

# Printer Models and Error Diffusion

Thrasyvoulos N. Pappas, *Member, IEEE*, David L. Neuhoff, *Fellow, IEEE*

**Abstract**—A new model-based approach to digital halftoning is proposed. It is intended primarily for laser printers, which generate “distortions” such as “dot overlap.” Conventional methods, such as clustered-dot ordered dither, resist distortions at the expense of spatial and gray-scale resolution. The proposed approach relies on printer models that predict distortions and, rather than merely resisting them, it exploits them to increase, rather than decrease, both spatial and gray-scale resolution. We propose a general framework for printer models, and find a specific model for laser printers. As an example of model-based halftoning we propose a modification of error diffusion, which is often considered the best halftoning method for CRT displays with no significant distortions. The new version exploits the printer model to extend the benefits of error diffusion to printers. Experiments show that it provides high quality reproductions with reasonable complexity. The new technique produces images that are sharper and have richer gray tones than those obtained with traditional techniques (e.g. “Classical” screening). The quality of images printed on a 300 dpi printer using the new technique is comparable to that of images printed on a 400 dpi printer using traditional techniques. The proposed modified error diffusion technique is compared to Stucki’s MECCA, a similar, not widely known, technique that accounts for dot overlap. MECCA is more efficient computationally, while the proposed algorithm has better performance.

Model-based halftoning can be especially useful in transmission of high quality documents using high fidelity gray-scale image encoders. As we show in a companion paper, in such cases halftoning is performed at the receiver, just before printing. Apart from coding efficiency, this approach permits the halftoner to be tuned to the individual printer, whose characteristics may vary considerably from those of other printers, for example, write-black vs. write-white laser printers.

## I. INTRODUCTION

Digital halftoning is the process of generating a pattern of binary pixels that create the illusion of a continuous-tone image. The term spatial dithering is also used to refer to this process. Digital halftoning is necessary for display of gray-scale images in media in which the direct rendition of gray tones is impossible. The most common example is printing of gray-scale images on paper. In this paper, we introduce new printer models and a halftoning method that exploits them to produce high quality images using standard laser printers.

As motivation for this work, we cite the increasing demand for digital storage and transmission of gray-scale images and the increasing use of laser printers to make hard copies. For

example, the burgeoning success of FAX for text and other black/white documents suggests there may be a similar large market for high fidelity transmission of gray-scale images, for example: photographs, art work, design renderings, magazine layouts, etc. Also, in the future there will likely be large image data bases, for which efficient storage and rapid transmission are needed. With these come the requirements for data compression and high quality printing. The former is addressed in a companion paper [1] (see also [2]), the latter is addressed here. The conventional approach to achieving the latter is to use a high resolution printer. We have estimated that the printer resolution required for visually transparent halftoning with conventional techniques is of the order of 1400 dots per inch (dpi). Such printers are slow and expensive, and likely to remain so in the near future. Thus, new halftoning techniques are needed that will permit transparent halftoning at much lower printer resolutions.

The issues of compression and high quality printing are coupled. Although gray-scale documents can be sent with present FAX (by halftoning before transmission), the quality is lacking and the transmission time is several times larger than for comparably sized black/white documents. Quality can be improved by increasing the resolution of the FAX scanner and printer. However, with present standardized FAX coding methods, transmission time will increase dramatically [1]. Even with proposed improvements to the coding algorithm [3] the transmission time will increase significantly. Fortunately, gray-scale image coding [4], [5] has advanced to the point where it can be used as the basis of image storage and transmission systems. As described in [1], gray-scale images can best be transmitted with high fidelity using gray-scale image encoders, and halftoned at the receiver, just before printing. Apart from coding efficiency, this approach permits the halftoner to be tuned to the individual printer. The latter is advantageous because printer characteristics vary considerably, for example, write-black vs. write-white laser printers. In other words, it permits the proposed model-based halftoning technique to exploit the characteristics of the specific printer.

The basic idea in halftoning is to represent a constant gray level  $x$  (on a scale of 0 = white to 1 = black) by a binary pattern in which the fraction of 1’s is approximately  $x$ . If the 1’s are printed as black spots and the 0’s are left as white spaces and if the distance between adjacent bits is sufficiently small, the eye averages black spots and white spaces, and perceives, approximately, gray level  $x$ . Thus, halftoning relies on the fact that the eye acts as a spatial low-pass filter.

Halftoning also relies on the assumption that the black area of a printed binary pattern is proportional to the fraction of ones in the pattern. This means that the area occupied by each black dot is roughly the same as the area occupied by each white dot. Thus, the “desired” shape for the black spots

Manuscript received May 4, 1992; revised October 4, 1993. Portions of this work were published in the *Proc. SPIE, vol. 1453, Human Vision, Visual Proc., and Digital Display II*, San Jose, CA, February, 1991. The associate editor coordinating the review of this paper and approving it for publication was Prof. Avidesh Zakhor.

T. N. Pappas is with the Signal processing Department, AT&T Bell Laboratories, Murray Hill, NJ 07974 USA.

D. L. Neuhoff is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA.

IEEE Log Number 9407036.

produced by a printer (in response to 1's) would be  $T \times T$  squares, where  $T$  is the bit spacing. However, most printers produce circular spots [6], [7]. As illustrated in Section III, the radius of the dots must be at least  $T/\sqrt{2}$  so that they are capable of blackening a page entirely. This has the unfortunate consequence of making black spots cover portions of adjacent spaces, causing the perceived gray level to be darker than the fraction of ones. We will refer to this phenomenon as “dot overlap.” Moreover, most printers produce black spots that are larger than the minimal size (this is sometimes called “ink spreading”), which further distorts the perceived gray level. The most commonly used digital halftoning techniques protect against dot overlap by clustering black spots so the percentage effect on perceived gray level is reduced. Unfortunately, such clustering constrains the spatial resolution of the perceived images and increases the low frequency artifacts, i.e. the visibility of the halftone patterns. As we will see in Section II, there is a tradeoff between the number of perceived gray levels and the visibility of low frequency artifacts. Rather than trying to resist printer distortions, as in the conventional approach, we here propose a technique that exploits them in order to improve both spatial resolution and low frequency artifacts. A key element in such a method is an accurate printer model. In this paper we propose a general framework for such models and develop a specific model for laser printers.

Having introduced the printer models, we propose a new halftoning method that exploits them. It is an extension of error diffusion [8] and was first presented in [9]. Error diffusion is generally considered to be the best halftoning technique for displays, such as CRT's, that do not suffer from substantial dot overlap distortions [10].<sup>1</sup> It produces images with high spatial resolution and visually pleasant textures, i.e. few low frequency artifacts. With our new model-based approach, the advantages of error diffusion are extended to printers, by incorporating a model for printer distortions. Experiments indicate that the modified error diffusion algorithm offers a substantial improvement over conventional clustered-dot ordered dither in spatial resolution, severity of low frequency artifacts, and even in gray-scale resolution.

The first one to propose an extension of error diffusion to account for printer dot overlap was Stucki [13], [14]. However, his multiple-error correction computation algorithm (MECCA) is not widely known. In fact, when this paper was originally written we were not aware of Stucki's work. In Section IV we review Stucki's algorithm and compare it to the algorithm we propose. We show that, while Stucki's algorithm is more efficient computationally, the proposed algorithm has better performance. In particular, we show that the proposed algorithm produces images that are sharper with more pleasant halftone patterns than Stucki's method.

Most of the halftoning literature assumes perfect printing [8], [10], [15]. Roetling and Holladay [6] proposed a dot-overlap printer model, like we do, but used it only to modify the ordered dither thresholds so that they result in a linear gray scale. Allebach [16] also proposed a modification of

<sup>1</sup>In fact, CRT's present different kinds of distortions which may require similar device-dependent models [11], [12].

