Urbana, IL 61801

Dr. Edward R. Jones, Chief
Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division, Box 516
St. Louis, MO 63166

Dr. Richard W. Pew
Exper. Psychology Dept.
Bolt Beranek & Newman Inc.
50 Moulton Street
Cambridge, MA 02138

Journal Supplement Abstract Serv.
American Psychological Association
1200 17th Street, N.W.
Washington, D. C. 20036

Dr. Thomas P. Piantanida
SRI International
BioEngineering Research Center
333 Ravensworth Avenue

Dr. Douglas Towne
University of Southern California
Behavioral Technology Laboratory
3716 South Hope Street
Los Angeles, CA 90007

Mr. Harold E. Price
BioTechnology, Inc.
3027 Rosemary Lane
Falls Church, VA 22042

Mr. A. Finelli, Librarian
Institut de Programmation
Tour 55-65
4, place Jussieu
75230 Paris, France