# Specifying Color in Office Documents

Robert R. Buckley*

Abstract: The current trend towards device-independent color in office documents is part of a larger trend towards open systems, with an emphasis on blind interchange and systems integration. Several vendors already offer device-independent color, and international standards-making organizations are developing device-independent standards for representing color in documents.

Device-independent color means representations based on human color vision, as opposed to device-dependent color representations such as ink dot percentages. Different approaches to device-independent color will be described, along with their implications and relationship to device-dependent color. The talk will also describe the current state of international color standards for office documents.

## Graphic Arts and Desktop Publishing

To set the stage for describing an office color document system, I will first describe the more familiar worlds of color desktop publishing and electronic prepress systems. The distinction between the way they use color and the way an office system would, is essentially the difference between device-dependent and device-independent color.

Figure 1 shows the block diagram of a simple color desktop publishing system, with a scanner, workstation, and color printer. The same basic components are also found in a graphic arts system, where they cost about two orders of magnitude more than their desktop counterparts. In a graphic arts system, the workstation is a color electronic prepress system while the printer could be either a direct digital proofer or a film plotter for making color separation films. I want to focus on how color is described in the files exchanged between the workstation and the printer in a system, and between different systems.

* Webster Research Center, Xerox Corporation

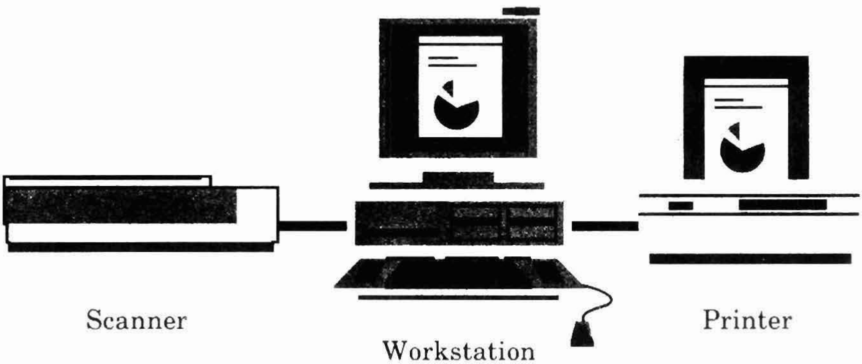


Figure 1. Block Diagram of a Desktop Publishing or Graphic Arts System

The files interchanged in these systems use cyan, magenta, yellow, and black (or CMYK) ink amounts to describe a color. Because not all printers use the same inks, or even when they do, apply them in the same way, the same CMYK values can lead to different results on different printers. For this reason, CMYK color values are device dependent: the desired result is linked to a specific printer or type of printer. The color model that relates values in the file to the appearance of the color is supplied by the printer.

Several file formats in use today offer device-dependent color. Among them are standards developed by the IT8 Committee of ANSI for the exchange of scanned images between color electronic prepress systems (ANSI, 1988) and between prepress systems and on-line digital color proofers. These standards support the exchange of CMYK, complementary RGB, and custom or user-defined color separations. The separation data is normally given as halftone dot percentages.

Among vendor-developed file formats, PostScript® (Adobe, 1989) and Printer Control Language or PCL 5 (Hewlett Packard, 1990) use device-dependent color: both support CMYK and complementary RGB. While device independent in many respects, both are designed for the device-dependent control of color printers.*

---

* Adobe has announced plans to add device-independent color to PostScript Level 2.

Device-dependent CMYK values occur in several situations. Because the printing press is a major investment in most graphic arts systems, different printers are not an issue and documents will always be printed on the same printer. Therefore, these systems use device-dependent CMYK values, tuned to the specific printer around which the system is built. Device-dependent CMYK are also used when the printers are nearly identical, or when the differences between printers (and their output) are within acceptable limits.

## Office Document Systems

Instead of systems with dedicated workstation-printer pairs, consider a network (Figure 2), connecting different color workstations to a wide variety of printing options, such as a desktop color printer, a shared color printer, an imagesetter, or even a black-and-white printer. The connections may use a local area network, magnetic tape transport, satellite links, phone lines, or a combination of these.