TABLE I: Classification of halftoning techniques

Block Replacement		
Screening	Random Dither	
	Ordered Dither	Dispersed Clustered
Error Diffusion		
Least Squares		

ordered dither that takes into account dot overlap to improve the tone scale of the printed images. However, as we will see in Section II, ordered dither is a highly constrained method that still lacks the high spatial resolution and visually pleasant textures of the proposed modified error diffusion. Another technique that takes into account dot overlap was proposed by Pryor et. al [17]. Its performance is also inferior to the proposed method.

A review of current halftoning techniques is presented in Section II. Our new printer models are presented Section III. Section IV presents the proposed extension of error diffusion. The conclusions are summarized in Section V.

## II. REVIEW OF HALFTONING TECHNIQUES

Digital halftoning is the process of transforming a continuous-tone image to a pattern of black and white pixels that, due to the eye's lack of high frequency resolution, have the appearance of the original continuous-tone image. This section reviews standard halftoning techniques and describes how they can be evaluated in terms of their performance in several domains. It will be assumed throughout that the image has been sampled so there is one pixel per dot to be generated.<sup>2</sup> We will also assume that the gray-scale of the original image has been adjusted so that it represents the desired reflectances in the printed image. Table I shows a classification of halftoning techniques. These techniques are described below.

In *block replacement*,<sup>3</sup> the image is subdivided into blocks (e.g.  $8 \times 8$  pixels) and each block is “replaced” by one of a predetermined set of binary patterns (having the same dimensions as the image blocks). Typically, the binary patterns have differing numbers of ones, and the pattern whose fraction of ones best matches the gray level of the image block is selected.

In *screening*, the image array is compared, pixel by pixel, to an array of image-independent thresholds. A black dot is placed wherever the image gray level is greater than the corresponding threshold. In so called *random dither*, the thresholds are randomly generated. In *ordered dither* they are periodic. More specifically, the threshold array is generated by periodically replicating a threshold matrix (e.g.  $8 \times 8$ ). Figure 1

<sup>2</sup>When number of samples of a given image is different from the number of dots to be generated, interpolation is necessary. *Bilinear* and *spline* interpolation can produce an image of any size. However, for sampling rate conversions by a rational factor, the best results are obtained by an *expander* followed by an equiripple low-pass FIR filter [18, pp. 105–109].

<sup>3</sup>Also called *pulse-surface-area modulation* [10, p. 77].

.576	.635	.608	.514	.424	.365	.392	.486
.847	.878	.910	.698	.153	.122	.090	.302
.820	.969	.941	.667	.180	.031	.059	.333
.725	.788	.757	.545	.275	.212	.243	.455
.424	.365	.392	.486	.576	.635	.608	.514
.153	.122	.090	.302	.847	.878	.910	.698
.180	.031	.059	.333	.820	.969	.941	.667
.275	.212	.243	.455	.725	.788	.757	.545

(a)

.513	.272	.724	.483	.543	.302	.694	.453
.151	.755	.091	.966	.181	.785	.121	.936
.634	.392	.574	.332	.664	.423	.604	.362
.060	.875	.211	.815	.030	.906	.241	.845
.543	.302	.694	.453	.513	.272	.724	.483
.181	.785	.121	.936	.151	.755	.091	.966
.664	.423	.604	.362	.634	.392	.574	.332
.030	.906	.241	.845	.060	.875	.211	.815

(b)

.917	.250	.583
.750	.083	.417

(c)

.917	.583	.250
.417	.083	.750

(d)

Fig. 1: Ordered dither threshold matrices: (a) “Classical-4” ( $8 \times 8$ , clustered); (b) “Bayer-5” ( $8 \times 8$ , dispersed); (c) “ $2 \times 3$  clustered;” (d) “ $2 \times 3$  dispersed.”

shows the matrices used with ordered dither in this report. The first two are standard matrices [10, pp. 86, 135].<sup>4</sup> The  $2 \times 3$  matrices are not likely to be used in practice; they are used to provide simple nontrivial examples of halftoning properties.

The matrices in Figure 1(a) and (c) are examples of *clustered* ordered dither. The thresholds are grouped according to size so that they produce dots that are clustered together to form bigger dots or *macrodots*. The matrices in Figure 1(b) and (d) are examples of *dispersed* ordered dither. The thresholds are arranged so that they produce dots that are dispersed. As we will see in this section, dispersed dither produces images with better spatial resolution and fewer low frequency artifacts, while clustered dither is more robust to printer distortions.

In *error diffusion*, the image pixels are also compared to thresholds, like in screening, but in this case the threshold for each image pixel is dependent upon “prior” image pixels (usually above and to the left). Error diffusion is described in more detail in Section IV, where we introduce an extension

<sup>4</sup>The designations “classical-4” and “Bayer-5” are consistent with Ulichney’s notation.

that accounts for printer distortions such as dot overlap.

In *least-squares* halftoning, an “optimal” halftone image is obtained by minimizing the squared error between the eye-filtered binary and the eye-filtered original gray-scale image [15]. A similar distortion measure was proposed by Allebach in [19]. Recently, the least-squares approach has received a lot of attention [20]–[22]. The least-squares approach with a printer model is considered in [23], [24] and also in a forthcoming paper.

The performance of a halftoning technique may be evaluated in four domains: regions of constant gray level, regions of slowly changing gray level, regions of rapidly changing gray level, and performance in the presence of printer distortions such as dot overlap. In the following we will discuss the basic principles and the specific techniques in the context of these domains. We presume no printer distortions until the subsection on such.

The images we use to demonstrate the various halftoning methods are shown in Figure 2. The resolution of the “Bank,” “Lena,” and “6386” images is  $256 \times 256$  pixels. The resolution of the “Ramp” image is  $788 \times 80$  pixels. The gray-scale resolution of the original images is 8 bits/pixel. The images in Figure 2 were printed using a LINOTRONIC 200P printer at 1270 dpi. The images in Figures 3, 4, 5, 13, 15, and 17 are magnified details of the halftone images printed on a 300 dpi HP LASERJET II printer. They are printed with simulated ink spreading at one third of the printer resolution. The images in Figure 16 are magnified details of the halftone images printed on a DATAPRODUCTS LZR 1560 printer at 400 dpi. They are also printed with simulated ink spreading at one third of the printer resolution.

#### A. Performance in Regions of Constant Gray Level

The two most important considerations in halftoning regions of constant intensity are the number of gray levels and the severeness of low frequency artifacts. Another important consideration, as we will see in Subsection II-D, is the accuracy of gray level rendition. In fact, there is a tradeoff between gray-scale resolution and low frequency artifacts. To see this, let us ignore printer distortions and the specific characteristics of the eye, and simply assume that the halftoning method generates periodic patterns whose perceived gray level equals the fraction of 1’s in the pattern. To obtain good gray-scale resolution it is desirable to allow the period to be large. On the other hand, periodic patterns with large period may cause undesirably visible artifacts. This limitation on gray-scale resolution, induced by the spatial resolution of the eye, also applies to the patterns produced by any halftoning technique, for it basically says that the number of perceivable levels is limited by the size of the region over which the eye averages. Dispersed-dot schemes produce less objectionable low frequency artifacts than clustered dither.

