The Color Gamut Limits of Halftone Printing with and without the Paper Spread Function

Peter G. Engeldrum*

Imcotek, Winchester, Massachusetts 01890, imcotek@aol.com

Abstract

Color printing using halftoning techniques is becoming ubiquitous on the desktop. Accuracy and consistency of the color printed by these devices depends on, among other things, the optical properties of the paper. The general influence of light scattering within the paper has been known for half a century, but calculation of the effect on the color gamut is complex and depends on the details of the paper optical spread function and of the halftone pattern. Two limiting cases are analyzed, no-paper-scattering and complete-paper-scattering (optical paper spread function is much larger than the halftone cell size). Simple models are presented for the limiting cases of halftone printed color images, and the predictions are compared with measurements of single-colorant samples from a wax thermal printer. Results show that the simple models bound the single-colorant colors produced by the wax thermal printer. The simple model predicts that the major effect of the paper scatter on single-colorant images is the shifting of colors along the CIE L*a*b* locus. This can result in color differences for the two scattering extremes of 20 CIE L*a*b* units for the same fractional dot area. Two different CIE L*a*b* loci are computed for the same colorant, under the no-light-scattering versus complete-light-scattering conditions, but this is a much smaller effect.

Introduction

The explosion of inexpensive computer technology is making color imaging widely available. This is particularly true on the desktop, where color printers, mostly ink jet, are replacing the black and white impact dot matrix varieties. This rapid expansion of color printing has not been without problems. Color printing has, historically, been controlled by experienced people who are knowledgeable about printing technology and who have developed substantial expertise on how to manipulate the technology. On the desktop, either in the home or office environment, most users have goals other than becoming experts in the technology used by their color devices. A major step toward making "good color" a reality for experts and nonexperts alike has been the incorporation of so-called color management systems (CMS) into new computer operating systems. To "manage the color," the CMS needs some information about the color printing device. This usually takes the form of a *device profile*, which translates some specification of color using CIE colorimetry into the colorant amounts required to render the color on the device.

The most successful type of device profile for color printers is some form of multidimensional look-up table (LUT). Generation of such a table can be accomplished primarily in two ways: (1) Print an array of colors with known amounts of colorants and measure the resultant color and (2) use a color imaging model. Printing and measuring an array of colors is time-consuming and prone to the random variations of the printer unless the measurements from some number of samples, both from printers and replicates from the same printer, are averaged. The imaging performance of most color imaging technologies varies with substrate, environment, past printing history, etc. For high-quality color imaging, device profiles that cover these variable factors must be available to the color matching system. A simpler, and possibly more accurate, approach would be model-based, provided that unbiased estimates of low variance can be made of the model parameters. The major impediment to the model-based approach to LUT generation is the lack of a good color imaging model for halftone printing.

The most popular and least expensive color printers place various colored dots of fixed or varying size on a substrate in a grid pattern of from 300 to 720 dots per inch. Color formation by these devices uses halftone printing, wherein the fractional area covered by the three or four colorants varies. For conventional color halftone printing that uses rotated halftone screen angles or the various forms of error diffusion or stochastic screening, the model first proposed by Neugebauer[1] in 1937 provides a starting point for the model-based approach. These equations describe the CIE tristimulus values of the halftone, given the fractional areas of cyan, magenta and yellow colorants and the tristimulus values of the eight "Neugebauer primaries": "white" (paper), cyan, magenta, yellow, red, green, blue, and three-color black. The basic tristimulus equations, where T is used for CIE X, Y, or Z, are:

$$
\begin{aligned}
T &= \sum_{i=1}^{8} f_i T_i, \\
f_1 &= (1 - c)(1 - m)(1 - y), \\
f_2 &= (1 - m)(1 - y), \\
f_3 &= (1 - c)(1 - y), \\
f_4 &= y(1 - c)(1 - m), \\
f_5 &= my(1 - c), \\
f_6 &= cy\,(1 - m), \\
f_7 &= cm\,(1 - y), \\
f_8 &= cmy.
\end{aligned} \tag{1}
$$

In Eq. 1 values of T_i are the tristimulus values for "white," cyan, magenta, yellow, red, green, blue, and black, respectively. These primaries could be measured from a transmittance image or a reflectance image. Although these equations are much maligned, they are quite robust for a wide array of image microstructures. Neugebauer's[1] original derivation was probabilistic in approach. However, the same set of equations can be derived assuming that the cyan, magenta, and yellow halftone layers are superimposed, or as trilinear interpolation, and, as shown in the Appendix, assuming complete randomness of the eight primaries.