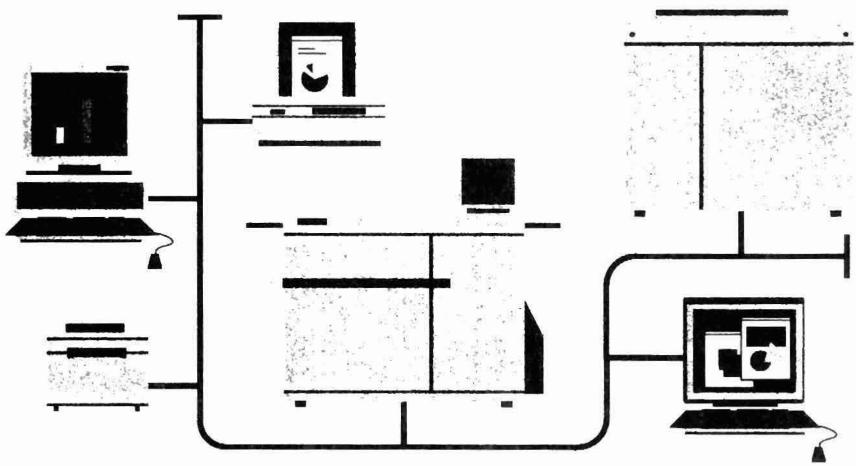

Figure 2. Block Diagram of an Office Document System

The network will have a shifting population of printers, as units are added, removed, upgraded or replaced. It will support a wide range of users whose needs can vary from simple, easy-to-use color to graphic arts-level precision. Workstation-printer pairs are continually being formed and dissolved to meet the needs of specific users and jobs. This scenario describes today's modern office.

Consider a user creating an office document with colored logos, highlighted text, some colored illustrations and possibly a color picture. The resulting document is distributed electronically to a large number of recipients either in editable form, as a mail note, or as a printable document. Some of the recipients may view the document on a color monitor; others, on a high-resolution black-and-white display. Some will print it on their local color printer while others will store it on a file server for retrieval and printing at a later date.

The goal in this scenario is, first of all, to capture the creator's intent as to what the document should look like, including the appearance of colors, and then preserve it systemwide across different recipients and printers. This raises several issues not encountered in the graphic arts or desktop publishing systems described earlier.

**Open System**: Implementations will be based on public-domain rather than proprietary standards. This will allow vendors equal access to the market for system components and give users the widest choice in selecting components.

**Blind Interchange**: The creator of a document will usually not know beforehand how, when, or where a document will be used or printed. Therefore, a document creator should make as few assumptions as possible about how the document will be processed.

**Systems Integration**: The portability of user applications and the interchangeability of system components will give users the greatest latitude in configuring seamless systems to meet their needs.

**Black-and-White Compatibility**: A shift to color in documents should not ignore the large existing base of black-and-white printers; they must be able to make credible prints of color documents.

**Color Expertise**: Accommodating casual as well as the sophisticated color users will shift the burden of color reproduction from the user to the printer. There will be more emphasis on easy access and low entry costs. The printer will have to know how to reproduce the colors the user wants, rather than the user having to know how to obtain the colors the printer reproduces.

**Color Document Lifespan**: A stored or archived document, when retrieved and printed, should look as it did the day it was created, even though some system components may have changed.

## Device Independence

Device independence is a way of enabling the different printers in an office document system to produce the same result from the same file. A device-independent file specifies the information content or appearance of a document (Warnock and Wyatt, 1982), describing the result the user wants to see, not how the printer would produce it. In operational terms, device independence means dependence on an ideal or universal device, which provides the reference coordinate system used to represent a document and the elements that it contains.

Each printer in the system would convert the ideal, device-independent description of the document into the device-dependent description it would actually use to produce the document. A device-independent approach to page layout, for example, could describe the ideal location of elements on a page in units such as inches using an orthogonal Cartesian coordinate system whose origin is the lower left corner of the page. A raster printer would convert this ideal position into a corresponding line number and pixel position on the line.

Device-independent color is based on a human color vision model, formalized by the CIE system of colorimetry (CIE, 1986): the ideal device is the CIE Standard Observer and the reference coordinates are XYZ tristimulus values (or a transformation of them). XYZ values can be computed from the spectral energy distribution of a color stimulus or measured directly using a colorimeter. The Y tristimulus value, or luminance, correlates with stimulus lightness. The CIE defines two standard observers: current color document standards all use the 1931 or 2° Standard Observer, rather than the 1964 or 10° observer.

XYZ values describe only one aspect of a color stimulus, and in the case of color hardcopy, for only one illuminant. The ultimate goal is specifying color appearance, which requires knowing several variables besides XYZ values; they include the surrounding stimuli, ambient illumination, light intensity, and stimulus size. Although the relationship between XYZ values and color appearance is still a research topic (ISCC, 1989), it can be considered fixed when viewing conditions are standardized.