Both block replacement and ordered dither produce periodic patterns when halftoning an image with constant gray level. With the former, the periodic patterns are those formed by replicating a “replacement pattern,” and the number of gray levels that can be produced equals the number of gray levels



Fig. 2: Original test images

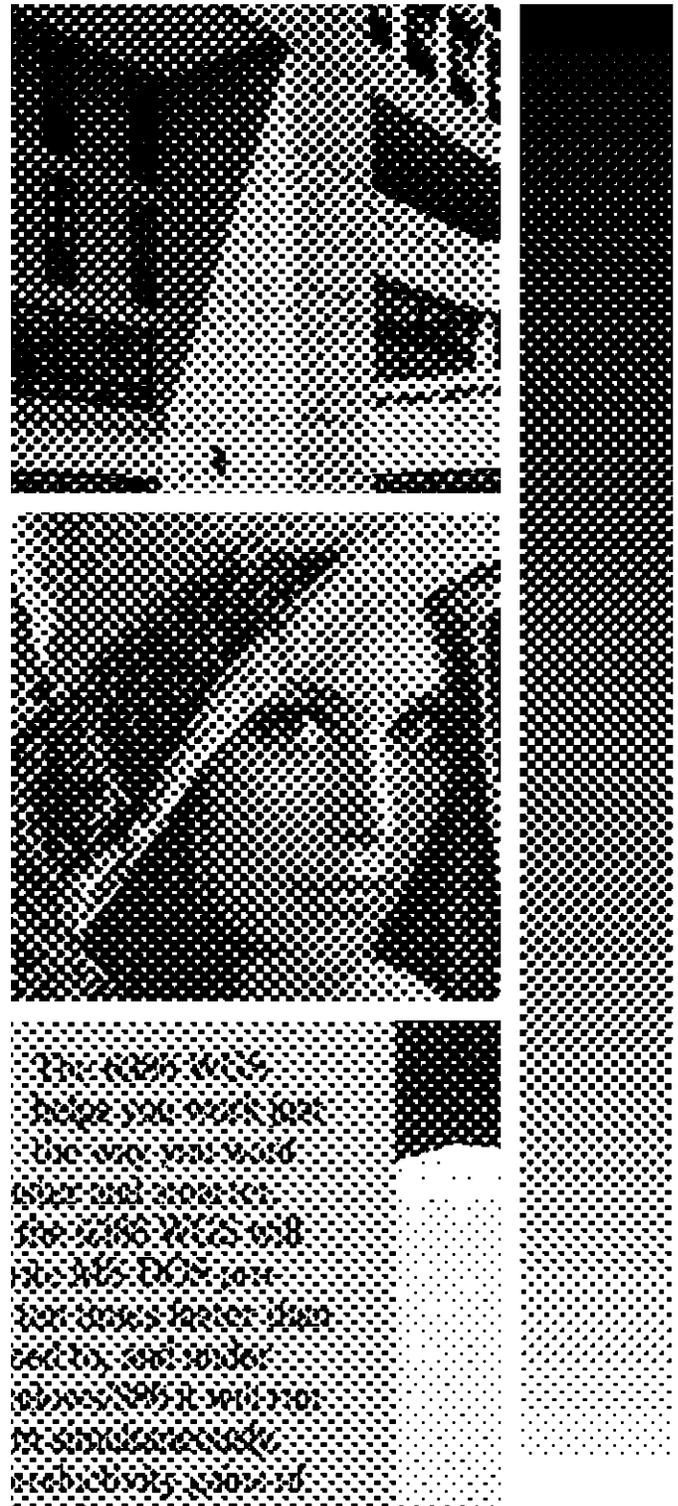


Fig. 3: Classical screen (no microdither) at 300 dpi

of such patterns. With the latter, the periodic patterns are those formed by replicating a “fundamental pattern,” obtained by “slicing” the threshold matrix. Specifically, for some value  $x$  between 0 and 1, we form a fundamental pattern with the same dimensions as the threshold matrix by placing a one wherever the matrix is less than  $x$  and a zero elsewhere. Thus, the number of fundamental patterns, and hence gray levels,

that the threshold matrix can produce, equals the number of distinct values in the matrix, plus one. For example, the “ $2 \times 3$  clustered” scheme of Figure 1(c) produces the 7 gray levels listed in Table II, along with the corresponding fundamental patterns.

For most observers, and for viewing 300 dpi printing at two feet, the matrices in Figure 1(a) and (b) represent a good

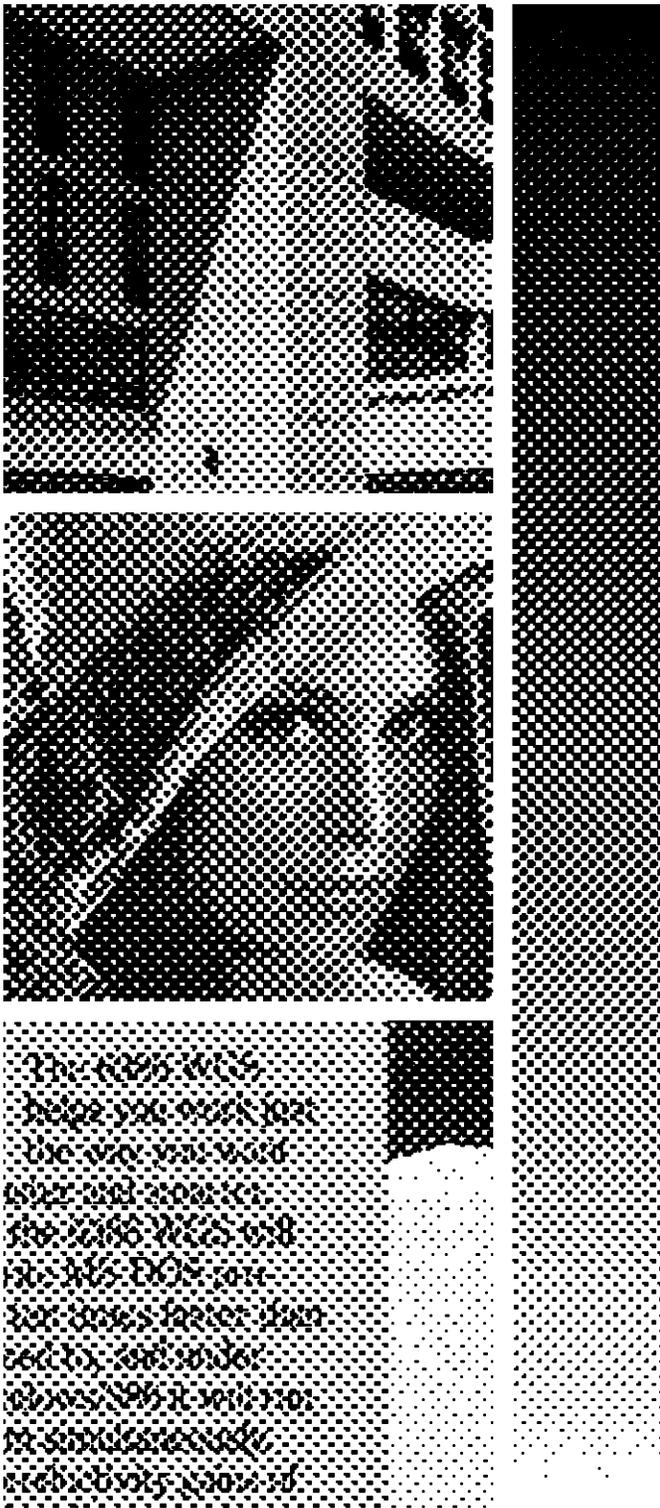


Fig. 4: Classical screen (with microdither) at 300 dpi

compromise between the conflicting requirements for more gray levels and for low visibility of artifacts. “classical-4” and “Bayer-5” produce 33 gray levels. Note that the first and fourth quadrant are the same and so are the second and third. Thus, they produce patterns with a diagonal structure and, even though they are  $8 \times 8$  screens, the “effective period” is approximately 6 pixels ( $4\sqrt{2}$ , the distance between the two

TABLE II: Gray levels produced by “ $2 \times 3$  clustered” ordered dither.

Pattern	000	000	010	010	011	011	111
	000	010	010	011	011	111	111
Gray Level	0	1/6	2/6	3/6	4/6	5/6	1

identical quadrants). Patterns with such diagonal structure are less objectionable to the eye because the eye is least sensitive to 45 degree sinusoids [25].