It is well known, however, that Neugebauer's equations often do not predict the measured tristimulus values of colored halftones on paper.[2-6] The primary reason, first proposed by Yule and Nielsen in 1951,[7] is the scattering of light within the bulk of the paper, a phenomenon that is now known as the *paper optical spread function*. Their "fix" was to propose a factor, the "n-factor," that modifies the tristimulus values of the eight primaries by raising them to a power of $1/n$ and then taking the n-th root of the result. (Actually the modified form of the Neugebauer equations was first described by Pobboravsky and Pearson.[3] Yule and Nielsen[7] gave the result for a single-color halftone.) Although the incorporation of the n-factor often works in accounting for the effect of light scattering, this formalism does not accurately portray the optical physics of the problem. Various approaches have been put forth to increase the accuracy of the equations, usually incorporating some form of n-factor.[8-10]

In 1974 Lehmbeck[11] described a model for images on paper that explicitly included the paper spread function and colorant layer spatial characteristics. Later, Ruckdeshel and Howser[12] conducted a complete exploration of a one-dimensional version of Lehmbeck's model and showed that the n-factor is theoretically bounded by one and two. Recent work[13-15] has shown that both the paper and the dot reflectance change with fractional dot area. This is contrary to the Yule–Nielsen single n-factor approach, which assumes the paper and the dot reflectance (tristimulus values) are constant for a fixed set of printing conditions. As Pearson[16] has noted, the optimum n-factor is a function of fractional area, colorant level, and the uniformity ("formation") of the sheet of paper.

A direct approach would be to calculate the spectral reflectance factor as a function of x-y position on the paper, but, unfortunately, with the current models this is a complex undertaking. Of more immediate practical value would be the determination of the colorimetric bounds on the halftone colors for two cases of light scattering by the paper: (1) the Neugebauer equations apply (an infinitely narrow paper optical spread function or no lateral scattering within the paper), or n-factor = 1.0, and (2) the paper spread function is much larger than the halftone screen spatial frequency, or n-factor = 2.0. Experimental data for the reflectance or color of the paper between the dots, and the dot color, have shown various relationships with fractional dot area.[13-15] With any scattering, or finite-sized paper spread function, reflectance limits of the paper between the dots are bounded by the

paper reflectance, at 0 fractional dot area, and the paper reflectance and colorant layer transmittance product at fractional dot area = 1.0.[13] Dot reflectance is bounded by the same paper reflectance–colorant layer transmittance product, but at dot area = 0.0, and by the dot reflectance at fractional dot area = 1.0. The upper bound of the paper reflectance in the absence of any paper spread function is, as assumed in the simple Neugebauer model, constant for all dot areas. But what is the lower bound on the paper reflectance–dot area function for the complete scattering case? Also what is the upper bound of dot reflectance versus area for the same conditions?

This report describes the relationships for these two limits and provides a comparison of the predictions of single-colorant halftone scales with experimental results from a wax thermal transfer printer.

Analysis

The functional relationship between the paper and dot reflectance, or tristimulus value, and the fractional area covered by colorants depends on the sequence of forming the halftone dot, the dimensions of the halftone cell, and the extent of the spread of light within the paper. In the case where the spread of light is very much *smaller* than the halftone cell dimension and the printer does not use dot-on-dot printing,[17] the Neugebauer equations accurately represent the printed color. Such practical realization might be, for example, a 65-halftone cells/inch pattern printed on a cast-coated paper. A halftone transparency would be another example. Knowing the equations that define the color when the spatial extent of the spread of light within the paper is much *greater* than the halftone cell size will enable us to define the bounds of the color: the Neugebauer equations for the very small scattering case, and the new set of equations for the complete scattering case.

