

COLOR IN ELECTRONIC DISPLAYS

Everything from the physics of light generation
to the human factors of color perception
must be considered in their design.

Lawrence E. Tannas Jr

When historians judge the 20th century, they may conclude that one of its most important technical accomplishments was the electronic color video display. Certainly there have been many other notable inventions, such as the transistor, the computer and the video camera, but these are all parts used to make electronic images. Electronic images are still many times more expensive than other forms of images, and they are limited in resolution. Nonetheless, they have an advantage over all other forms of images that completely eclipses the limitations: the speed and ease with which they can be captured, distributed, stored, retrieved, edited and displayed in real time. One can sit in Los Angeles and see the finish of an Olympic event in Barcelona with nearly the same sense of reality that someone actually in the grandstands experiences. Some television viewers feel they even have better coverage of the event because of the use of stop action, replay, zoom and multiple camera angles. When such an event is displayed with the resolution of high-definition television and with stereo sound, one can easily imagine being there. In this and many other ways, electronic images are helping to propel the human race into the information age.

I will begin this article by discussing how an electronic display uses many small, discrete, spatially independent building blocks called picture elements, or pixels, to create an image. Next I will describe how this is done in two important examples of electronic displays, shadow-mask cathode-ray tubes and matrix-addressed flat-panel displays. The former have been in use in televisions for 40 years; the latter are among the most recent and exciting developments in color display technology. (See figure 1.) Several different techniques are used to generate the frequencies of light needed for such displays. Finally I will consider the impact human factors of color perception have on the quality of electronic displays and the difficult problem of achieving device independence—that is, obtaining the same image on both a print and a display screen, despite the vagaries of ambient lighting.

Electronic pointillism

An electronic display is made up of many pixels, a system directly analogous to the technique employed by the 19th-

century French Impressionist painter Georges Seurat when he created "Sunday Afternoon on the Island of La Grande Jatte" by painting numerous colored dots on the canvas. In a display, the pixels are individually controllable and are arranged in an orthogonal array or matrix. On a US television set built to National Television System Committee (NTSC) standards, the image is made from approximately 480 horizontal scan lines with approximately 380 pixels per line. In the video graphic adapter (VGA) computer monitor format, the image is typically made with 480 rows of 640 pixels each.

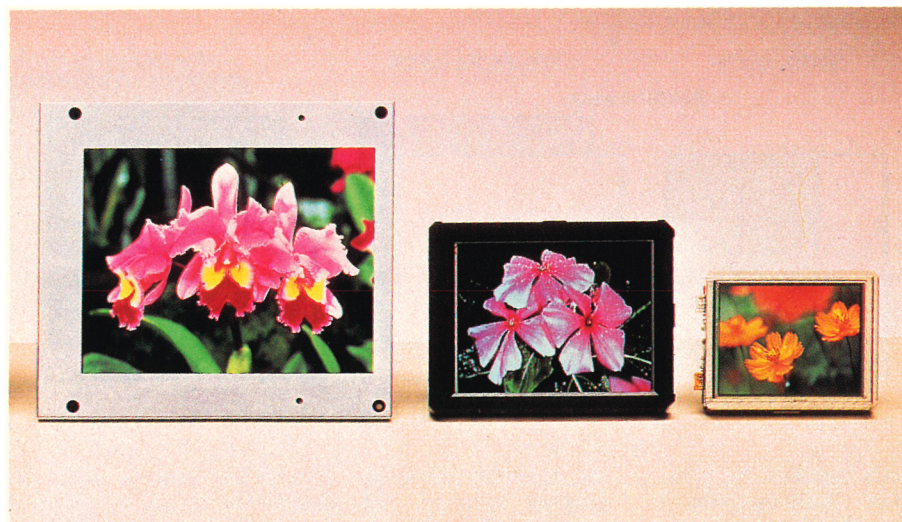
Each pixel costs money to build, requires memory for temporary storage and requires bandwidth in the refresh signal. Therefore electronic displays are designed to minimize the pixel count, and electronic images are always pixel limited. In addition, we are limited by display technology. Electronics, phosphors and other required materials do not yet have the properties needed to make higher-performance displays. We physically cannot make a display with as many pixels as exist in a photographic print or a picture made by a printing process. Displays are getting better every year, and electronics are becoming less expensive, but the same is true for the photographic and printing arts.

