

Color Gamut of Halftone Reproduction*

Stefan Gustavson^{†‡}

Department of Electrical Engineering, Linköping University, S-581 83
Linköping, Sweden

Abstract

Color mixing by a halftoning process, as used for color reproduction in graphic arts and most forms of digital hardcopy, is neither additive nor subtractive. Halftone color reproduction with a given set of primary colors is heavily influenced not only by the colorimetric properties of the full-tone primaries, but also by effects such as optical and physical dot gain and the halftone geometry. We demonstrate that such effects not only distort the transfer characteristics of the process, but also have an impact on the size of the color gamut. In particular, a large dot gain, which is commonly regarded as an unwanted distortion, expands the color gamut quite considerably. We also present an image processing model that can describe and quantify the effects of physical and optical dot gain on different media and with different halftoning methods.

Introduction

Any halftone reproduction is subject to dot gain. The effect takes its name from the fact that the halftone dots end up larger in the reproduction than in the original if no compensation is performed.

Physical dot gain occurs because the ink that forms the halftone dot usually does not stay entirely within its bounds, but spreads somewhat physically on the printed surface. The printed halftone dots thus get somewhat larger than intended, and the resulting image appears darker.

Physical dot gain is illustrated in Fig. 1. From Fig. 1(a) to 1(d), we see the effect of an increasing physical dot gain on two small halftone dots.

Optical dot gain, or, as it is often called, the “Yule-Nielsen effect”,¹ is slightly harder to explain. It occurs because most printing substrates used for halftone imaging are translucent and scattering. Examples of such substrates are paper, plastic, and cloth. Metal surfaces are not subject to optical dot gain, but such surfaces are rarely used for imaging.

Figure 2 shows the basic principle behind optical dot gain. Light enters from above and passes through the ink layer. Light is absorbed where there is ink, and passes straight through where there is no ink. The printed pattern casts a shadow of itself on the surface of the scattering substrate. Due to multiple scattering within the substrate, incident light is not reflected only from the point of incidence, but from a small region around that point. Thus, the reflected image is a diffused version of the incident pattern of light. This diffused reflected pat-

tern then gets attenuated once more by the pattern of ink that resides on the surface, and the finally reflected light is the result of these three effects combined: transmission through the ink film, diffused reflection from the substrate, and transmission through the ink film again. The left-hand side of Fig. 2 shows an exploded view of the ink layer and the substrate, with the diffused reflected pattern shown on the substrate. The final viewed image is a view from the top of these two layers, as shown to the right in Fig. 2. The dots do not really increase in size, but they have a shadow around the edge that makes them appear larger, and the image is darker than what would have been the case without optical dot gain.

Model

We have developed a simple but effective model for both physical and optical dot gain, which has proven to be successful both in terms of explanation and prediction

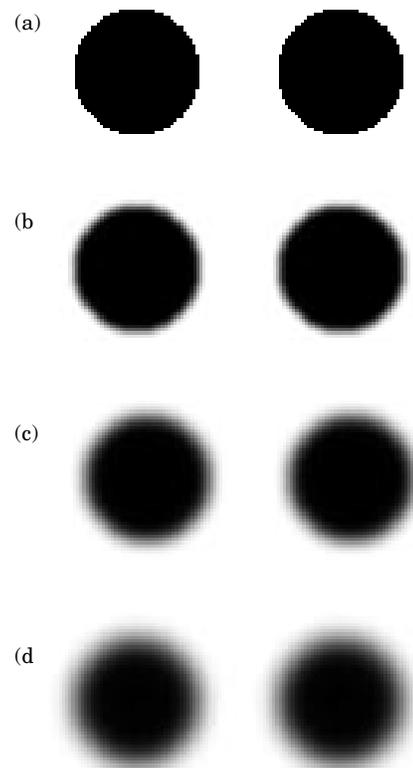


Figure 1. The effect of increasing physical dot gain.