Complete Scattering Limit

Imagine that we have three separate halftone colorant layers (cyan, magenta, and yellow) randomly combined in a sandwich. We do not require, and in fact do not want, the three layers to be spatially coherent, which is one reason for rotating the screen angles in commercial printing. The space average spectral transmittance of the sandwich, $t(\lambda)$, is given by Eq. 2, the product of the transmittances of each of the three halftone layers;

$$t(\lambda) = [t_b(\lambda)(1 - c) + ct_c(\lambda)]$$
$$\cdot [t_b(\lambda)(1 - m) + mt_m(\lambda)][t_b(\lambda)(1 - y) + yt_y(\lambda)], \quad (2)$$

where c, m, and y are the fractional dot areas of cyan, magenta, and yellow; t_b is the base spectral transmittance; and t_c, t_m, and t_y are the colorant spectral transmittances. It can be readily shown that the expansion of Eq. 2 yields the spectral transmittance version of the Neugebauer equations.[4,8] Assume that the transmittance pattern modulates the incident light propagating toward a paper substrate of spectral reflectance $R_p(\lambda)$. We assume that the incident light is completely scattered, or averaged, by the paper before being reflected. When the light emerges from the paper, the fraction returned at each wavelength

is just the spectral transmittance, given by Eq. 2, times the spectral reflectance of the paper. Equation 3 summarizes the reflectance of the paper under the halftone pattern:

$$R_1(\lambda) = t(\lambda)R_p(\lambda). \qquad (3)$$

Because the halftone pattern is small compared with the spatial extent of the light scattering within the paper, the periodic structure is lost, and the surface of the paper, under the halftone sandwich, is uniform, with spectral properties given by Eq. 3. Upon reflection from the paper, this uniform color light is modulated by the spectral transmittance of the halftone sandwich, $t(l)$, given by Eq. 2. Finally, the average reflectance is given by the combination of Eq. 3 and Eq. 2. Thus,

$$R(\lambda) = t(\lambda)^2 R_p(\lambda). \qquad (4)$$

Note that we have not considered first surface reflection effects in this formulation.

For a single-colorant halftone, the paper spectral reflectance between the dots is just a linear mixture of the base transmittance and the transmittance of the colorant weighted by the fractional area of the colorant.[13] Equation 5 describes the details. Here R_p is the paper reflectance, $t_b(l)$ is the base transmittance, $t_c(l)$ is the colorant spectral transmittance, and a is the fractional area covered by the colorant:

$$R_{\text{paper-dots}} = R_p(\lambda)[t_b(\lambda)(1-a) + at_c(\lambda)]t_b(\lambda). \quad (5)$$

This equation illustrates that the limiting spectral reflectance of the paper between the dots has, as its lower bound, a line starting at the paper spectral reflectance, for $a = 0$, and ending at the spectral reflectance given by the paper and colorant transmittance, for $a = 1.0$. The paper reflectance curve for any combination of halftone dot pattern and paper spread function can be found in the triangle bounded by the line, given by Eq. 5, and the constant R_p.

Dot reflectance also has a bounded region. It has been argued before that Eq. 3 also gives the limiting dot reflectance, as the fractional colorant area, a, approaches zero.[13] Thus the bounding relationship for the dot reflectance is given by Eq. 6.

$$R_{\text{dots}} = R_p(\lambda)t_c(\lambda)[t_b(\lambda)(1-a) + at_c(\lambda)]. \qquad (6)$$

The usual assumption is that the dot has constant reflectance independent of the fractional area of the halftone and this defines the lower bound of the dot reflectance region. The upper bound, as a function of fractional area of the colorant covered, is given by Eq. 6.