The number of gray levels that the eye can resolve depends on a variety of factors like signal-to-noise ratio, sampling density, and subject matter [26, p. 99]. According to Schreiber this number is typically between 64 and 128 gray levels. In the following discussion, we will assume that the eye can resolve 64 gray levels [26, p. 99]. Thus halftoning with 300 dpi at two feet (using “classical-4,” as we discussed above) can at best attain about 50% of the eye’s gray-scale resolution. Alternatively, one must go to about 400 dpi printing at two feet<sup>5</sup> or 800 dpi at one foot to attain the full gray-scale resolution of the eye. For visually transparent halftoning (i.e. no low frequency artifacts) with conventional techniques, we estimate that the required printer resolution is of the order of 1400 dpi.

Random dither, error diffusion, and least-squares halftoning do not produce periodic patterns. But, as mentioned above, they are subject to the same limitations when halftoning an image with constant gray level. In the absence of printer distortions, the number of gray levels they produce is essentially the same as in screening. As we will see in the next section, in the presence of dot overlap these methods can actually produce more gray levels than screening. The patterns produced by random dither are totally random, and because of this they have such visible and objectionable low frequency artifacts that random dither is almost never used in practice [10].

### B. Performance in Regions of Slowly Changing Gray Level

When gray level changes smoothly over a region, the halftoned output may change abruptly from one representable gray level to another. If this happens simultaneously along a contour of constant image gray level, the contour appears as an edge. Such “false contouring” is an undesirable artifact of many halftoning techniques. For example it is a noticeable problem with block replacement and ordered dither.

To reduce false contouring, a small amount of noise, called *microdither*<sup>6</sup>, can be added to each pixel of the image, before halftoning. The noise must be white, uniformly distributed, and its amplitude must be equal to half the quantization level spacing. This is essentially Roberts’ technique [27] for removing false contouring in quantized images. Roberts adds the noise signal before quantization and subtracts an identical noise signal after quantization. In this case the quantizer is the

<sup>5</sup>A  $5 \times 10$  “classical” screen at 375 dpi (50 gray levels) and a  $6 \times 12$  “classical” screen at 450 dpi (72 gray levels) have the same macrodot spacing as the  $4 \times 8$  “classical-4” screen at 300 dpi.

<sup>6</sup>Since the term (spatial) dither is also used to denote halftoning, we use the term microdither to avoid confusion.

halftoning technique. However, because of the binary display it is impossible to subtract the noise signal after quantization. In the case of ordered dither with a threshold matrix having  $M$  distinct and uniformly spaced thresholds, it suffices to add an independent sample of noise, uniformly distributed in the interval  $[-1/2M, 1/2M]$ .<sup>7</sup> Figures 3 and 4 show the test images halftoned using “classical-4” ordered dither, without and with microdither, respectively. The images in Figure 5 were halftoned using the “Bayer-5” screen with microdither.

As we will see in Section IV, error diffusion does not suffer from false contouring. The same is true for least-squares halftoning.

### C. Performance in Regions of Rapidly Changing Gray Level

A halftoning method should also be judged by its spatial resolution; i.e., its ability to follow rapid changes in gray level, for example where there is a step change at an edge. In this regard, we mention that block replacement produces noticeable blurring [28]. Since screening uses dot by dot thresholding, it can more accurately follow sizable changes in image gray level. Dispersed-dot dither produces sharper images than clustered-dot dither [10]. This is because the thresholds of similar size are spread out and any rapid change is more likely to be detected. The images in Figure 5 are halftoned by “Bayer-5” and are clearly sharper than the images in Figure 3 and 4 that are halftoned by “classical-4.” As we will see in Section IV, error diffusion allows more flexibility in the placement of dots, thus resulting in images that are sharper than those produced by any ordered dither technique. The same observation holds for the least-squares approach [23].

### D. Performance in the Presence of Printer Distortions

As indicated in Section III, printer distortions may have a large effect on the appearance of a halftoned image. They distort the linearity and even the monotonicity of the gray scale; that is, the produced gray level in a region may not correspond well to the fraction of ones therein. Dispersed-dot techniques, such as ordered dither with “Bayer-5,” produce halftoned output that looks good when displayed on a CRT screen (with little or no display distortions) but looks bad when printed with distortions such as dot overlap. Figure 5 shows the test images halftoned with “Bayer-5.” The distortions to gray scale show most clearly in the halftoned reproduction of a linear ramp image.

A standard way to reduce the effects of dot overlap and other printer distortions is to produce dots in clusters or macrodots. Such clusters are less affected by printer distortions than individual dots, and the gray scale is less distorted. With ordered dither, macrodots may be formed by appropriate choice of the elements in the threshold matrix. For example, it may be seen that a macrodot will emerge from both the lower left and upper right corners of the “classical-4” threshold matrix. The result is that gray level is represented by the size of the macrodots, as opposed to their spacing or frequency, which

<sup>7</sup>In the presence of dot overlap distortions, the spacing of the quantization levels that ordered dither generates is not uniform. Thus, the amplitude of the noise must be equal to half the maximum quantization level spacing.



Fig. 5: Dispersed ordered dither (Bayer) at 300 dpi

are fixed. This approach mimics traditional analog halftoning techniques used in printing.

We now summarize our review of halftoning techniques. When little or no dot overlap is present, as with most CRT displays, it has been found that error diffusion gives the best results [10]. Ordered dither is a simpler method that works fairly well. Random dither and block replacement are almost

never used. When there is dot overlap, as with laser printers, clustered ordered dither schemes like the “classical” family are the only regularly used techniques [10]. While the clustered dot approach is successful in reducing the effects of dot overlap, it sacrifices spatial resolution and generates more low frequency artifacts. In contrast, it is the purpose of this paper to show that dot overlap can be exploited, rather than avoided, to increase the number of gray levels, and, at the same time, increase the spatial resolution and reduce the low frequency artifacts. In Section IV we show that this can be accomplished by error diffusion which can be adapted to account for dot overlap.

### III. PRINTER MODELS

In this section we introduce a framework for printer models and a specific model for laser printers. The purpose of a model is to accurately predict the gray levels produced by a printer. Our printer models are independent of the specific characteristics of the human visual system, which should be considered separately.

Our work is oriented to high resolution laser printers. We have used “write-black” laser printers with 300 dpi resolution (such as the HP LASERJET II or the DATAPRODUCTS LZR 1260) as our test vehicles. However, the discussion and methods are intended to apply directly, or be adaptable, to any printer whose resolution is not too small (e.g. to a “write-white” laser printer such as the DATA PRODUCTS LZR 2665). To a first approximation, such printers are capable of producing black spots (more commonly called dots) on a piece of paper, at any and all sites of a Cartesian grid with horizontal and vertical spacing of  $T$  inches. The reciprocal of  $T$  is generally referred to as the *printer resolution* in dots per inch (dpi). The printer is controlled by sending it an  $N_W \times N_H$  binary array  $[b_{i,j}]$ , where  $b_{i,j} = 1$  indicates that a black dot is to be placed at “site  $(i, j)$ ” which is located  $iT$  inches from the left and  $jT$  inches from the top of the image, and  $b_{i,j} = 0$  indicates that the site is to remain white. We’ll refer to the latter as a “white” dot.

As we saw, printers produce round rather than square black dots [6]. As illustrated in Figure 6, the black dots produced by an “ideal” printer are black circles (no shading) with radius  $T/\sqrt{2}$ . The latter is the smallest radius such that black circles placed at all sites completely cover the page. The area of such a dot is  $1.57T^2$ , i.e., 57% larger than a  $T \times T$  square. Accordingly, horizontally or vertically (but not diagonally) neighboring black dots overlap, and white dots are darkened by neighboring black dots. Specifically, if a white dot has  $d$  horizontally or vertically neighboring black dots, then 14.3  $d$ % of it is blackened.