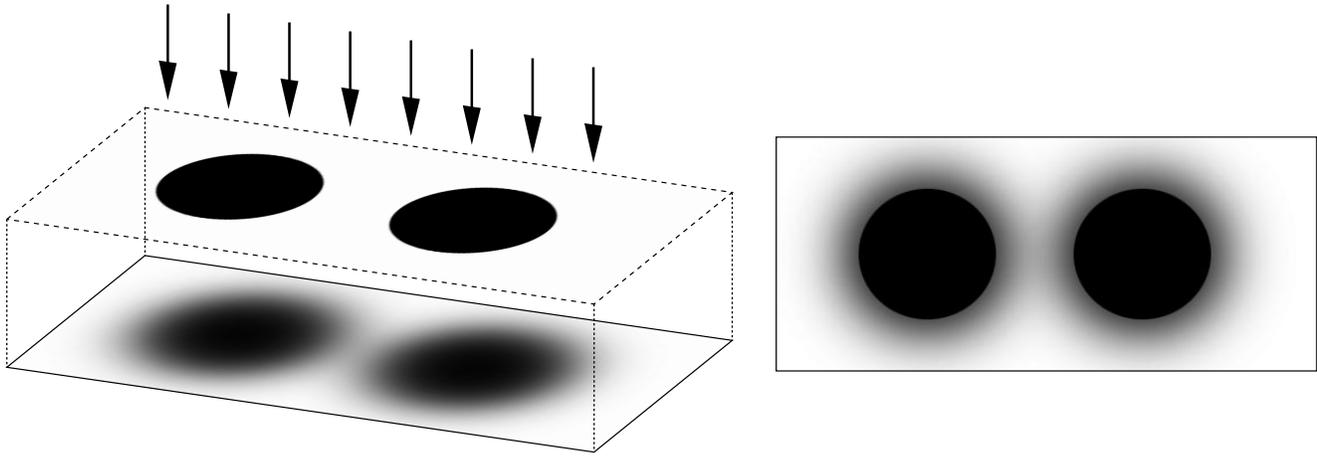


Figure 2. Optical dot gain. Principle to the left, appearance to the right.

of dot gain effects. The model has a direct relation to the physics and optics of the actual imaging system.

We assume that the print substrate (most often paper) is flat, smooth, and reasonably uniform, that the ink is placed in a thin layer entirely on top of the substrate, and that the ink is properly characterized by its absorption properties only. The pattern of ink on the surface can then be described by an absorption function, or, more conveniently, a transmission function $T(x, y)$, defined over the surface coordinates x, y . This transmission function can only take on values between 0 and 1, inclusive, where 0 means total absorption and 1 means unhindered transmission. The transmission is in turn dependent on the thickness, or density, of the ink layer. More specifically, the transmission is an exponential function of the ink density.

Instead of assuming perfect halftone dots with sharp edges, we model a smearing of the ink by first calculating a perfectly sharp simulated halftone image $H(x, y)$, which takes on the values 0 or 1 only. To this image we apply a linear blurring (low-pass) filter $B(x, y)$. If the blurring filter kernel is properly normalized, this operation does not change the total amount of ink on the surface, but merely redistributes it by smearing out sharp edges. After the smearing, we exponentiate the result to get our transmission image $T(x, y)$ according to Eq. 1.

$$T(x, y) = 10^{-D_{\max}(H(x, y) * B(x, y))}. \quad (1)$$

In this equation, D_{\max} is the full-tone transmissive density, not the reflective density.

For color halftone images, a dependence of wavelength needs to be incorporated, and we also need to calculate one transmission image for each primary ink, but the basic model remains the same (2):

$$\begin{aligned} T_C(x, y, \lambda) &= 10^{-D_C(\lambda)[H_C(x, y) * B(x, y)]} \\ T_M(x, y, \lambda) &= 10^{-D_M(\lambda)[H_M(x, y) * B(x, y)]} \\ T_Y(x, y, \lambda) &= 10^{-D_Y(\lambda)[H_Y(x, y) * B(x, y)]} \\ T_K(x, y, \lambda) &= 10^{-D_K(\lambda)[H_K(x, y) * B(x, y)]}. \end{aligned} \quad (2)$$

The final transmission function is then calculated as the product of the individual transmission images (3):

$$T(x, y, \lambda) = T_C(x, y, \lambda)T_M(x, y, \lambda)T_Y(x, y, \lambda)T_K(x, y, \lambda). \quad (3)$$

For the optical dot gain, we need to model the lateral diffusion of light due to multiple scattering before reflection. This can be described by a convolution operation with a point spread function (PSF) for diffuse reflection, $P(x, y, l)$. The total integral of this PSF is the diffuse reflectance R_0 of the print substrate, and the spatial extent of the PSF describes the amount of lateral spreading of light. Light that enters through the ink layer is first attenuated according to a point-wise multiplication with the transmission function $T(x, y, l)$ and then diffused by multiple scattering. The reflected light then has to pass once more through the ink layer to reach the viewer. This corresponds to a final point-wise multiplication with the transmission function. If we denote the incident light intensity with $I(l)$, the reflected image $R(x, y, l)$ is thus described by Eq. 4:

$$R(x, y, \lambda) = I(\lambda)[T(x, y, \lambda) * P(x, y, \lambda)]T(x, y, \lambda). \quad (4)$$

This is a nonlinear model for optical dot gain expressed in image processing terms. It allows for direct simulation of the reflected image from an arbitrary halftone pattern, provided that the PSF is known. We have shown that it is possible to calculate the PSF by direct simulation of the multiple scattering optical system of a model paper sheet using modern computers. A typical simulated PSF for diffuse incident light is closely approximated by a simple exponential function (5):

$$P(x, y, \lambda) \approx R_0(\lambda) \frac{a(\lambda)}{2\pi r} e^{-a(\lambda)r}. \quad (5)$$

Admittedly, the function in (5) has a singularity at the origin, but it is in effect a probability distribution function, and as such it is valid, since the integral over any finite area is well defined.