Experimental Test of Theory

Experimental verification of Eq. 4 was undertaken, using a wax thermal transfer printer. Measurements were made of the spectral reflectance of eleven patches of single-colorant (cyan, magenta, and yellow) constant fractional area. The patches included three cases of colorant coverage: no colorant, the paper, and complete or 100% coverage. From the 100% patch, the spectral transmittance was determined by solving Eq. 4 for $t(\lambda)$, using the measured paper and 100% area spectral reflectance data for cyan, magenta, and yellow. In this sense the spectral transmittance of the colorant is defined using Eqs. 2, 4, and actual measurements [note that $t_b(\lambda)$ for this case = 1.0; i. e., no base]. With the spectral transmittance for cyan, magenta, and yellow determined in this manner and using Eqs. 1, 2, and 4, the CIE tristimulus values and CIE L*a*b* coordinates can be calculated for arbitrary amounts of fractional area covered.[18] For
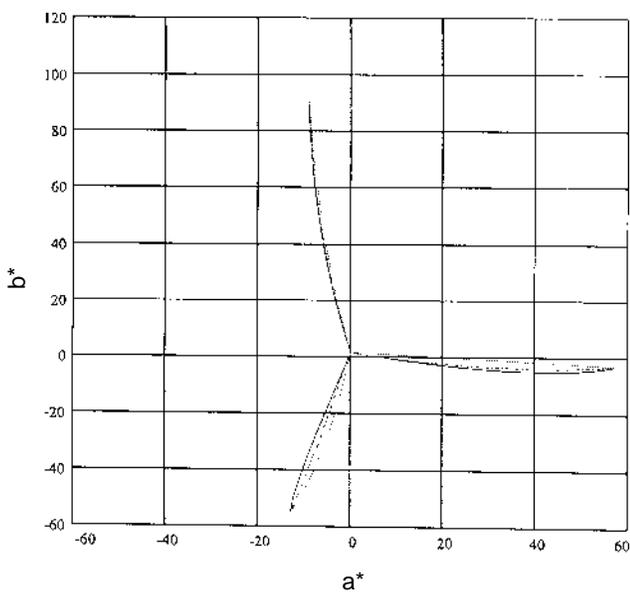

Figure 1. Cyan (lower left), magenta (right) and yellow (top) single-colorant loci. The solid lines are total scattering; dotted lines are nonscattering, and dash-dot lines are measured data.
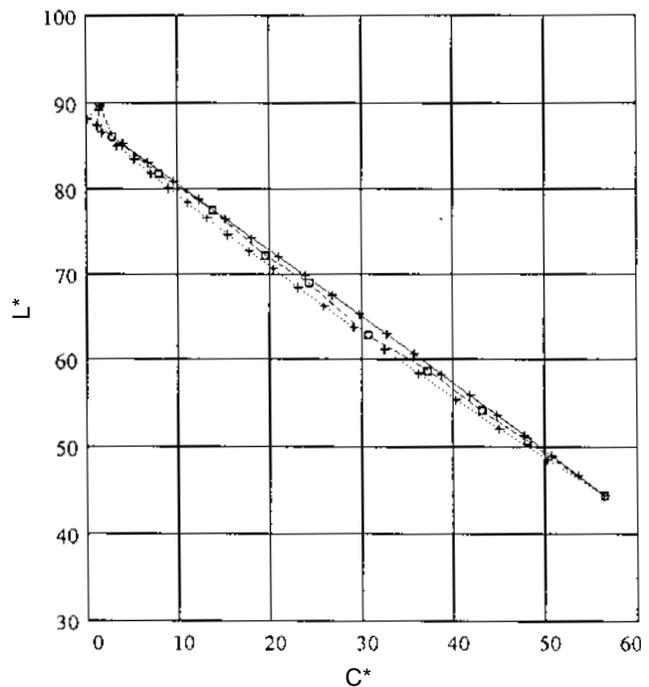

Figure 2. L*–C* diagrams for cyan colorant. The solid line with +'s is the total scattering limit. The dotted line with +'s is the nonscattering limit, and the dashed line with squares is measured data. The + marks are computed L*a*b* values using increments of 0.05 fractional area.

single-colorant fractional areas, we can plot two loci. Both loci are plotted as projections; one projection is in the a*-b* plane and the other is a projection in the lightness (L*)-metric chroma (C*) plane. One locus is for the no-scattering, or Neugebauer case, using measured spectral reflectances of the 100% coverage patches. The second locus is for complete light scattering, given by Eq. 4. These limiting loci can be compared with measured loci of single-color halftones. This locus approach was chosen to avoid the difficulty of estimating the fractional areas covered by the halftones. If the measured locus of a single-colorant lies between the two limiting loci, in the two CIE L*a*b* projections, we can reasonably conclude that Eqs. 1, 3, and 4 describe the limiting cases.

Results and Discussion

Wax thermal printer technology uses heat to melt a ribbon coated with a wax in which a colorant is dispersed. Within the heated area, all of the colorant layer is transferred to a paper receiver sheet, thus assuring a constant colorant layer thickness. The wax thermal printer results are shown in Figs. 1 through 4. The a*–b* plots are the projections of the loci onto the a*–b* plane (Fig. 1), and the L*–C* plots are projections of the loci onto an approximately constant hue plane, (Figs. 2 through 4). In the CIE L*a*b* system, C* is the distance from the L* axis, termed metric chroma.