Because of their limited number of pixels, electronic displays have an optimum viewing distance, which depends on the lighting and on the visual acuity and accommodation of the observer. Typically, the optimum distance is that at which the pixels are separated by 1 minute of arc. According to that criterion, the optimum viewing distance of an NTSC television is seven screen heights. As one moves closer to the television set, the image does not become any clearer and is not seen in more detail, because there is no further detail to be seen. With a photographic print one can move closer and even use a magnifying glass to see more detail. This is not so with electronic displays.

A pixel of an electronically displayed image contains no spatial information. All the spatial information is represented in the difference from pixel to pixel. The image is portrayed by the pattern of pixels, which are turned on to various gray shades, or intensities. It is up to the viewer to use cognitive processes and cultural experience to perceive and interpret the pattern.

Early experiments on human factors in the viewing process found that three and only three orthogonal metrics determine how we detect a pattern of pixels at

Lawrence Tannas Jr is a consultant and president of Tannas Electronics, in Orange, California.



First flat-panel displays to match the cathode-ray tube's ability to display full color with gray shades, good contrast ratios and high brightness at video speeds are active-matrix liquid crystal displays using amorphous silicon thin film transistors. Despite remaining five to ten times more expensive than CRTs and having a more limited viewing angle and overall size, AMLCDs can be used in many new products that would be impractical with a bulky CRT. The three displays shown here have diagonal sizes of 9, 6 and 4 inches. (Courtesy of Sharp Electronics.) **Figure 1**

various relative gray shades. The three metrics are resolution, luminance and contrast. That is, the pixels must be large enough and bright enough to be seen and different enough to be discerned.

Electronic displays have three basic ways to impart color to a pixel: spatially, temporally and coincidently. (See figure 2.) The spatial technique subdivides the pixel into independently controllable areas for the primary colors red, green and blue. To display a red pixel, one turns on only the red subpixel portion to the intensity needed to give the entire pixel the desired gray shade of red. One produces colors other than the primary colors by turning on more than one subpixel. Optimum viewing for such a display is at the distance where the pixel-to-pixel difference can be seen but the subpixel structure cannot. In a properly designed display at the proper viewing distance, the images of the subpixels are below the resolving power of the eye, and it is immaterial that different frequencies of light are coming from different parts of the pixel. The subpixels fuse at the retina, and the human visual system perceives a single color according to the rules of the additive color system. (See the article by Alan R. Robertson on page 24 for more detail on how the eye-brain system ascribes a hue and saturation and brightness levels to a given stimulus.)

Seurat created a variety of colors on the canvas in a manner similar to spatial electronic color. However, in electronic color we can employ the two additional means of temporal color and coincident color. In the temporal technique, the primary colors are flashed in the pixel area faster than the critical fusion frequency (the rate of flashing below which an image will appear discontinuous), and again the eye-brain system of the viewer sees a single hue at particular saturation and brightness levels. The coincident technique uses layers of transparent phosphors that shine through one another. The primary-color photons emerge simultaneously from the entire pixel area.

Electronic control

One gives a display pixel its color attributes by controlling the red, green and blue subpixels, which add according to Grassman's laws of color addition.^{1,2} With monochrome displays independent pixel control is necessary, but with color displays independent *subpixel* control is necessary. As a consequence, color electronic displays can be up to three times as complex as monochrome displays. If 64 shades of gray are to be shown—the minimum required for

live video—then each subpixel must be controllable to 64 discernible levels of luminous intensity. The full pixel color gamut would then be 64^3 , or about 260 000, colors.