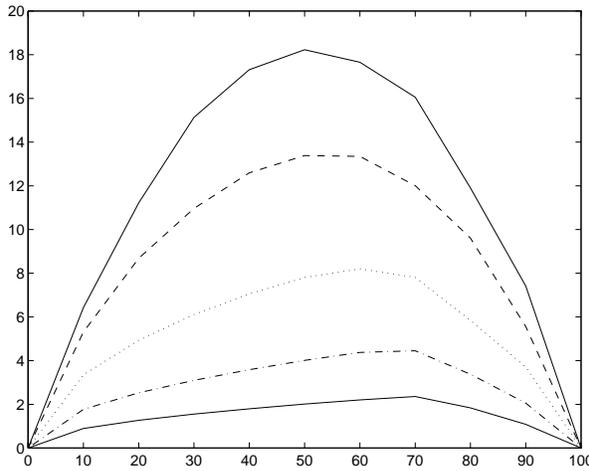


Figure 3. Dot gain increase from increasing half-tone frequency. From bottom to top, the simulated screen rulings increase by a factor of 2 for every curve. The top one corresponds to a very fine screen ruling or a stochastic screen, and the bottom one to a coarse conventional screen.

The shape of the PSF, and of course therefore also the parameters R_0 and a in the approximation above, depend in a nontrivial way on the scattering and absorption cross sections and the thickness of the substrate. The cross sections are in turn related to the K and S parameters of the famous Kubelka-Munk theory,² although not in such a simple way as one would have hoped. For more information on our light diffusion model and the relation to the approximative Kubelka-Munk theory, we refer to previously published work.³⁻⁵

That concludes the summary of our dot gain model. Note that the equations contain no noise model. For the moment, we model the deterministic average distortion only. However, incorporating a noise model is perfectly possible and indeed very close at hand, and this is an extension we are planning to do. The problem is that the properties of the noise are not very well known right now, but we are looking into it.

Monochrome Dot Gain

Dot gain in monochrome prints manifests itself as a nonlinear transfer function. If the reflectance of the substrate is R_0 and the reflectance of the full-tone ink film is R_S , a 50% dot coverage in the halftone pattern yields a reflectance that is less than $(R_0 + R_S)/2$. The usual way of expressing the dot gain is by measuring the reflectance R_A for a certain area coverage A and calculate the apparent area coverage A' according to Eq. 6:

$$A' = \frac{R_0 - R_A}{R_0 - R_S}. \quad (6)$$

The dot gain G is then defined as the difference $G = A' - A$. Both A and A' , and therefore also G , are mostly expressed in percent. Our model readily explains the commonly known effect that an increased half-tone frequency increases the dot gain. This is seen in Fig. 3, where the simulated dot gain is plotted against the area

coverage for a number of different halftone frequencies. What is less obvious is that the halftone geometry also plays an important role in this respect. A few plots of our model's simulations of physical and optical dot gain due to different halftone geometries are shown in Figs. 4 and 5. Note that the physical and optical dot gain have very similar characteristics, although they occur for entirely different reasons.

Color Dot Gain

Dot gain in color prints is, perhaps somewhat unexpectedly, a far more complicated issue than monochrome dot gain, but our model as we have described it previously handles it straight off.

The final spectral distribution $R(x, y, l)$ according to Eq. 4 is evaluated by the human eye according to the tristimulus principle, and the perception of color may be presented in a three-dimensional space, for example the CIE $L^*a^*b^*$ color space. For a monochrome print, the range of reproducible colors, the color gamut, is just a line from black to white, and the signal is a one-dimensional property, the reflectance. Any distortion shifts the signal along one dimension only and does not change the color gamut. For color prints, the impact on the signal from the dot gain is more complicated. The color gamut is now a three-dimensional volume in the color space, and distortions due to dot gain that occur in the spectral domain can shift the signal along any direction, and actually change the extent of this color gamut. For the simulations below, we used spectrophotometric reflectance measurements for the CMYK primary colors used in four-color offset printing as a starting point, and made the (admittedly somewhat dubious) assumptions that the four ink films placed on top of each other are placed mutually independent of one another, and that the ink is nonscattering. Thus, the ink films are characterized by their absorption properties only, and a combination of several colors on top of each other can be modeled by a multiplication of the transmission through each separate ink film.

It should be noted that the validity of the assumptions before are by no means crucial to our qualitative results, namely that the extent of the color gamut does change under dot gain. However, to perform a proper quantitative analysis, a more thorough investigation based on more data is required.