With an actual printer the black dots aren’t perfectly round, they’re not perfectly black, they’re not the ideal size, and they may be slightly misplaced. There are other intriguing phenomena, as well. For example, a white line surrounded by several black lines is brighter than when surrounded by two single lines. There are all sorts of potential causes for such distortions, e.g., ink spreading, spreading of the laser beam, interaction of the laser and the charge applied to the drum,

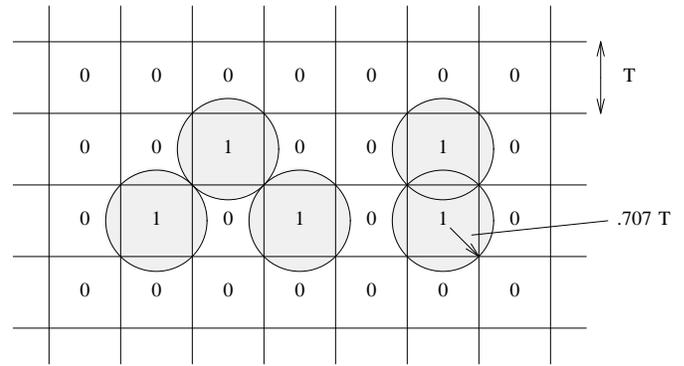


Fig. 6: Ideal black spots.



Fig. 7: Periodic test patterns.

the movement of toner particles in reaction to charge, the heat finishing, reflections of light within the paper, and so on.

Some idea of the distortions introduced by a printer may be obtained by considering a set of periodic patterns like those shown in Figure 7. A set of such patterns (with period 6) was printed on a 300 dpi write-black HP LASERJET II printer. The different patterns, represented by one period, are listed in Table III. The table also lists the frequency of ones in each pattern, the gray level predicted by the circular dot-overlap model we will consider below, and the measured reflectance density of each printed pattern. The higher the measured density of the pattern the darker it appears to the eye. Notice that patterns with the same frequency of ones differ substantially in their reflectance densities. Such effects are also evident when we compare the images in Figures 3, 4, and 5 that have been halftoned using two traditional techniques. Notice that images in Figure 5 are much darker than the images in Figures 3 and 4, even though in any region the two methods present patterns with essentially the same number of ones.

As a result of phenomena such as those mentioned above, the gray level produced by the printer at any point in the image depends in some complicated way on the surrounding bits. Let  $u(s, t)$  be the gray level produced by the printer at point  $(s, t)$  located  $s$  inches from the left and  $t$  inches from the top of the image. Then,

$$u(s, t) = f(s, t; B_{s,t}) \quad \begin{aligned} T/2 \leq s \leq N_W T + T/2, \\ T/2 \leq t \leq N_H T + T/2 \end{aligned} \quad (1)$$

where  $B_{s,t}$  denotes the set of bits in a neighborhood of the point  $(s, t)$  and  $f$  is some function.<sup>8</sup> However, due to the close spacing of the dots and the limited spatial resolution of the eye,

<sup>8</sup>Note that the function  $f$  could be deterministic or probabilistic, as suggested in [14].

TABLE III: Periodic patterns at 300 dpi

Pattern	Frequency of 1's	gray level predicted by circular dot-overlap model ( $\alpha = 0.33$ )	Measured density
000000	0.00	0.00	0.00
100000	0.17	0.28	0.29
100100	0.33	0.55	0.76
101000	0.33	0.55	0.62
110000	0.33	0.44	0.41
101010	0.50	0.83	1.46
101100	0.50	0.72	0.87
111000	0.50	0.61	0.57
110110	0.67	0.89	1.33
101110	0.67	0.89	1.30
111100	0.67	0.78	0.75
111110	0.83	0.94	1.15
111111	1.00	1.00	1.57

the gray level  $u(s, t)$  of the printed image can be modeled as having a constant value  $p_{i,j}$  in the vicinity of site  $(i, j)$  as follows

$$\tilde{u}(s, t) = p_{i,j} \quad \begin{aligned} iT - T/2 \leq s \leq iT + T/2, \\ jT - T/2 \leq t \leq jT + T/2 \end{aligned} \quad (2)$$

for all  $1 \leq i \leq N_W$  and  $1 \leq j \leq N_H$ . Although the gray level is not actually constant, the eye responds, essentially, only to the average gray level over the site. It is this average gray level that  $p_{i,j}$  represents, namely

$$p_{i,j} = \frac{1}{T^2} \int_{iT-T/2}^{iT+T/2} \int_{jT-T/2}^{jT+T/2} f(s, t; B_{s,t}) ds dt \quad 1 \leq i \leq N_W \quad 1 \leq j \leq N_H \quad (3)$$

It follows from Eq. (3) that the average level  $p_{i,j}$  depends on the neighboring bits. Thus

$$p_{i,j} = \mathcal{P}(W_{i,j}) \quad 1 \leq i \leq N_W, \quad 1 \leq j \leq N_H \quad (4)$$

where  $W_{i,j}$  is a window that consists of the bits in some neighborhood of  $b_{i,j}$  and  $\mathcal{P}$  denotes some function thereof. Note that the printer model is described by the array  $[b_{i,j}]$ . Thus, given the binary array  $[b_{i,j}]$  that specifies the dot pattern to be printed, our printer model generates a new array  $[p_{i,j}]$  of gray levels which has the same dimensions as the binary array. This simplifies processing considerably because we only have to work with a discrete set of image values instead of the complicated description of Eq. (1).

In tailoring a model of the above form to a given printer, the main task is to identify how the function  $\mathcal{P}$  specifying  $p_{i,j}$  depends on the bits in the neighborhood of  $b_{i,j}$ . As mentioned earlier, there are a variety of phenomena that may contribute to this dependence, and the only place to account for such is in the function  $\mathcal{P}$ . For computational efficiency, it is essential that  $p_{i,j}$  be entirely determined by the bits in a small window around  $b_{i,j}$ , e.g. a  $3 \times 3$  window. In this case, the possible values of  $\mathcal{P}$  can be listed in a table, e.g. with  $2^9$  elements. The individual elements of this table might be derived from

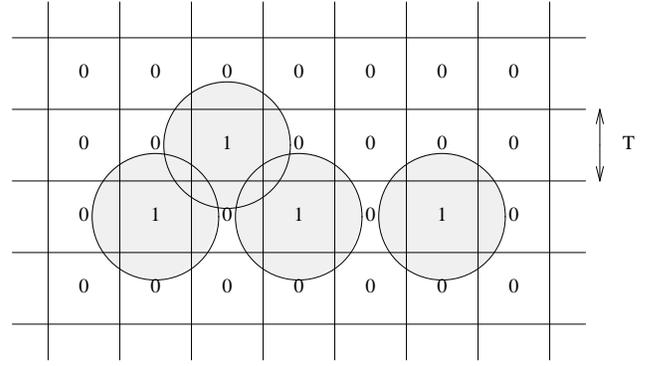


Fig. 8: Actual dot overlap.

a detailed physical understanding of the various phenomena effecting gray level or from measurements of the gray level that results when various dot patterns are printed. In this paper we consider an example of the first approach. The second approach is explored in [29], [30].