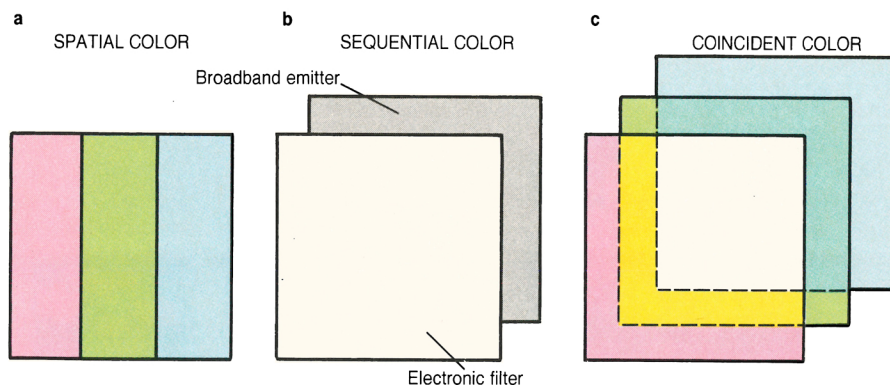
The matrix of pixels can be controlled by any of several basic addressing schemes.³ The two most common are scan addressing, used for cathode-ray tubes, and matrix addressing, used for most flat-panel displays, such as liquid-crystal displays, electroluminescent displays, plasma display panels and light-emitting diode displays.⁴

Figure 3 shows scan addressing of a color CRT. Three independently variable electron beams electronically excite three phosphors. The three beams are synchronously deflected in a raster—a pattern of closely spaced horizontal lines—that scans the image area. Typically the electron beams are deflected magnetically. The beams pass through a metal shadow mask, which has a pattern of holes that are carefully aligned with the phosphors and the beams. The three beams arrive at a given shadow-mask hole on slightly different angles so that the part of each beam that passes through the hole will strike its own corresponding phosphor area. Different phosphor formulations are used for each color. Silver, terbium and europium are typically used as activators in blue, green and red phosphors, respectively.

It is very important that the electrons in the beam intended to excite the red phosphor do not also excite, for example, the blue phosphor. When this happens we say there has been loss of color purity. To ensure proper alignment and avoid loss of color purity, the phosphor dot locations are defined by a photolithographic process using the actual metal shadow mask as the image and a point light source near the origin of the electron beam. However, loss of color purity will still occur if the shadow mask moves due to thermal expansion or vibration.

The video amplifier controls the modulation and cutoff of the electron beams. The beams' diameters and the amplifier control the "spot," or pixel size. It is not possible to aim the beam with sufficient precision to hit a given shadow mask hole: If the television were rotated, the change in orientation with respect to the Earth's magnetic field would throw the beam out of alignment. Thus as the set of beams is deflected to sweep out a raster, the electron-beam spot is made large enough to encompass $2\frac{1}{2}$ shadow mask holes. (This doesn't cause a loss of color purity; the shadow mask still ensures that each beam illuminates only the correct subset of phosphors.) Note that this means that for CRTs a pixel is defined by the elec-

Three techniques for generating color in color electronic displays use red, green and blue subpixels and the principle of additive color. **a:** In the spatial technique the subpixels occupy different subareas of the pixel. **b:** In the sequential technique a broadband emitter is set to three successive intensities while an electronic color filter cycles through the primary colors. **c:** In the coincident technique transparent subpixels shine through one another. **Figure 2**



tron beams themselves and typically contains $2\frac{1}{2}$ phosphor spots of each primary color. Encompassing a nonintegral number of holes ensures that moiré patterns caused by interference between the raster lines and the shadow-mask hole pattern will be avoided.