The maximum transmissive densities of each of the the primary colors were calculated from the spectral reflectances of the full-tone primary colors, as presented in Eq. 7:

$$\begin{aligned} D_C(\lambda) &= -\log \sqrt{\frac{R_C(\lambda)}{R_0(\lambda)}} \\ D_M(\lambda) &= -\log \sqrt{\frac{R_M(\lambda)}{R_0(\lambda)}} \\ D_Y(\lambda) &= -\log \sqrt{\frac{R_Y(\lambda)}{R_0(\lambda)}} \\ D_K(\lambda) &= -\log \sqrt{\frac{R_K(\lambda)}{R_0(\lambda)}} \end{aligned} \quad (7)$$

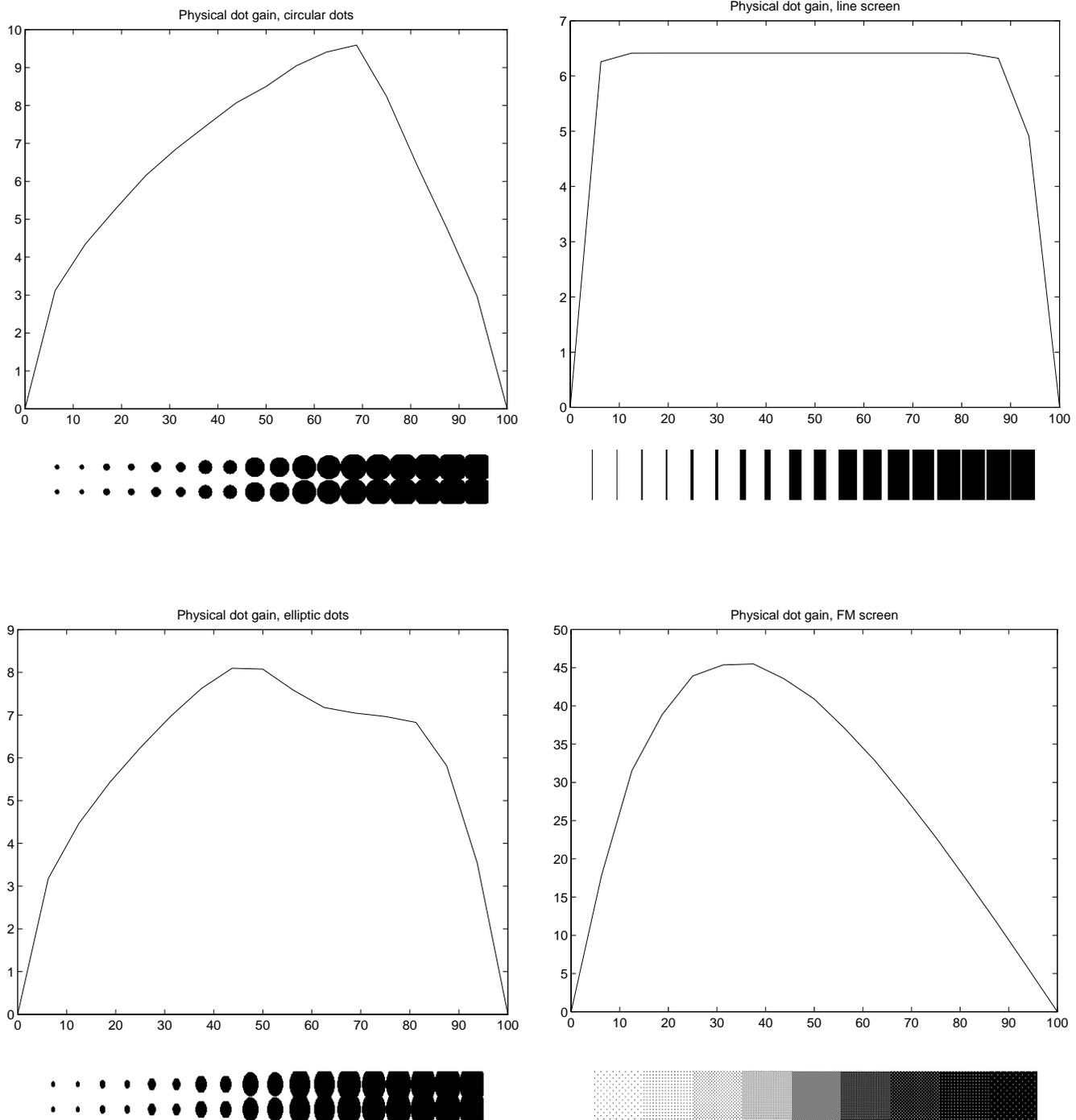


Figure 4. Physical dot gain for different halftone geometries.

The corresponding transmission properties of the full-tone primary colors are shown in Fig. 6.