Figure 3 shows a simple model for the device-independent interchange of documents between multiple workstations and printers. The workstation is interfaced to the printer via reference coordinates, either directly by using reference coordinates in the file, or indirectly by using the workstation user's coordinates but including the transformation $T_i$ to reference coordinates. In the direct case, the conversion to reference coordinates is applied at the workstation; in the indirect case, it is deferred to the printer where it can be combined with the transformation $S_j$ from reference to printer coordinates to form a single overall transformation $S_jT_i$ from user to printer coordinates.

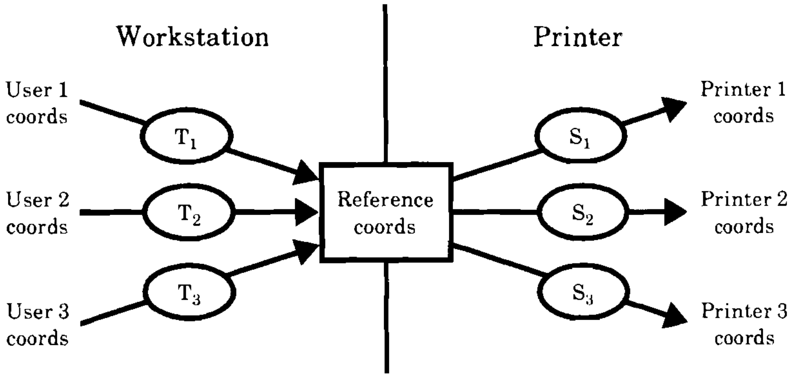


Figure 3. Model for Device-independent Interchange

In the case of device-independent color, the reference coordinates are XYZ or a transformation of them. The printer transformation ($S_j$ in Figure 3) converts XYZ values to the device-dependent printer coordinates (CMYK values in the case of a process color printer) that lead to equivalent XYZ values. The printer transformation is usually determined empirically by calibrating the printer—measuring the XYZ values of the results printed with various combinations of CMYK values.

For the printer to convert XYZ to equivalent CMYK values, the XYZ values corresponding to the possible combinations of CMYK must be known and predictable. Printer calibration is an essential step in implementing device-independent color: it establishes and then maintains a desired level of printer performance by using CIE standards for measuring the XYZ values of the printer output.

The range of XYZ values resulting from all possible CMYK combinations on a process color printer defines the printer's gamut. Complications arise when a user requests XYZ values outside the printer gamut and which can't be printed. Nonprintable colors can be handled either by the user at the workstation as part of the color selection process, or by the printer as part of the transformation $S_j$. While nonprintable colors are an important practical issue, most interchange standards take the narrow view that they are not an interchange problem.

While there is general agreement to base device-independent color on XYZ tristimulus values, they are themselves seldom used directly for interchange. Instead, transformations of XYZ values are preferred, which offer added value such as RGB compatibility, better uniformity, or true chrominance coding.

Among the transformations of XYZ values that are candidates for reference coordinates are RGB—several versions in fact, depending on the choice of primaries and white point; luminance and chrominance coordinates, including YIQ, YUV, and YES; luminance and chromaticity coordinates, such as Yxy; and CIE-recommended uniform color spaces, CIELAB and CIELUV, and their variants, such as TekHVC.

## Industry Standards for Device-Independent Color

Several industry standards support device-independent or CIE-based color. Among them is the Tag Image File Format or TIFF, developed by Aldus and Microsoft for the interchange of scanned color images. Published in 1988, Version 5.0 defines an RGB color model for interchange (Aldus, 1988). The RGB values in a TIFF file are related to XYZ values as follows:

$$[\,X\ Y\ Z\,] = [\,f_R(R)\ f_G(G)\ f_B(B)\,]\,\mathbf{T} \qquad (1)$$

where **T** is a $3 \times 3$ matrix and $f_R$, $f_G$, $f_B$ are 1-dimensional lookup tables. The parameters of the transformation—**T**, $f_R$, $f_G$, and $f_B$—are specified in the TIFF file. Rather than giving **T** directly, the file specifies the chromaticities of the R, G, B primaries and white point, from which the matrix can be reconstructed. The default values are based on the SMPTE standards for studio monitors (SMPTE, 1986).