To prevent flicker, the image is typically recreated approximately 60 times a second. European and Russian television sets operate at 50 Hz, while some computer graphics displays operate at 70–80 Hz. Each horizontal line that creates the image is traced from the viewer's left to his or her right, and then the beams return to the left edge to trace the next row. As the beams are deflected to trace a row, they are amplitude-modulated by the video amplifier in an analog fashion to reproduce the color intensity of the image. In television the signal for each row is itself analog. In computers the signal is digital but is converted to analog at the last CRT amplification stage. The electron beams that excite the red, green and blue phosphor dots on the faceplate of the CRT are independently modulated in parallel. The maximum rate at which the beams can be modulated from full "off" to full "on" as they are slewed from left to right defines the theoretical maximum number of pixels that can be written in a row. For color TVs the beam is modulated at 5 MHz, while VGA monitors use 12 MHz.

The beams are stepped digitally from row to row. The spacing of the rows is typically designed so that the bell-shaped profiles of beams from two adjacent rows overlap at the 50% intensity point. In NTSC television, odd and even rows are traced on alternate scans of the image; this is called 2:1 interlace and saves a factor of two in bandwidth. Computer images rarely use interlace because in graphics images it introduces artifacts such as line crawl (diagonal lines appear to have something crawling along them) and broken motion (odd and even lines show slices of a fast-moving object at slightly different locations).

The pixel matrix is made up of the number of rows in the vertical direction and the maximum number of beam modulations in the horizontal direction. In computer graphics, pixels are usually designed to be square; in television images, they are usually oblong. Today, Sony manufactures the highest-resolution color CRT, which will be used in the new US air route traffic control display. It has 2000 rows and 2000 columns and measures 20 inches by 20 inches.

Flat-panel displays use a different addressing system from that used in CRTs. (See figure 4.) In matrix-addressed FPDs, a row and a column electrode are associated with each subpixel. The equivalent electronic circuit for each subpixel is a "lossy capacitor." The external electronic row and column drivers must charge

the capacitor up to the appropriate level to make the subpixel turn on. As always, the image is typically refreshed about 60 times per second to prevent flicker.

In spite of this scheme's apparent simplicity, charging a matrix of identical lossy capacitors with different voltages to correspond to an image is an extremely complex task. First, by application of Kirchhoff's laws, we know that when one subpixel capacitor is charged to the desired voltage, there is electrical cross-coupling to all the other subpixel capacitors. Second, because of losses in the capacitor, the charge will not stay and must be refreshed. Third, a minimum time is needed to charge each capacitor because of the line resistance, the RC charging time constant and the distributed nature of the rows and columns. To minimize these effects, active elements such as thin film transistors are often added at each subpixel, and all the columns are addressed in parallel. (See figure 5.)

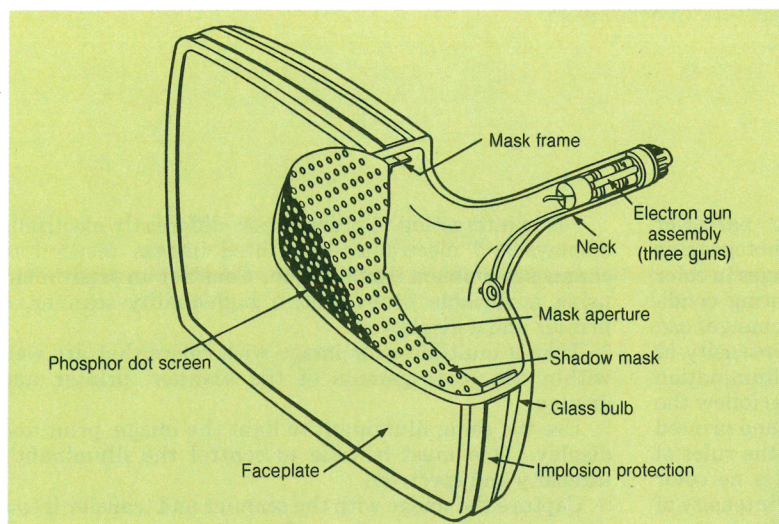
The most successful full-color FPD has been the active-matrix liquid crystal display that uses amorphous silicon as the semiconductor in thin film transistors at each pixel.⁵ (See figure 1.) All large AMLCDs actually in production use amorphous silicon. Displays using polycrystalline silicon are currently in the research and development phase. (Some small polycrystalline silicon AMLCDs are in production in Japan.) To date the highest-resolution AMLCDs in production have 480 rows and 640 columns, with each pixel made up of three subpixels, one for each primary color. The largest color video FPDs in production today measure about 6 inches by 8 inches. At the Japan Electronics Show in October of this year, Sharp demonstrated a 17-inch diagonal, 1280×1024 -pixel AMLCD with performance comparable to that of a 19-inch CRT workstation.