#### A. Circular Dot-Overlap Model

The most elementary distortion introduced by most printers is that, as illustrated in Figure 8, their dots are larger than the minimal covering size, as would occur if “ink spreading” occurred (there may be other causes as well). A *circular dot-overlap* model that accounts for such has

$$p_{i,j} = \mathcal{P}(W_{i,j}) = \begin{cases} 1, & \text{if } b_{i,j} = 1 \\ f_1\alpha + f_2\beta - f_3\gamma, & \text{if } b_{i,j} = 0 \end{cases} \quad (5)$$

where the window  $W_{i,j}$  consists of  $b_{i,j}$  and its eight neighbors, as indexed below

$$W_{i,j} = \begin{bmatrix} b_{nw} & b_n & b_{ne} \\ b_w & b_{i,j} & b_e \\ b_{sw} & b_s & b_{se} \end{bmatrix} \quad (6)$$

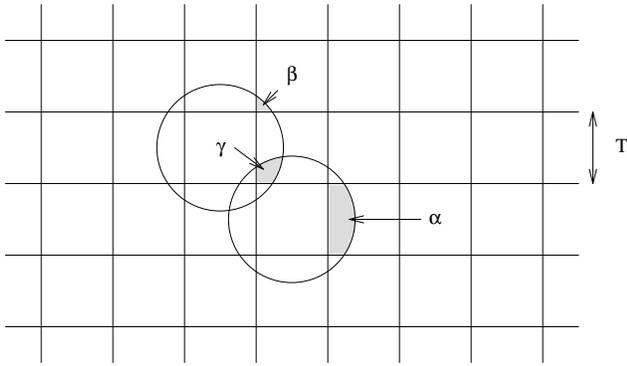
where  $f_1$  is the number of horizontally and vertically neighboring dots that are black (i.e., the number of ones in the set  $\{b_n, b_e, b_s, b_w\}$ ),  $f_2$  is the number of diagonally neighboring dots (i.e., among  $\{b_{nw}, b_{ne}, b_{se}, b_{sw}\}$ ) that are black and not adjacent to any horizontally or vertically neighboring black dot (e.g.  $b_{nw} = 1$  and  $b_n = b_w = 0$ ), and  $f_3$  is the number of pairs of neighboring black dots in which one is a horizontal neighbor and the other is a vertical neighbor (e.g.,  $b_n = b_w = 1$ ), and where  $\alpha$ ,  $\beta$ , and  $\gamma$  are the ratios of the areas of the shaded regions shown in Figure 9 to  $T^2$ .

The parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  can be expressed in terms of the ratio  $\rho$  of the actual dot radius to the ideal dot radius  $T/\sqrt{2}$  as follows:

$$\alpha = \frac{1}{4} \sqrt{2\rho^2 - 1} + \frac{\rho^2}{2} \sin^{-1} \left( \frac{1}{\sqrt{2\rho}} \right) - \frac{1}{2} \quad (7)$$

$$\beta = \frac{\pi\rho^2}{8} - \frac{\rho^2}{2} \sin^{-1} \left( \frac{1}{\sqrt{2\rho}} \right) - \frac{1}{4} \sqrt{2\rho^2 - 1} + \frac{1}{4} \quad (8)$$

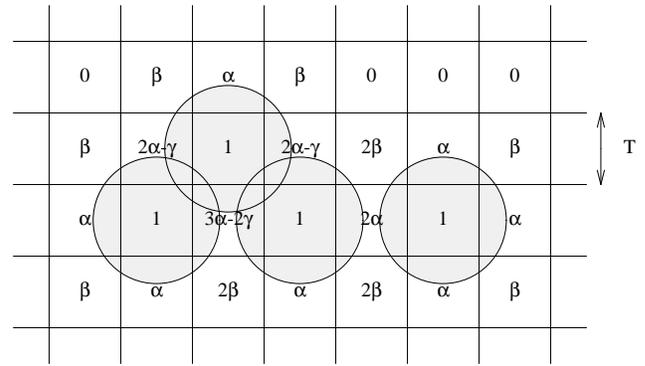
$$\gamma = \frac{\rho^2}{2} \sin^{-1} \left( \sqrt{\frac{\rho^2 - 1}{\rho^2}} \right) - \frac{1}{2} \sqrt{\rho^2 - 1} - \beta \quad (9)$$

Fig. 9: Definition of  $\alpha$ ,  $\beta$ , and  $\gamma$  for printer model.

The above presumes  $1 \leq \rho \leq \sqrt{2}$ ; i.e., the black dots are large enough to cover a  $T \times T$  square, but not so large that black dots separated (horizontally or vertically) by one white dot would overlap. Since the parameter  $\alpha$  is the largest of the three factors, we will refer to this model as  $\alpha$  dot overlap. It represents the fraction of a horizontally or vertically neighboring site covered by a black dot. Note that the model is not linear in the input bits, which is due to the fact that the paper saturates at black intensity. For a printer with the ideal dot size  $\rho = 1$ , the minimum value, and  $\alpha = .143$ ,  $\beta = 0$ , and  $\gamma = 0$ . For  $\rho = \sqrt{2}$ , the maximum value,  $\alpha = .46$ ,  $\beta = .079$ , and  $\gamma = .21$ . We will refer to a hypothetical printer with no dot overlap (square dots) as the  $\alpha = 0$  dot-overlap printer.

For the HP printer, our experience indicates that for  $\rho \approx 1.25$  the circular dot-overlap model matches the HP printer about as well as it can.<sup>9</sup> This value results in  $\alpha = .33$ ,  $\beta = .029$ , and  $\gamma = .098$ . Figure 10 illustrates how the dot pattern in Figure 8 is modeled with these values. Table III shows the gray levels predicted by such a model for various periodic patterns. Since the patterns are horizontally invariant, the gray level of a white dot depends only on the presence or absence of vertically neighboring black dots. Specifically, the gray level of a white dot is  $\alpha$ ,  $2\alpha$  or 0 depending on whether there are one, two or no vertically neighboring black dots. One can see from the gray levels predicted in Table III that the circular dot-overlap model does much to explain how patterns with the same numbers of ones can have different gray levels. For example, it predicts the relative gray levels among the patterns with 3 ones. On the other hand it doesn't explain why the pattern 100100 is darker than the pattern 101000, nor why 101010 is darker than 111110. Such effects can only be captured by printer models with window sizes larger than  $3 \times 3$ .

Roetling and Holladay [6] used a circular dot-overlap model to modify the ordered dither thresholds so that they result in a linear gray scale. They also used the model to optimize screen design.<sup>10</sup> However, clustered ordered dither sacrifices spatial resolution and generates more low frequency artifacts. In the next section, we show that our printer models can be used to obtain a linear gray scale and, at the same time, a substantial

Fig. 10: Circular dot-overlap model with  $\alpha = .33$ ,  $\beta = .029$ , and  $\gamma = .098$ .TABLE IV: Gray levels produced by “ $2 \times 3$  clustered” ordered dither.

Pattern	000	000	010	010	011	011	111
Gray Level ( $\alpha = 0$ )	0.0	.17	.33	.50	.67	.83	1.0
Gray Level ( $\alpha = .33$ )	0.0	.41	.55	.69	.88	.99	1.0

improvement in spatial resolution and low frequency artifacts. Moreover, the printer model can even be exploited to improve the gray-scale resolution as we show below.

### B. Halftoning in the Presence of Printer Distortions

Assuming that the above printer model is valid, we now show that dot overlap can be exploited to obtain more gray levels than could be obtained without it. As an indication of what is possible, we observe that without dot overlap there are 7 gray levels that can be produced by  $2 \times 3$  periodic patterns. As an example, consider periodic repetitions of the fundamental patterns generated by the “ $2 \times 3$  clustered” ordered dither matrix of Figure 1(c). The resulting gray levels are shown in Table IV, assuming a printer model of the form (4) with  $\alpha = 0$  (no dot overlap) and with  $\alpha = .33$  (the usual dot overlap). The gray levels of a dispersed-dot ordered dither are distorted more severely by dot overlap, as can be seen in Table V.

In the presence of dot overlap, the perceived gray level is determined by, not only the fraction of ones in the pattern, but also by their relative locations. Thus,  $2 \times 3$  periodic sequences can produce 10 gray levels, as shown in Table VI. Since ordered dither allows only one pattern for a given set

<sup>9</sup>As the printer cartridge ages, lower values of  $\rho$  provide a better match.

<sup>10</sup>They actually calculated the amount of overlap by subdividing the area of each pixel into a  $33 \times 33$  grid.

TABLE V: Gray levels produced by “ $2 \times 3$  dispersed” ordered dither.