For two of the three single colorants, the measured loci shown in Figs. 1 through 4 are found to lie within the two boundary loci. An exception was the L*–C* rep-
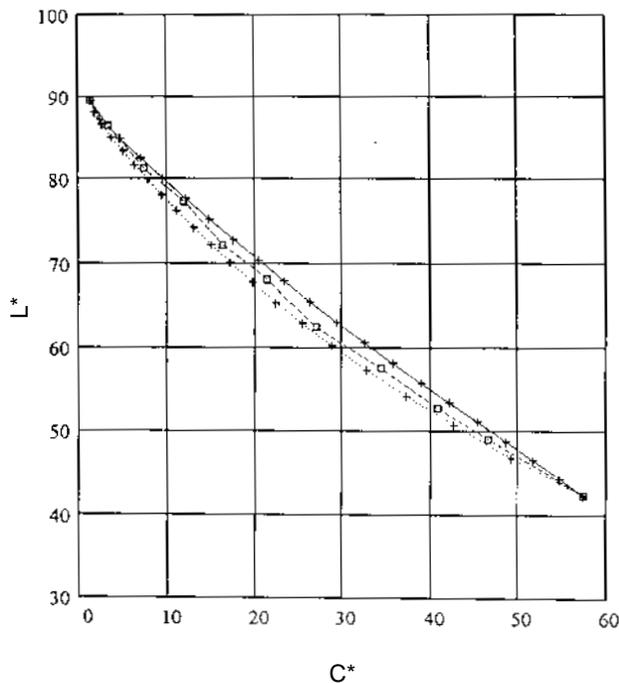
resentation for the yellow colorant (Fig. 4), where the space between the loci is so small that an L* variation of less than about 0.25 causes points to fall outside the boundaries. The a*–b* projection for the yellow colorant fits between the two limiting loci.

Figures 2 through 4 illustrate that for all single-colorant halftones, the complete scattering limit has higher L* for a given C* than with the Neugebauer equations (no scattering). This confirms the observations made more than 40 years ago by Yule and Nielsen;[7] papers with large spread functions gave lighter colors. Additional model calculations lead us to expect that this observation also holds for dark colors.

The crosses on the loci in Figs. 2 through 4 represent the computed L*a*b* values for every 5% fractional dot area from 0 to 100%. Comparison of the points on the two loci shows that the distributions are markedly different. The no-scattering condition results in small differences between the colors at the highlight (lighter) end of the scale compared with the dark, or higher chroma, end of the scale. With complete scattering the colors are more uniformly spaced along the loci.

Points are defined along these loci according to the amount of fractional area covered by the colorant. Rear-
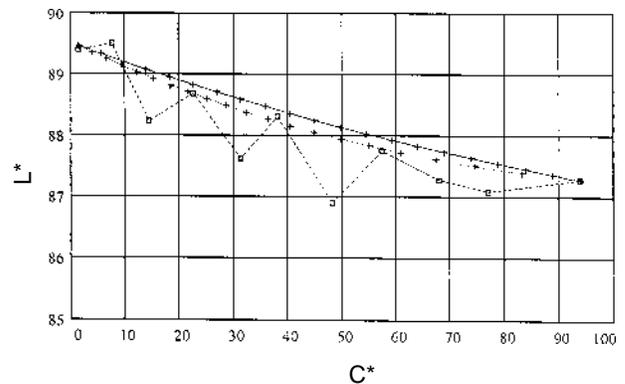
Figure 4. L*–C* diagrams for yellow colorant. The solid line with +'s is the total scattering limit; the dotted line with +'s is the nonscattering limit, and the dashed line with squares is measured data. The + marks are computed L*a*b* values using increments of 0.05 fractional area.