Materials for generating light

Almost all electronic color displays create a colored image by the light additive process using red, green and blue as the primaries. Most displays—including CRTs, LEDs and electroluminescent displays—do this directly: Electron bandgap transitions emit visible photons that stimulate the viewer's retina.

Another important FPD technology, called plasma panels, uses ultraviolet-excited phosphors. In this type of display the subpixel electrodes ionize mercury atoms in a gas mixture. Ultraviolet light emitted by the mercury excites an appropriate phosphor in the immediate proximity of the subpixel. The uv emission is blocked from exciting the phosphors for other colors associated with neighboring subpixels. The viewer sees the light from the individual phosphor dots associated with each subpixel.

In active-matrix liquid crystal displays, the photons



Shadow-mask color CRT has been the standard design for televisions and computer displays for 40 years. Each pixel is created by three electron beams that excite red, green and blue phosphor dots. The mask ensures that each beam hits only one type of phosphor. This technique is an example of scan addressing. **Figure 3**

seen by the viewer come from a classical fluorescent backlight that is continuously projected uniformly over the back side of the display. (See figure 5.) The light results from mercury uv excitation of a mix of phosphors that have appropriate emission peaks in red, green and blue. Nature has been kind to display designers by making three appropriate phosphors readily available in fluorescent lamps. The AMLCD acts as a controllable array of subpixel apertures, and each subpixel has a red, green or blue filter to allow passage of the corresponding color when electrically opened.

One display not based on additive color is the triple-cell liquid crystal projection display of Infocus Systems Inc, of Tualatin, Oregon.⁶ At each pixel, light from a broadband white backlight passes through three liquid crystal cells in series. Each cell can absorb a broadband portion of the spectrum—one can absorb the cyan portion of the spectrum, the second the yellow and the third the magenta. The opacity of each cell to its particular band can be controlled over a range from transparent to opaque. The color of the pixel is determined by the amount of each frequency that is transmitted through the triplet of cells. Thus the sequence of cells controls the color by the rules of subtractive color.

Visible-light-emitting diodes have often been used in displays. Individual colored LEDs can be assembled in rows and columns for this purpose. However, the development of full-color LED displays has been inhibited by the lack of an efficient blue LED. Blue LEDs that use SiC are becoming available in limited quantities. The use of LEDs in displays will be enhanced by the development of highly efficient AlGaAs LEDs for red and AlInGaP LEDs for orange, yellow and, most recently, green. These materials are five to ten times more efficient than previous LED materials.

The luminous efficiency of the light-generating process is very important in displays. Most of the power in a

television or personal computer is consumed by the display's light-generation process. As a rule of thumb, a display should output at least 1 lumen toward the viewer per watt of electrical power input to the display. At lower efficiencies dissipation of the heat generated in the display becomes a problem. The fluorescent lamps used with AMLCDs have luminous efficiencies of over 60 lumens per watt. However, the transmittance through the AMLCD and color filter is about 4%, giving a net efficiency of 2 lm/W.

Human factors and device independence

The very large color gamuts, or ranges of reproducible colors, of electronic displays exceed those of printing and photography and match what is necessary to portray most natural scenes. The color gamuts of electronic displays have very good brightness and color purity characteristics. However, they are weak in portraying dark yellows, dark cyans and dark magentas. Another problem is that the saturated (as opposed to pastel) blues and reds cause chromatic stereopsis: A red image distinctly appears to be in front of a blue image. This phenomenon occurs because of the difference in wavelength between blue and red and the lack of chromatic correction in the human eye lens.