Implications for the Color Gamut

Using the spectral measurements described previously, the effects of physical and optical dot gain were evaluated with our model. Quite surprisingly, the dot gain,

which has been regarded as an unwanted distortion, actually increases the size of the color gamut, and quite considerably so. Without any dot gain, the color gamut of a standard halftone process is that of Fig. 7. The extreme cases of dot gain occur when the extents of the spread functions are large compared to the details of the halftone pattern. These extreme cases are shown in Figs. 8 and 9. Figure 8 shows the color gamut under extreme

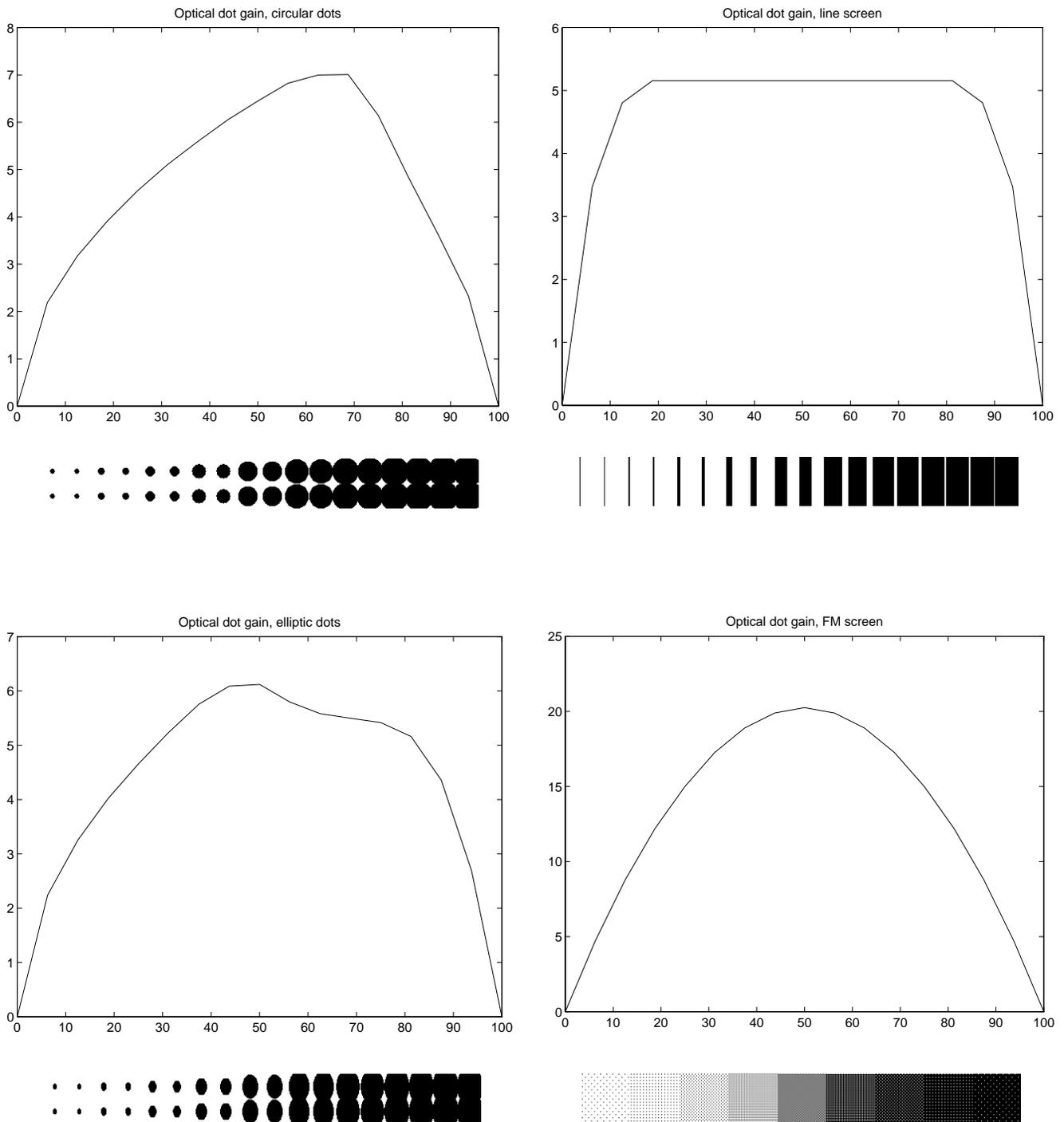


Figure 5. Optical dot gain for different halftone geometries.

optical dot gain only, and Fig. 9 shows the color gamut under extreme physical dot gain (in this case, optical dot gain has no further effect). It is clearly seen that both types of dot gain increase the available range of colors for light tones (the top half of the gamut, between white and the chromatic colors). A region where this effect is prominent is the cyan plus yellow (CY) colors, the light green part of the gamut, for which we show a two-di-

mensional chromaticity plot in the $L^*C_{ab}^*$ plane in Fig 10. On a very close look, the color gamut shrinks somewhat in dark tones (the bottom half of the gamut, between the chromatic colors and black) under the influence of dot gain, but the effect is very slight in comparison to the marked expansion in light tones.