The RGB color model in TIFF is an example of the use of indirect reference coordinates: the workstation places RGB values in the file, along with the parameters of the RGB-to-XYZ transformation. Therefore, the file specifies the color model that defines the color corresponding to the values in the file.

Although TIFF was clearly intended for use with tone-corrected red, green, and blue values, a TIFF file can contain any values that satisfy Equation (1). With the right choice for the f-functions and T, the TIFF file could use CMY values that are the complements of red, green, blue, or even XYZ values. Because a printer cannot assume that R, G, B mean red, green, blue, it will in general have to examine the parameters of T, looking for a clue to the true identity of the RGB values, or else blindly go ahead and apply T, either by itself or in combination with the transformation to printer coordinates.

In 1988, Tektronix introduced the TekHVC™ color model (Taylor et al., 1989). It is a variant of the CIELUV uniform color space, using the cylindrical coordinates H, V, and C, correlated with hue, value (or lightness), and chroma, respectively. The TekColor™ Color Management System uses the TekHVC color values as both user and reference coordinates to characterize printer gamuts and in the user interface for selecting printable colors. Whether the transformation from TekHVC values to printer coordinates occurs at the workstation or the printer depends on the system configuration.

First described in 1986, the Xerox Color Encoding Standard or CES is designed for all document interchange applications, including mailing, filing, and printing (Buckley, 1986; Xerox, 1989). It offers device-independent RGB, YES*, and CIELAB color models.

The CES RGB color model is based on the SMPTE standard phosphors for studio monitors but uses the D50 standard illuminant from the graphic arts (ANSI, 1989). Making the graphic arts illuminant part of the RGB definition eliminates color conversions between calibrated graphic arts systems, where the color accuracy requirements are likely to be the most exacting.

The CES YES color model is a simple luminance-chrominance color model; Y is the luminance or gray component and E and S are the red-green and yellow-blue color differences, respectively. YES is an uninflected color model, based on the modern SMPTE phosphors and without built-in gamma correction, unlike luminance-chrominance models such as YIQ and YUV.

* YES is the name Xerox uses for an independently-developed variant of the ARgYb color space (Naiman, 1985), which is itself a modified version of a color space based on opponent color theory.

These RGB and YES values are related to XYZ as follows:

$$[\,X\;Y\;Z\,] = [\,f_R(R)\;\; f_G(G)\;\; f_B(B)\,]\;\mathbf{T_1} \tag{2a}$$

$$= [\,f_Y(Y)\;\; f_E(E)\;\; f_S(S)\,]\;\mathbf{T_2} \tag{2b}$$

The f-functions are 1-dimensional lookup tables; $\mathbf{T_1}$ and $\mathbf{T_2}$ are $3\times3$ constant matrices, defined in the Xerox standard. Compared to device-dependent values, whose relationship to XYZ is defined by the printer, and TIFF RGB values, whose relationship is defined by the file, the Xerox RGB and YES values have a fixed relationship to XYZ, defined by the standard. Therefore, there is no ambiguity about what color the RGB and YES values represent.

Because the Y in YES is the luminance, the YES color model is black-and-white compatible: a black-and-white printer can use the Y value to obtain the gray equivalent of a color without extra computation. The CES also offers the CIE-recommended CIELAB uniform color space as a nonlinear black-and-white compatible color space for more advanced requirements. CIELAB is used instead of CIELUV because of the perceived preference of the graphic arts industry for CIELAB.

Besides these device-independent color models, CES also offers a device-dependent color model for process control applications, CMYK compatibility and printer calibration. This model is an extended version of the ANSI IT8 color model for graphic arts applications.

Besides providing the basic infrastructure for device-independent color, the CES also addresses advanced topics in color interchange. It allows the user to specify color tolerances based on CIELAB color differences. The tolerance parameter makes it possible, for example, for a printer to distinguish between two reds that have the same numeric values, although one is meant to be a very specific custom red color while the other can be anything, as long as it looks "red." A loose tolerance turns the colorimetrically-exact RGB model into the uncalibrated RGB model popular in computer graphics, where any red phosphor, for example, can be substituted for the SMPTE-standard red phosphor.

The standard also offers appearance hints, optional information that capable printers can use to preserve color appearance under different viewing conditions, as when a document is transferred from a monitor to hardcopy. Because not all workstations and printers in a system will use the same white point or illuminant, the CES also describes a simple method for adjusting color values between system components with different white points.