Pattern	000	000	001	001	011	011	111
	000	010	010	110	110	111	111
Gray Level ( $\alpha = 0$ )	0.0	.17	.33	.50	.67	.83	1.0
Gray Level ( $\alpha = .33$ )	0.0	.41	.66	.92	.98	.99	1.0

TABLE VI: Gray levels produced by  $2 \times 3$  periodic sequences.

Pattern	000	000	010	001	010	011	001	011	011	111
	000	010	010	010	011	011	110	110	111	111
Gray Level ( $\alpha = 0$ )	0.0	.17	.33	.33	.50	.67	.50	.67	.83	1.0
Gray Level ( $\alpha = .33$ )	0.0	.41	.55	.66	.69	.88	.92	.98	.99	1.0

of zeros and ones, it cannot exploit this added flexibility; changing the screen we change the gray levels we get, not their number. On the other hand, block replacement can exploit this flexibility at the expense of spatial resolution. Thus, we have to consider other techniques that can increase the gray-scale resolution while maintaining, or improving spatial resolution of the printed images. One such technique is the modified error diffusion presented in the next section.

#### IV. ERROR DIFFUSION

Error diffusion is a popular method for generating sharp halftone images for displays, such as CRT’s, that do not suffer from substantial dot overlap or other distortions [8], [10]. In this section we show that the models we introduced in Section III make it possible to extend the advantages of error diffusion to printers.

In error diffusion each image pixel is compared to a threshold which depends upon “prior” image pixels, usually above and to the left. Alternatively, each image pixel is compared to a fixed threshold, after a correction factor is added to its original gray level to account for past “errors.” Let  $[x_{i,j}]$  be a gray-scale image (after possible interpolation), where  $x_{i,j}$  denotes the pixel located at the  $i$ -th column and the  $j$ -th row. Without loss of generality, we assume that the image is scanned left to right top to bottom. The binary image  $[b_{i,j}]$  produced by error diffusion is obtained by the following set of equations

$$v_{i,j} = x_{i,j} - \sum_{m,n} h_{m,n} e_{i-m,j-n} \quad (10)$$

$$b_{i,j} = \begin{cases} 1, & \text{if } v_{i,j} > t \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$e_{i,j} = b_{i,j} - v_{i,j} \quad (12)$$

Here  $v_{i,j}$  is the “corrected” value of the gray-scale image. The error  $e_{i,j}$  at any “instant”  $(i,j)$  is defined as the difference between the “corrected” gray-scale image and the binary image. The “past” errors are filtered and subtracted from the current image value  $x_{i,j}$  before it is thresholded to obtain the binary value  $b_{i,j}$ , where  $[h_{m,n}]$  is the impulse response of the linear filter. Thus errors are “diffused” over the image.

A diagram of the algorithm is shown in Figure 11. The threshold  $t$  is fixed at .5, the middle of the gray-scale range. The linear filter  $[h_{m,n}]$  has non-symmetric half-plane support, which is the two-dimensional equivalent of causality. The filter coefficients are positive and their sum is equal to one. This

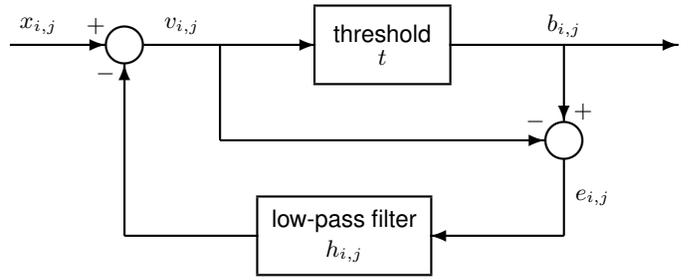


Fig. 11: Standard error diffusion

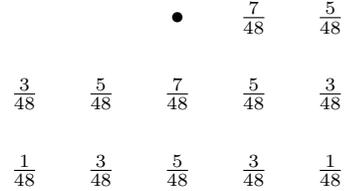


Fig. 12: Impulse response of Jarvis, Judice and Ninke error filter

guarantees stability. The decisions at any point in the image are instantaneous. This is possible because of the “causality” of the filter. Error diffusion requires only one pass through the data.

Various error diffusion filters have been suggested in the literature (see [10]). In the following examples we will use the filter proposed by Jarvis, Judice and Ninke [31], whose impulse response is shown in Figure 12.

Like most of the known halftoning schemes, error diffusion makes implicit use of the fact that the eye acts as a spatial low-pass filter. It shapes the noise, i.e. the difference between the gray-scale image and the halftone image, so that it is not visible by the eye. Error diffusion produces noise with most of its energy concentrated in the high frequencies; this is called blue noise.

In *regions of rapidly changing gray level*, error diffusion accomplishes high resolution by spreading the dots. It is thus very sensitive to dot overlap, in contrast to the clustered-dot schemes like “classical” screening. In the *presence of dot overlap*, error diffusion produces very dark images, as can be seen in Figure 13. This has limited its application to cases with no dot overlap. However, it is actually easy to correct this problem by using the printer models we developed in Section III. The *modified error diffusion* algorithm that compensates for dot overlap is shown in Figure 14. The error is now defined as the difference between the “corrected” gray-scale image  $v_{i,j}$  and the output of the printer model, rather than the binary image. Thus, it accounts for printer distortions as well as quantization effects. The modified error diffusion equations are

$$v_{i,j} = x_{i,j} - \sum_{m,n} h_{m,n} e_{i-m,j-n}^{i,j} \quad (13)$$

$$b_{i,j} = \begin{cases} 1, & \text{if } v_{i,j} > t \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

$$e_{m,n}^{i,j} = p_{m,n}^{i,j} - v_{m,n} \quad \text{for } (m,n) \prec (i,j) \quad (15)$$

where  $(m, n) \prec (i, j)$  means  $(m, n)$  precedes  $(i, j)$  in the scanning order and

$$p_{m,n}^{i,j} = \mathcal{P}(W_{m,n}^{i,j}) \quad \text{for } (m, n) \prec (i, j) \quad (16)$$

where  $W_{m,n}^{i,j}$  consists of  $b_{m,n}$  and its neighbors as in Eq. (4), but here the neighbors  $b_{k,l}$  have been determined only for  $(k, l) \prec (i, j)$ ; they are assumed to be zero (i.e. white) for  $(k, l) \succeq (i, j)$ . Since only the dot-overlap contributions of the “past” pixels can be used in Eq. (16), the “past” errors keep getting updated as more binary values are computed. Hence the dependence of the error and the printer model output on the “instant”  $(i, j)$ . One would expect that the assumption that the undetermined pixels are white would lead to a bias in the gray scale of the printed image. This is indeed the case, but this bias is very small and difficult to detect.<sup>11</sup> Figure 15 shows the results of the modified error diffusion algorithm based on the circular dot-overlap model of Eq. (5). The gray scale of the images has now been restored and, at the same time, they have the sharpness of error diffusion.

An alternative way to compensate for printer distortions, is to directly measure the output of the printer in response to a gray-scale ramp halftoned by error diffusion, and then derive a compensating gray-scale transformation to be applied to the image before halftoning as in [10, p. 36]. To avoid direct measurement, which can be quite painful, one can actually use our printer model to get an estimate of the effect of the printer on the gray-scale ramp, and then derive the transformation. Both of these approaches were explored in [32]. As Ulichney points out, however, this cannot be done when the printer introduces a nonmonotonic nonlinearity. Moreover, this approach is imprecise because the patterns that error diffusion produces depend on the gray level of the present as well as the previous pixels, while the compensating transformation is applied to each point independently.