Figure 3. L*–C* diagrams for magenta colorant. The solid line with +'s is the total scattering limit.; The dotted line with +'s is the nonscattering limit; and the dashed line with squares is measured data. The + marks are computed L*a*b* values using increments of 0.05 fractional area.
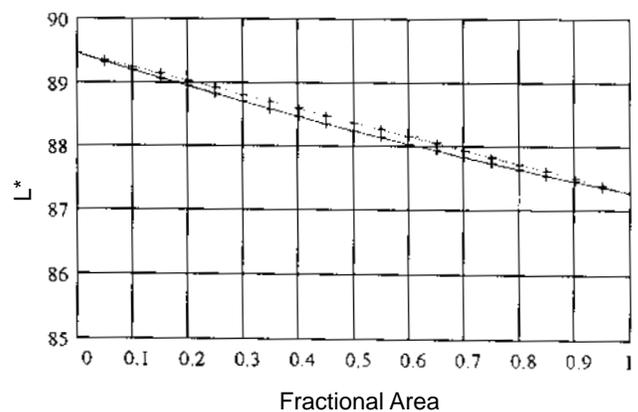
Figure 5. L* versus fractional area covered for yellow colorant. Dotted curve is without scattering and the solid curve is with complete scattering.

rangement of the areas via an LUT in the printer processing stream can alter the spacing of the points with respect to some input. However, physically, any given fractional area printed on a specific paper will yield a specific color, and no LUT can change this. Here is where the paper spread function and the specifics of the halftone pattern interact. A good strategy would be to space the colors uniformly. To achieve uniform spacing, these results suggest that large paper spread functions (high scattering) are preferred. This seems to be against the current practice of using specially coated paper or plastic substrates, which often have relatively narrow spread functions.

Maximum distance (color difference) between colors generated by *the same amount of colorant (fractional area),* under the two diffusion conditions, can be as high as 20 L*a*b* units. The loci are in fact closer in space than this number suggests; the maximum is actually around 3 to 5 unit color differences. By adjustment of the printed areas in the halftone pattern, the large color difference can be reduced to the maximum distance between the loci. This is the primary colorimetric function of so-called "dot gain curves." However, these results suggest that there is a limit to the reduction in the color difference using this dot-gain or 1-D LUT color correction strategy, that limit being the distance between the curves.

Figures 5 through 7 show calculated L* versus fractional area covered for the two cases. Note the difference in the L* for the complete-scattering and no-scattering conditions. For the complete-scattering case the relationship between L* and fractional area is almost linear for all three colorants. We have often observed this relationship for a variety of printers, both monochrome and color. It now appears that this is a consequence of the ratio of paper spread function to halftone cell size, or imaging element size, being such that nearly complete scattering is observed. The no-scattering case produces lighter images for the same amount of colorant.

The curves of Figs. 5 through 7 also suggest the magnitude of dot area changes needed for constant L*, the so-called dot gain curves. The dot area is usually assumed
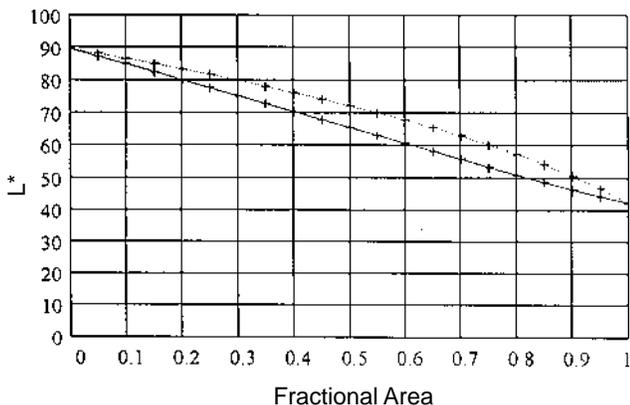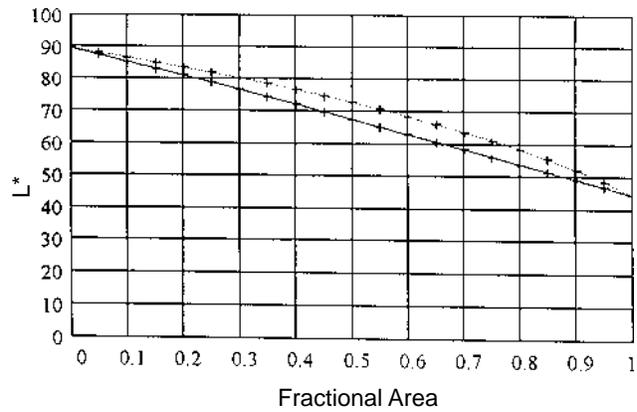

Figure 7. L versus fractional area covered for cyan colorant. Dotted curve is no scattering and the solid curve is with complete scattering.*

to be equal to the nonscattering curve in these figures, but the measurements are made from scattering images that follow the complete scattering curve. Because there is a decrease in L* for the same colorant, it is *assumed* to come from an increase in fractional dot area. Hence a "gain" in the dot area. An upper bound to the dot gain curves, at least on a lightness basis, can be computed directly from these curves by computing the difference in fractional dot area for constant L* between the nonscattering and complete scattering cases.