The extreme cases of dot gain demonstrated here are in fact never observed in reality, but for very fine con-

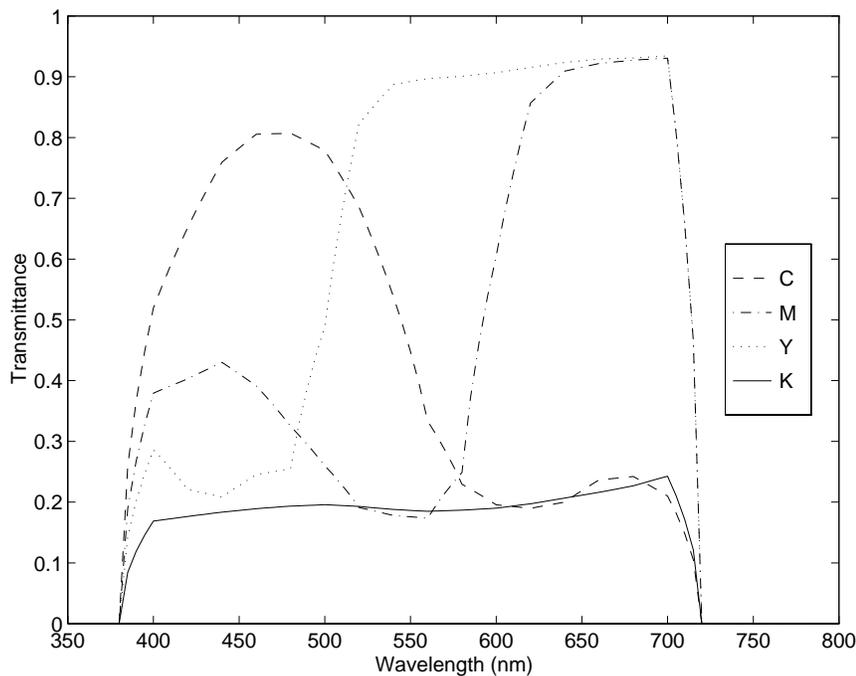


Figure 6. Measured spectral transmittance for CMYK inks. The spectrophotometer used measured only the range 400–700 nm, and outside of that range the reflectance curves were interpolated to zero for the CIE $L^*a^*b^*$ conversion. This is not correct, but the influence on the actual $L^*a^*b^*$ values were insignificant.

ventional halftone screens and for FM screens or stochastic screens, the details of the halftone pattern are on the same order of magnitude as the extent of the reflectance point spread function of paper, and the optical dot gain is actually quite close to the maximum. For some print methods where the halftone dot is not sharply defined, the physical dot gain can also be very large, even though the extreme case shown here is actually equal to a continuous-tone reproduction following Beer's law, where all colorants are spread out uniformly over the surface. In any case, the impact of dot gain is significant for most halftone reproduction processes, especially with the advent of new screening methods and the ever increasing resolution for digital printing devices.

Since the spectral absorption curves of the primary colors show a great deal of overlap, it is to be expected that it matters whether dots are placed on top of each other or next to each other. The normal way of making halftone separations is to make the amount of dot overlap more or less random, either by printing conventional halftones in different angles, or by making uncorrelated stochastic halftone separations. This is the assumption that underlies the color reproduction model of Neugebauer⁶ and its modern successors, and a random dot overlap is often the best thing to strive for when dealing with conventional printing processes, as there is often a considerable amount of inaccuracy in the registration between the different colors. However, if we have detailed control over the dot placement, which is often the case with digital color printers, it is possible to increase the color gamut and reduce the color noise or "graininess" of the image by correlating the halftone patterns of the primary colors. Uncorrelated halftones, where the dots of each primary color are placed in a random fashion, yields a color gamut like the one previously shown in Fig. 7. A positively correlated placement, where dots are placed on top of dots of other colors as much as possible, yields a color gamut as shown in Fig.

11. This type of halftone pattern is often called dot-on-dot. On the other hand, a negatively correlated dot placement, dot-off-dot, where dot overlap is avoided as much as possible, yields a color gamut according to Fig. 12. The mesh on the lower half of the gamut plot in Fig. 11 and 12 is a bit distorted due to a very different impact from the black ink for the two cases, but it is in the upper part that the gamuts actually differ to any significant amount. The dot-on-dot halftoning yields desaturated light tones in the secondary colors, where the dot-off-dot halftoning gives a lot more saturation. This is most clearly observed in the light blue to purple region to the right in Figs. 11 and 12, where the gamut in Fig. 11 has an indentation, but the gamut in Fig. 12 comes out to a cusp. A chromaticity plot in the $L^*C_{ab}^*$ plane for these cyan plus magenta (CM) colors is shown in Fig. 13.

It is clear that, at least for the spectral characteristics of the offset inks used in this example, quite a lot can be gained in terms of color gamut by printing halftone patterns in a dot-off-dot fashion instead of with a random placement. Again, we see that the effect is most pronounced in light tones, which is rather self-evident, since it is only in lighter tones that we actually have any real choice of whether to print the dots with or without overlap.

It should be noted that the difference between dot-on-dot and dot-off-dot halftoning decreases with increasing dot gain, since the placement of the dots becomes less well defined with increasing dot gain. A bad registration on a large scale, or a bad dot placement accuracy on a small scale, can also take out the difference.