Most types of electronic display do not need any ambient illumination. (The only exceptions are LCDs that lack a backlight, and these have very limited color gamuts.) For color consistency and device color independence, color electronic displays are best viewed under dark conditions. Surfaces such as the air-glass interface and the phosphor surface reflect the ambient illumination, and reflected light always reduces the contrast between pixels, desaturates or washes out the color and reduces the color gamut of a display. The phosphors of CRTs can reflect as much as 70% of the ambient light. CRTs cannot be viewed in bright ambient lighting unless they are built with heavy filtering, as are aviation-grade CRTs.

The primary colors in electronic displays tend to have

Effect of changing illumination on three types of images

	Original image	Display image	Printed image
Intensity changed	No change	Changed	No change
Spectrum changed	Changed	Different change from above	Changed
Spectrum and intensity changed	Same change as above	Further change from above	Same change as above

narrow-band spectral characteristics—much narrower than those of the dyes and pigments used in photographic prints and in printing. As a consequence, changes in color in electronic displays caused by varying lighting conditions tend to be more pronounced. Such color changes can occur in electronic displays when either the intensity or the chromaticity coordinates of the ambient illumination change. (See Robertson's article.) The changes follow the rules of additive color. In photographic prints and printed images, the colors change in accordance with the rules of subtractive color. For printed images there is no color change or reduction in color gamut if only the intensity of the illumination changes. However, color shifts do occur in prints when the spectral content of the illuminator is changed.

More and more users want to be able to take a given colored image from an electronic camera or scanner and reproduce it on an electronic display and an electronic printer. The biggest challenge is faithfully reproducing the same colors with both of the latter devices, given that almost all electronic displays use additive color and electronic printers use subtractive color. This difference, combined with the narrower-band spectral characteristics of displays, means that apparently identical displayed and printed images are in fact metameric—that is, they emit different spectra that give rise to the same perceived colors.⁷ (See Robertson's article.) A compounding complication is the different responses of displays and prints to the same ambient illumination. An electronic display is seen best and with the widest color gamut in dark ambient illumination; an electronically printed image is seen best in bright ambient light with a broadband spectrum.

It is interesting to observe how differently electronic displays and electronically printed images respond to changes in ambient illumination. Consider an experiment using a variable illuminant, a high-quality scanner, a printer and a display:

▷ Take a multiple-color image with colors that are well within the color gamuts of the scanner, printer and display.

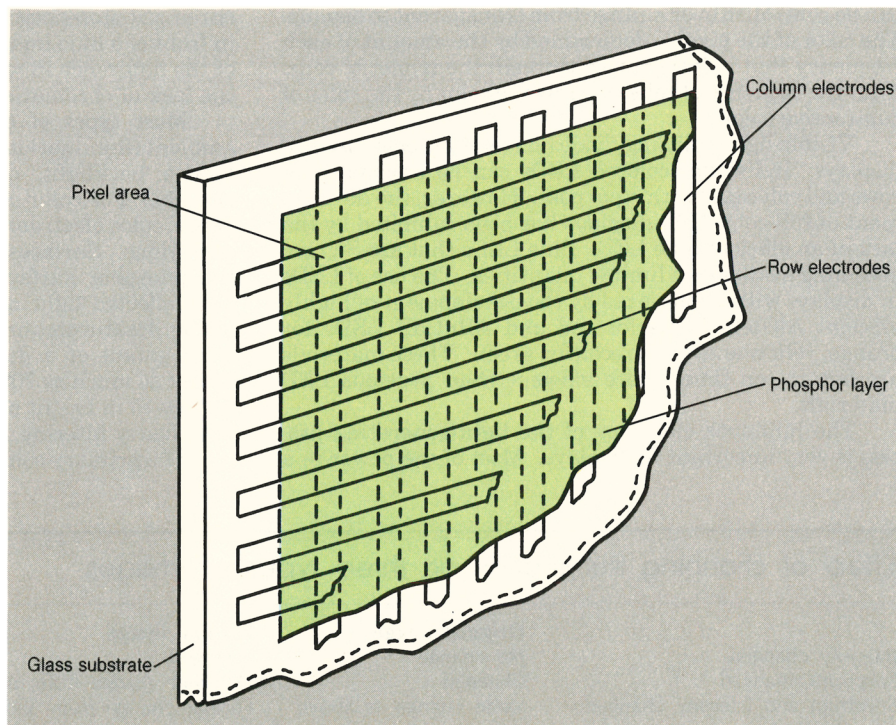
▷ Use the same illuminant to light the image, print and display. One must be able to control the illuminant's intensity and spectrum.