The difference in the L*a*b* loci implies that there is a difference in the color gamut. Using the Neugebauer equations as a model, which is equivalent to assuming no scattering, can result in "out of gamut" colors. This can occur when one inverts the equations to determine the amounts of *c*, *m*, and *y*, colorants needed for a colorimetric match. This problem is due to the paper scattering, which alters the dot and paper-between-the-dot reflectance that is not considered by the equations.

Conclusions

A proposed spectral model can be used in the determination of the colorimetric limits of Neugebauer halftone imaging for two cases. The first case, which follows the well-known Neugebauer equations, is for the no-paper-scattering case. Complete scattering defines the second case.

Single-colorant halftone patches from a wax thermal transfer printer showed that these theoretical limits bound the measured data.

The biggest single effect of scattering is to change the color for a fixed amount (fractional area) of printed colorant. Color differences as high as 20 CIE L*a*b* units have been calculated between single-colorant colors printed with the same colorant amounts under the two scattering cases. However, actual maximum distances of the colorimetric loci are only about 3 to 5 CIE L*a*b* units apart. The biggest effect of scattering is to alter the position of colors along the colorant locus. For no scattering, the distances between colors for equal colorant amounts at the low-colorant end of the locus

Figure 6. L versus fractional area covered for magenta colorant. Dotted curve is no scattering and the solid curve is with complete scattering.*

are closer together. At the higher colorant end of the locus the distances become greater. Complete scattering tends to even out the distance between the equal-colorant increment points on the locus. Dot gain corrections and tone reproduction LUTs in color management systems are used to adjust the points on the locus relative to some input value, but these simple corrections cannot change the locus that is defined by a paper spread function and printing conditions to a locus defined by another set of conditions. Multiple-dimension LUTs must be used for this purpose.

Also, the color quantization inherent in the small halftone cell printing is greater at high colorant levels with no paper scattering. In other words, there are greater distances along the colorant loci, at high colorant levels, for contrast dot area differences. This suggests some advantage for printing on noncoated paper or substrates with large paper spread function (scattering). One possible penalty with this strategy is a loss of image sharpness.

Appendix

This appendix illustrates how to derive the Neugebauer equations from set and probabilistic concepts.

Figure Al shows a Venn diagram of the set of all the possible "events" when overprinting three colors, Ca = cyan, Ma = magenta, and Ya = yellow. The events are W = white paper, K = black, R = red, G = green, B = blue, C = cyan, M = magenta, and Y = yellow, eight in total.

The approach to the derivation is to determine the probabilities of events (colors) W, C, M, Y, R, G, B, and K in terms of the colorant amounts Ca, Ma, and Ya. We assume all events to be independent. We therefore have the following probabilities:

$$p[R] = p[Ma \ll Ya] - p[Ca \ll Ma \ll Ya]$$
$$= MaYa - CaMaYa = MaYa(1 - Ca), \qquad (A1)$$

$$p[G] = p[Ca \ll Ya] - p[Ca \ll Ma \ll Ya]$$
$$= CaYa - CaMaYa = CaYa(1 - Ma) \qquad (A2)$$

$$p[B] = p[Ca \ll Ma] - p[Ca \ll Ma \ll Ya]$$
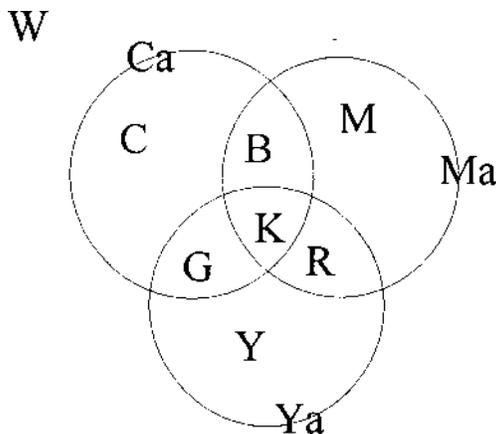$$= CaMa - CaMaYa = CaMa(1 - Ya), \qquad (A3)$$