Conclusion

The color gamut of a halftone reproduction is far from trivial to investigate. It is dependent on many factors. Apart from the obvious dependence of the spectral characteristics of the primary colors used, there is a consid-

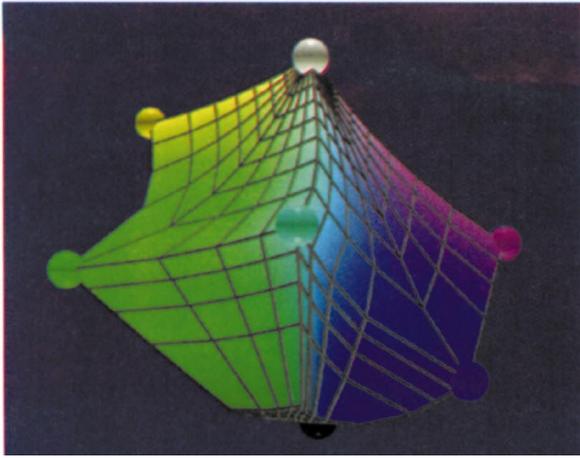


Figure 7. Color gamut, no dot gain.

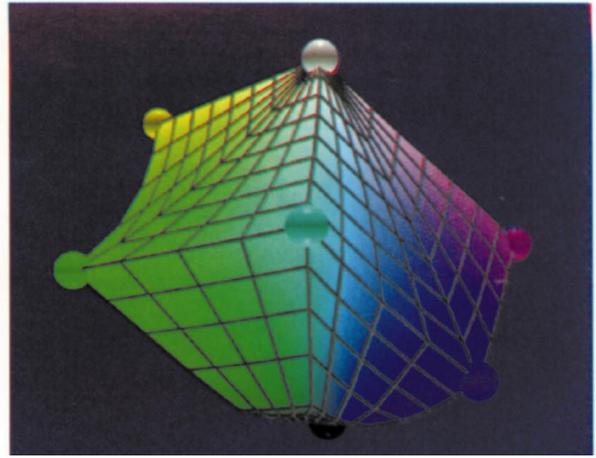


Figure 8. Color gamut, maximum optical dot gain.

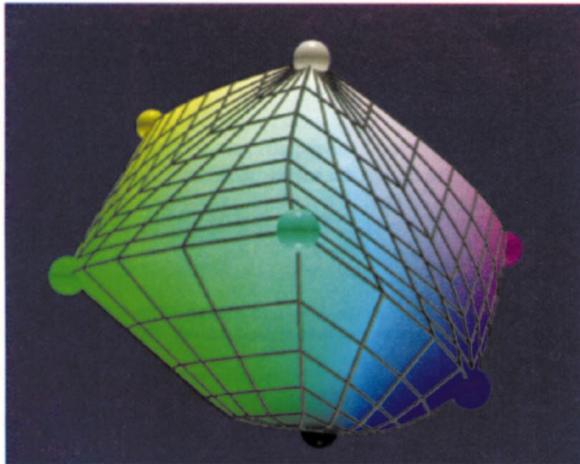


Figure 9. Color gamut, maximum physical dot gain.

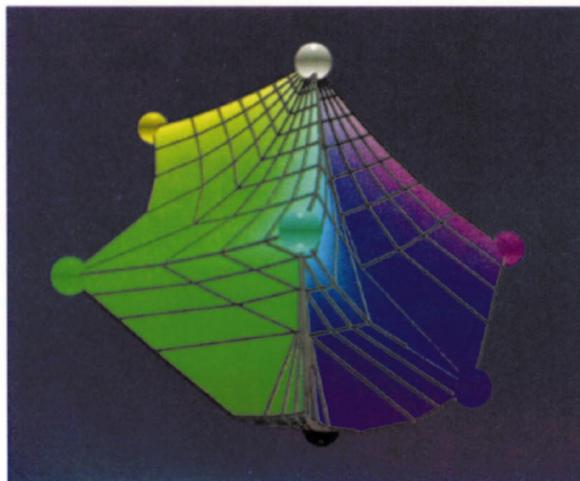


Figure 10. Color gamut, dot-on-dot halftone geometry.

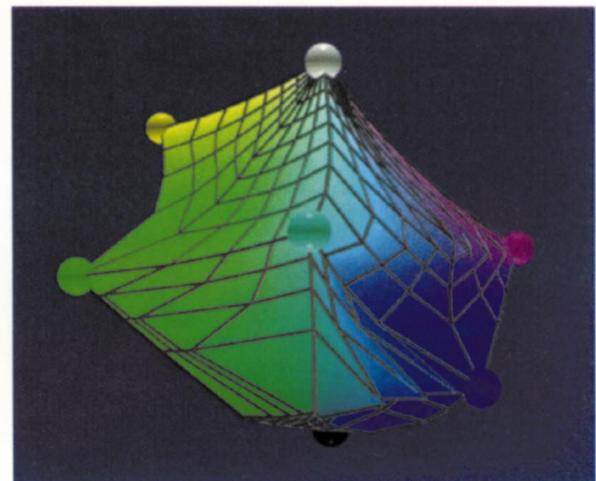


Figure 11. Color gamut, dot-off-dot halftone geometry.