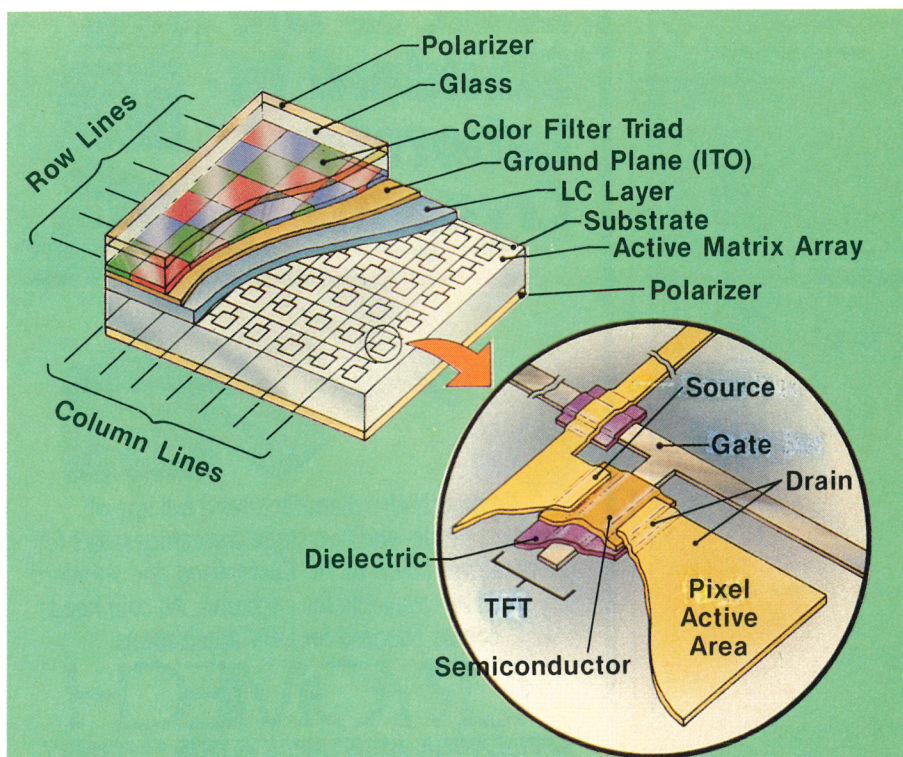
▷ Capture the image with the scanner and transfer it to a computer with appropriate software.

▷ Display the image and print it. View the display and the print under the illuminant and adjust the color balance algorithms in the computer software, the display and the printer until both images are perceived to be identical to the original. For a single image and a fixed illuminant, this should always be possible in a laboratory environment provided that none of the respective gamuts are exceeded. The three images will be metamers of one another.

▷ Now change the illuminant in three ways: Change its intensity; change its spectrum; and change its spectrum *and* intensity. The effects of these changes on the three colored images dramatize the difficulty of achieving color invariance. The perceived color changes for each illuminant change are summarized in the table on page 55. In general, there is no principle or condition that can be used to ensure that any of the colors in one image will be changed the same way as the corresponding colors in another of the images.