Figure Al. Venn diagram of cyan, magenta, and yellow overprints and the "events" (colors) they generate.

$$p[K] = p[Ca \ll Ma \ll Ya] = CaMaYa, \qquad (A4)$$

$$p[C] = p[Ca] - p[B] - p[G] - p[K]$$
$$= Ca - CaMa(1 - Ya) - CaYa(1 - Ma) - CaMaYa$$
$$= Ca(1 - Ma)(1 - Ya), \qquad (A5)$$

$$p[M] = p[Ma] - p[B] - p[R] - p[K]$$
$$= Ma - CaMa(1 - Ya) - MaYa(1 - Ca) - CaMaYa$$
$$= Ma(1 - Ya)(1 - Ya), \qquad (A6)$$

$$p[Y] = p[Ya] - p[G] - p[R] - p[K]$$
$$= Ya - CaYa(1 - Ma) - MaYa(1 - Ca) - CaMaYa =$$
$$Ya(1 - Ca)(1 - Ma), \qquad (A7)$$

$$p[W] = p[Ca' \ll Ma' \ll Ya']$$
$$= (1 - Ca)(1 - Ma)(1 - Ya). \qquad (A8)$$

To compute the tristimulus value of the color, we take the expected value, or average, over all the colors. From elementary statistics we know that the expected tristim-ulus value, T_{color}, is just the tristimulus value of the events (colors) times the probability of occurrence. Thus

$$T_{color} = T_w p[W] + T_c p[C] + T_M p[M]$$
$$+ T_Y p[Y] + T_R p[R] + T_G p[G] + T_B p[B] + T_K p[K]. \quad (A9)$$

Equation A9, when combined with the probabilities given by Eqs. A1 through A8, yields the familiar Neugebauer equations. Therefore, the Neugebauer equations imply that all the colors in a halftone image are independent of each other. No other assumptions were made in this derivation.

References

1. H. E. J. Neugebauer, *Zeitschr. Wiss Photogr.* **36**: 75 (1937).
2. I. Pobboravsky, *TAGA Proc.* p. 10 (1964).
3. I. Pobboravsky and M. Pearson, *TAGA Proc.* p. 65 (1972).
4. J. A. S. Viggiano, *TAGA Proc.* p. 647 (1985).
5. R. Rolleston and B. Balasubramanian, *Proceedings of 1st IS&T/SID Color Imaging Conference,* p. 32 (1993).
6. H. R. Kang, *Proceedings of 1st IS&T/SID Color Imaging Conference,* p. 78 (1993).
7. J. A. C. Yule and W. J. Neilsen, *TAGA Proc.* p. 65 (1951).
8. J. A. S. Viggiano, *TAGA Proc.* p. 44 (1990).
9. Y. Liu, *TAGA Proc.* p. 154 (1991).
10. S. Yoshinobu and T. Mizuno, *J. Imaging Sci.Technol.* **37**: 385 (1993).
11. D. R. Lehmbeck, in *IS&T 28th Annual Conference and Seminar on Quality Control,* p. 155, IS&T (May 1975).
12. F. Ruckdeschel and O. G. Hauser, *Appl. Opt.* **17**: 3376 (1978).
13. P. G. Engeldrum, *J. Imaging Sci. Technol.* **38**: 545 (1994); *see pg. 383, this publication*.
14. J. S. Arney, P. G. Engeldrum, and H. Zeng, *TAGA Proc.* p. 353 (1995).
15. J. S. Arney and P. G. Engeldrum, *Proceedings of IS&T's 11th International Congress on Advances in Non-Impact Printing Technologies*, p. 497 (1995); *see pg. 397, this publication*.
16. M. Pearson, *TAGA Proc.* p. 415 (1980).
17. P. G. Engeldrum, *J. Imaging Technol.* **12**: 126 (1986).
18. CIE Publication 15.2 and ASTM 308E-1987.