## International Color Standards

International standards-making bodies are now also developing device-independent color standards for electronic documents. The international forum for this activity is ISO/IEC JTC1/SC18/WG5.* SC18 is responsible for text and office systems and WG5, for document content. WG5 meetings are attended by representatives from the standards-making bodies of member countries. The countries that have been most active in color standards work are Canada (Chair), the Federal Republic of Germany, France, Japan, the United Kingdom, and the United States, whose member body is the American National Standards Institute or ANSI.

The US position on SC18/WG5 color activities is developed by ANSI/X3V1/TG5, in liaison with IT8. The ANSI corporate members that have been active in the color work are IBM, 3M, Eastman Kodak, Hewlett Packard, and Xerox.

The focus of the color standards activity within SC18/WG5 is ODA or the Office Document Architecture (ISO 8613), a joint ISO/CCITT standard for document interchange (ISO, 1988). The current standard only supports documents with black-and-white characters, geometric objects, and rasters; under development is an addendum to the standard that would add color (ISO, 1990b).

The ODA Color Addendum defines XYZ as the reference model and proposes four color interchange models: RGB, CMYK, CIELAB, and CIELUV. All parts of a document may use the RGB and CMYK models, while only raster data can use CIELAB and CIELUV. The addendum also defines color tolerances, similar to those in the Xerox Color Encoding Standard.

According to the Color Addendum, an ODA data stream or file would contain calibration data, defining how the color values it uses for interchange are related to the XYZ reference values. Calibration

* International Standards Organization/International Electrotechnical Commission Joint Technical Committee 1/Subcommittee 18/Working Group 5

data always includes the XYZ values of the white point. For CIELAB and CIELUV, this is all that is needed to relate them to XYZ.

For the ODA RGB color model, the calibration data would include three 1-dimensional lookup tables $f_R$, $f_G$, $f_B$ and a 3×3 matrix **T**. The RGB values in an ODA file would be related to XYZ values as follows:

$$[\,X\;Y\;Z\,] = [\,f_R(R)\;\;f_G(G)\;\;f_B(B)\,]\,\mathbf{T} \qquad (3)$$

This is similar to the TIFF RGB model, except that the 9 coefficients of the matrix T are given directly. With the proper choice of matrix T, the color coordinates labeled RGB can in fact be XYZ, YES, or various kinds of RGB values. But YIQ and YUV are not instances of the ODA RGB model because they are based on gamma-corrected RGB values (Pritchard and Gibson, 1980). The default matrix is based on SMPTE phosphors and CIE Illuminant D65.

The CMYK calibration data would include three 4-dimensional lookup tables $f_X$, $f_Y$, $f_Z$. CMYK values are related or transformed to XYZ using table lookup with interpolation:

$$X = f_X(C, M, Y, K) \qquad (4a)$$

$$Y = f_Y(C, M, Y, K) \qquad (4b)$$

$$Z = f_Z(C, M, Y, K) \qquad (4c)$$

The user can control the precision (and amount) of the calibration data by specifying the number of index values along each dimension: the more index values, the finer the interpolation grid, the larger the table, and the more accurate the conversion to XYZ. Setting the number of index values to zero makes it possible to reduce the dimensionality of the table and of the color values in the file. To indicate CMY color values and three 3-dimensional lookup tables, for example, the user would set the number of index values for K to 0. No default lookup tables are defined.

The orientation of the ODA Color Addendum is preserving existing color data bases and current device-dependent interchange practices by retrofitting device independence (in the form of calibration data) to existing color data.

The calibration mechanisms for the two main ODA color models—a matrix and a lookup table with interpolation—are general enough to accommodate most device-independent but more importantly most device-dependent color models used for interchange today. This means, for example, that ODA can use already-existing CMYK color data by simply inserting the CMYK values in the ODA data stream, along with the appropriate calibration data. The RGB and YES models of the Xerox Color Encoding Standard are compatible with ODA color; because the TekHVC color model uses cylindrical coordinates, it is not.

The generality of the calibration mechanisms means that a printer can't take the RGB and CMYK models literally, as they can be used with any color values that satisfy Equations (3) and (4). The only way a printer can know for sure what color values it has is by examining the calibration data or by executing the transformation $T_i$ based on the calibration data to get XYZ values. The ODA Color Addendum has proposed using a printable string in the data stream as an aid in identifying color values, but there are no conventions describing its use.