Matrix addressing is the usual technique used in flat-panel displays such as AMLCDs. Row and column electrodes deliver signals to a matrix of pixels that act somewhat like lossy capacitors and that may each have an active element such as a thin film transistor. (Adapted from ref. 3.) **Figure 4**





AMLCDs in production have an amorphous silicon thin film transistor (TFT) at each subpixel. A classical fluorescent backlight (not shown) shines uniformly and continuously through the display. Each transistor controls the state of a small region of the liquid crystal (LC) layer, which, in conjunction with the polarizer sheets and the color filter, controls in turn the intensity of light transmitted by a red, green or blue subpixel. **Figure 5**

Electronic color displays can be made almost completely independent of the ambient illumination by eliminating as much of the reflections as possible. This is achieved by adding antireflective coatings, polarizers, and neutral-density and narrow-band filters to the fronts of the displays. Inside the displays, electrodes, structures and nonemitting areas such as the mask are blackened.

Achieving independence from the ambient illumination is expensive for a shadow-mask color CRT but is relatively inexpensive for a color AMLCD. CRT phosphors are typically broadband 70% Lambertian reflectors—that is, the reflection is uniformly diffuse from every viewing angle. (A monochrome CRT has more Lambertian reflection than a color CRT: The black mask in a color CRT does not contribute to reflections.) Furthermore, because a CRT is curved, the user will always pick up specular reflections of light sources from the first glass surface of a CRT. In contrast to this an AMLCD is typically a black, flat display, making it easier to view in bright light than a CRT.

Display evolution

Broken down into its elements, an electronic display is a relatively simple device: a matrix of independently controllable light-emitting pixels and subpixels. However, in a VGA computer display, for example, there are nearly a million subpixels, each typically controlled to 1 part in 64 in intensity and reproduced at nominally 60 times a second. It turns out that electronic displays are the most complex electronic subsystems in existence, drawing on a wide variety of disciplines: They involve the physics of converting electrical power to photometric light, the chemistry of exotic materials, electro-optics, the physiology and psychology of human vision, circuit design and so on. And electronic displays push the state of the art in each of these disciplines. For example, the scientific community still does not have a sufficiently detailed analytical model of polycrystalline phosphors to enable the synthesis of new ones. This is not a new situation for

display technology: The CRT was invented in the late 1890s, before the electron had been characterized.

The developmental evolution of an electronic display from research breadboards to production units is measured in decades. It has taken engineers almost 40 years and billions of dollars to get a color video flat-panel display into production. The only technically successful color video FPD in high-volume production today is the AMLCD, and some analysts still question its economic viability. Nevertheless it is anticipated that by 1995 1 million high-performance AMLCDs will be produced per month in Japan for portable PCs and related products. (At present, Japanese companies are the leading manufacturers of AMLCDs.) Although such displays will remain much more expensive than CRTs, their flatness will allow the development of many new products that are impractical with CRTs. A new chapter in the evolution of displays has been made possible by the ability of LCDs to incorporate full color.

References

1. G. Wyszecki, W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., Wiley, New York (1982).
2. H. G. Grassmann, *Ann. Phys.* **89**, 69 (1853); *Philos. Mag.* **7**, 254 (1853).
3. L. E. Tannas Jr, *Flat-Panel Displays and CRTs*, Van Nostrand Reinhold, New York (1985).
4. *Display Devices*, suppl. to *J. Electron. Eng.*, **6** (Fall 1992).
5. L. E. Tannas Jr, W. E. Glenn, T. Credelle, J. W. Doane, A. H. Firester, M. Thompson, Japanese Technology Evaluation Center Report on Display Technologies in Japan, report PB92-100247, Natl. Technical Inf. Service, US Dept. of Commerce, Washington, D. C. (June 1992).
6. A. R. Conner, in *1992 SID Int. Symp. Digest of Technical Papers, Applications Notes*, Soc. for Information Display, Playa del Rey, Calif. (1992), p. 109.
7. F. W. Billmeyer Jr, M. Saltzman, *Principles of Color Technology*, Wiley, New York (1981). ■