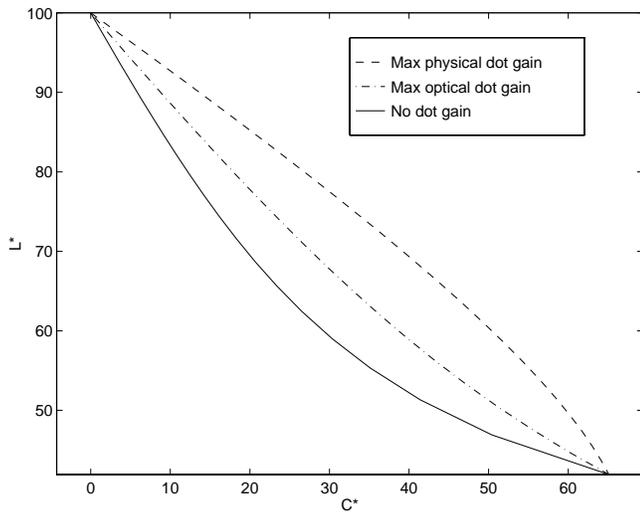


Figure 12. L^*C^* chromaticity plot for green colors ($C+Y$) for, from bottom to top, no dot gain, maximum optical dot gain, and maximum physical dot gain.

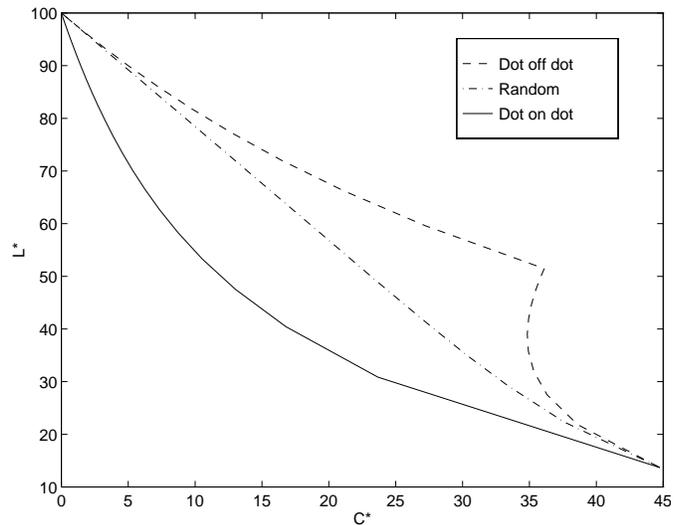


Figure 13. L^*C^* chromaticity plot for blue ($C+M$) colors for dot-on-dot halftoning (bottom), dot-off-dot halftoning (top), and random dot placement (middle).

erable influence from physical and optical dot gain, which both depend heavily on the halftone geometry, and also quite some influence from the exact placement of the halftone dots. Specifically, a large dot gain increases the range of reproducible colors in light tones, and a correlated placement of the halftones for the primary colors in a dot-off-dot fashion also has that effect. The model we have presented is capable of predicting the effect of the various parameters of a halftone reproduction process, and can be used to gain some further insight into how to make good halftones for color reproduction.

Acknowledgments

The work presented here has been performed within the Swedish national print research program, PFT, with funding from the Nordic Industrial Fund. The financial support and the fruitful collaboration within PFT is gratefully acknowledged. The raytraced images of color gamuts in this paper were rendered with the Blue Moon Rendering Tool (BMRT) written by Larry Gritz.

References

1. J. A. C. Yule, D. J. Howe J. H. and Altman, "The Effect of the spread-function of paper on halftone reproduction", *TAPPI J.* **50**, (7) 337–344 (1967).
 2. P. Kubelka and F. Munk, "Ein Beitrag zur Optik der Farbanstriche," *Zeitschrift für technische Physik* **11**(a), 593–601 (1931).
 3. S. Gustavson and B. Kruse, "3D modelling of light diffusion in paper," *TAGA Proc.* **47**, (2) 848–855 (1995).
 4. S. Gustavson and B. Kruse, "Evaluation of a light diffusion model for dot gain," *TAGA Proc.* **48** (1996).
 5. S. Gustavson, "Modelling of light scattering effects in print," *Thesis LIU-TEK-LIC-1995:52, Linköping Studies in Science and Technology*, 1995. ISBN: 91-7871-623-3. URL: <http://www.isy.liu.se/~stefang/lic.html>.
 6. H. E. J. Neugebauer, "Die theoretischen Grundlagen des Mehrfarbenbuchsdrucks", *Z. wiss. Phot.* **36** (4), (April 1937).
- * Previously published in the *Journal of Imaging Science and Technology*, **41**(3) pp. 283–290, 1997.