While the user-defined ODA color models will save work for the document creator, they will almost always mean more work for the printer. Using calibration data defers the transformation $T_i$ to reference coordinates from the workstation to the printer, where it can be implemented as part of one overall transformation $S_iT_j$. While this is usually more work for the printer, there will be occasions when the overall transformation is the identity transformation, in which case the printer can use the color values directly without a color conversion.

As this is being written, the ODA Color Addendum is in the Draft stage and out for balloting. Draft is the next-to-last stage in becoming a standard; the major technical issues have been settled and the only changes made at this stage are editorial. Once the addendum is accepted, the ODA standard will be amended to include it.

As the ODA Color Addendum is winding down, a new activity underway within SC18 is TOSCA (ISO, 1990a). TOSCA is the acronym for "Text and Office System Color Architecture": its goal is to establish a single approach to color for all text and office system applications. Besides ODA, this would include the Standard Generalized Markup Language or SGML (ISO 8879) and the Standard Page Description Language or SPDL, still under development.

The TOSCA project is still at the very first stage in the development of an international standard, and only user requirements have so far been discussed. The progression from user requirements to standard goes via working draft, draft proposal, and then draft; a draft proposal is expected in 1992.

## Conclusions and Trends

The emphasis on device independence in both industry and international standards has led to color standards for document interchange based on the CIE system of colorimetry and XYZ values. Implemented in this way, color device independence requires printer calibration and smart printers that are able to transform the device-independent color values they receive into the device-dependent color values they would use to print a document. Device-independent color standards use XYZ-based color models that can either be defined in the file or in the standard.

Device independence will enable the clean separation of the design and printing of office documents. This will have important implications for the links being forged between desktop publishing and color electronic prepress systems. While the latter have traditionally used device-dependent CMYK values, desktop publishing is evolving along with office color towards device-independent CIE-based color.

## Literature Cited

Adobe Systems Inc.

1989. "Color Separation Conventions for PostScript® Language Programs," Technical Note #5044, June 21, 1989

Aldus Corp.

1988. "Tag Image File Format Specification, Revision 5.0," Aldus/Microsoft Technical Memorandum, August 8, 1988

ANSI, American National Standards Institute, Inc.

1988. "User Exchange Format for the Exchange of Color Picture Data Between Electronic Prepress Systems via Magnetic Tape," IT8.1-88

1989. "Graphic Arts and Photography–Color Prints, Transparencies, and Photomechanical Reproductions–Viewing Conditions," PH2.30-89

Buckley, R. R.
1986. "Interpress Standard for Color Printers and Workstations," Laser in Graphics Conference Proceedings, Vol. II, pp. 24-28

CIE, Commission Internationale de l'Éclairage
1986. "Colorimetry," (Central Bureau of the CIE, Vienna), 2nd ed., Publication CIE No. 15.2

Hewlett Packard Co.
1990. LaserJet III Technical Reference Manual, Part No. 33449-90903, Edition 1E0390

ISCC, Inter-Society Color Council
1989. Vision and Color Appearance Symposium, Chicago, Illinois, April 12, 1989

ISO, International Standards Organization
1988. "Information Processing–Text and Office Systems–Office Document Architecture (ODA) and Interchange Format," ISO 8613
1990a. "Proposal for a New Work Item: Text and Office Systems Colour Architecture," Document No. ISO/IEC JTC1 N656
1990b. "Revised Text of PDAD to ISO 8613 to Add Colour Capability," Document No. ISO/IEC JTC1/SC18 N2208

Naiman, A.
1985. "Color spaces and color contrast," The Visual Computer, (Springer Verlag, Berlin) vol. 1, pp. 194-201

Pritchard, D. H. and Gibson, J. J.
1980. "Worldwide Color Television Standards-Similarities and Differences," SMPTE Jnl., vol. 89, no. 2, pp. 111-120

SMPTE, Society of Motion Picture and Television Engineers
1986. "Color Monitor Colorimetry," Proposed Recommended Practice 145, SMPTE Jnl., vol. 95, no. 9, p. 980

Taylor, J. M., Murch, G. M., and McManus, P. A.
1989. "TekHVC™: A Uniform Perceptual Color System for Display Users," Proc. of the SID, vol. 30, no. 1, pp. 15-21.

Warnock, J. E. and Wyatt, D. K.
1982. "A Device Independent Graphics Imaging Model for Use with Raster Devices," Computer Graphics, vol. 16, no. 3, pp. 313-319

Xerox Corp.
1989. "Color Encoding Standard," (Xerox Systems Institute, Sunnyvale, CA), XNSS 288811