We now examine the performance of the modified error diffusion algorithm in *regions of constant gray level*. As we discussed above, error diffusion produces blue noise, which has most of its energy in the higher frequencies where it is not visible. Thus, it minimizes the low frequency artifacts and produces images that are very pleasant to the eye [10, p. 338]. Error diffusion produces patterns that are quite regular, but not periodic. However, error diffusion is not entirely free of artifacts. It is well known [8], [10] that for some images error diffusion produces artifacts and asymmetries, like the sparse dots at the bottom of the “Ramp” image and the right-hand side of the “6386” image in Figure 15, which are very visible. Comparing the images of Figure 15 to those of Figure 4, especially the “Ramp” images, we see that the modified error diffusion algorithm produces at least as many gray levels as “classical-4.” In fact, as we saw in the previous section, dot overlap can be exploited to obtain more gray levels than could be obtained without it.

In *regions of slowly changing gray level*, error diffusion does not suffer from the false contouring problem, as is evident from the “Ramp” image of Figure 15. Thus there is no need

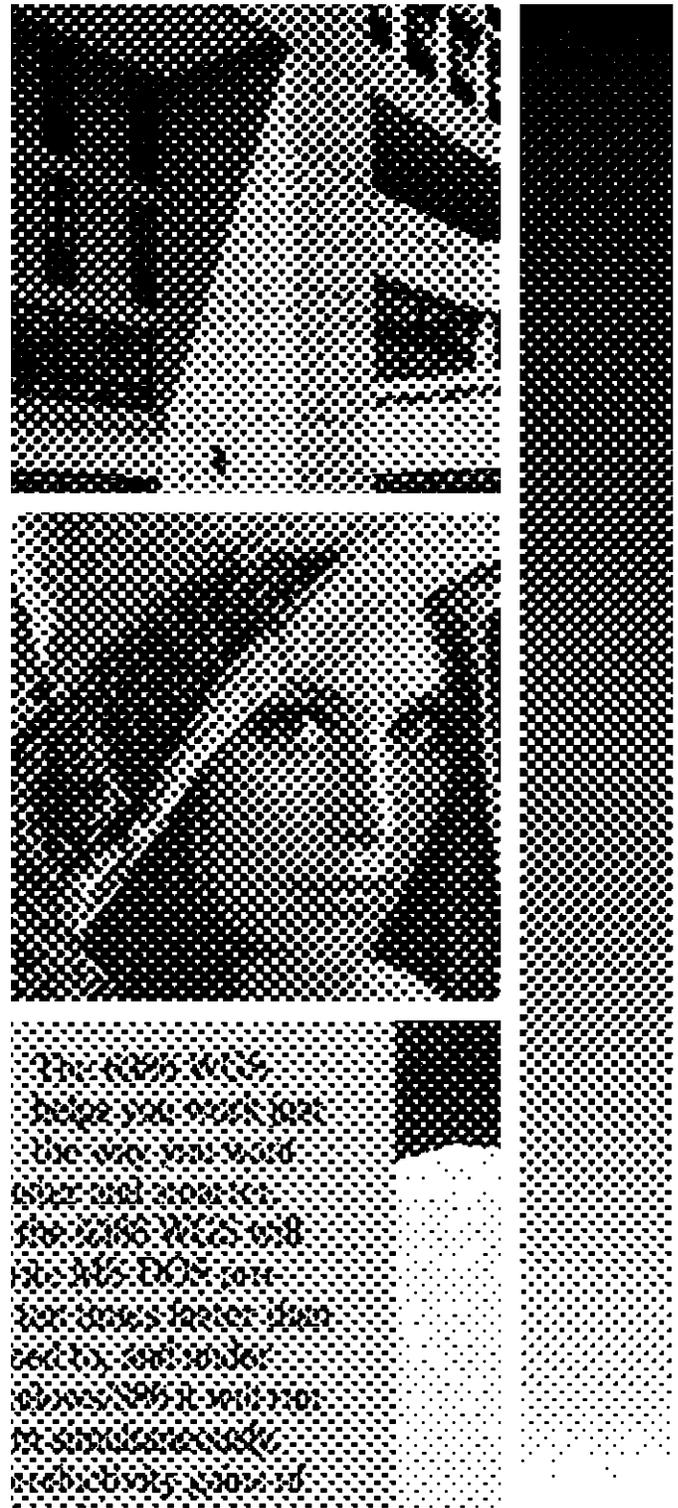


Fig. 13: Standard error diffusion at 300 dpi

to add microdither to the image. As a result the printed images of Figure 15 are less noisy than those of Figure 4 which were halftoned using “classical-4” with microdither.

A comparison of Figures 4 and 15 indicates that at 300 dpi the modified error diffusion algorithm produces images that are sharper, less noisy, and with richer gray tones than the commonly used “classical” screening technique. Actually,

<sup>11</sup>In fact, Dong proposed a multi-pass version of the modified error diffusion algorithm that eliminates any remaining bias [29], [30].

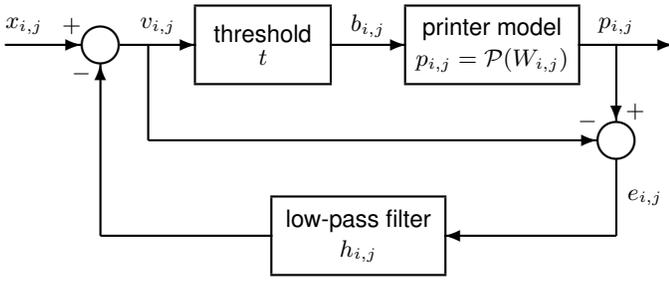


Fig. 14: Modified error diffusion

the quality of printed images obtained with the modified error diffusion on a 300 dpi printer is even better than the quality obtained with “classical” screening on a 400 dpi printer. Figure 16 shows the test images halftoned with the “classical-4” screen without microdither printed on a 400 dpi printer.

We now summarize the performance of the modified error diffusion algorithm in the four domains of Section II. In *regions of constant gray level*, the modified error diffusion produces patterns that are visually pleasant with very few low frequency artifacts. It produces at least as many gray levels as the “classical” screen. In *regions of slowly changing gray level*, the modified error diffusion algorithm does not produce false contours. In *regions of rapidly changing gray level*, error diffusion produces images that are sharper than any screening technique. In the *presence of printer distortions*, the modified error diffusion algorithm produces the correct gray scale without any sacrifice in spatial resolution compared to the standard error diffusion. Moreover, it exploits the printer distortions to produce an even richer set of gray tones. In summary, the modified error diffusion algorithm provides high quality reproductions with reasonable complexity.

#### A. Stucki's Algorithm

Stucki was the first one to propose an extension of error diffusion to account for printer dot overlap in [13], [14]. Like the modified error diffusion algorithm, the multiple-error correction computation algorithm (MECCA) accounts for dot overlap as well as quantization effects. The MECCA equations are

$$v_{i,j} = x_{i,j} - \sum_{m,n} h_{m,n} e_{i-m,j-n} \quad (17)$$

$$b_{i,j} = \begin{cases} 1, & \text{if } v_{i,j} > t \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

$$e_{m,n} = \lambda_{m,n} b_{m,n} - v_{m,n} \quad (19)$$

where

$$\lambda_{m,n} = \begin{cases} 0, & \text{if } b_{m,n} = 0 \\ \frac{A_{m,n}^{\text{eff}}}{A_{\text{pixel}}^{\text{eff}}} = \frac{A_{m,n}^{\text{dot}} - A_{m,n}^{\text{overlap}}}{A_{\text{pixel}}^{\text{eff}}}, & \text{if } b_{m,n} = 1 \end{cases} \quad (20)$$

Here  $A_{m,n}^{\text{pixel}}$  is the area of one pixel and  $A_{m,n}^{\text{eff}}$  is the effective area of white paper newly covered by the dot at pixel  $(m,n)$ , which is equal to the area of the dot  $A_{m,n}^{\text{dot}}$  minus the area of overlap  $A_{m,n}^{\text{overlap}}$  of this dot and previously placed dots. Thus  $\lambda_{m,n} = 0$  when no dot is placed at the pixel  $(m,n)$

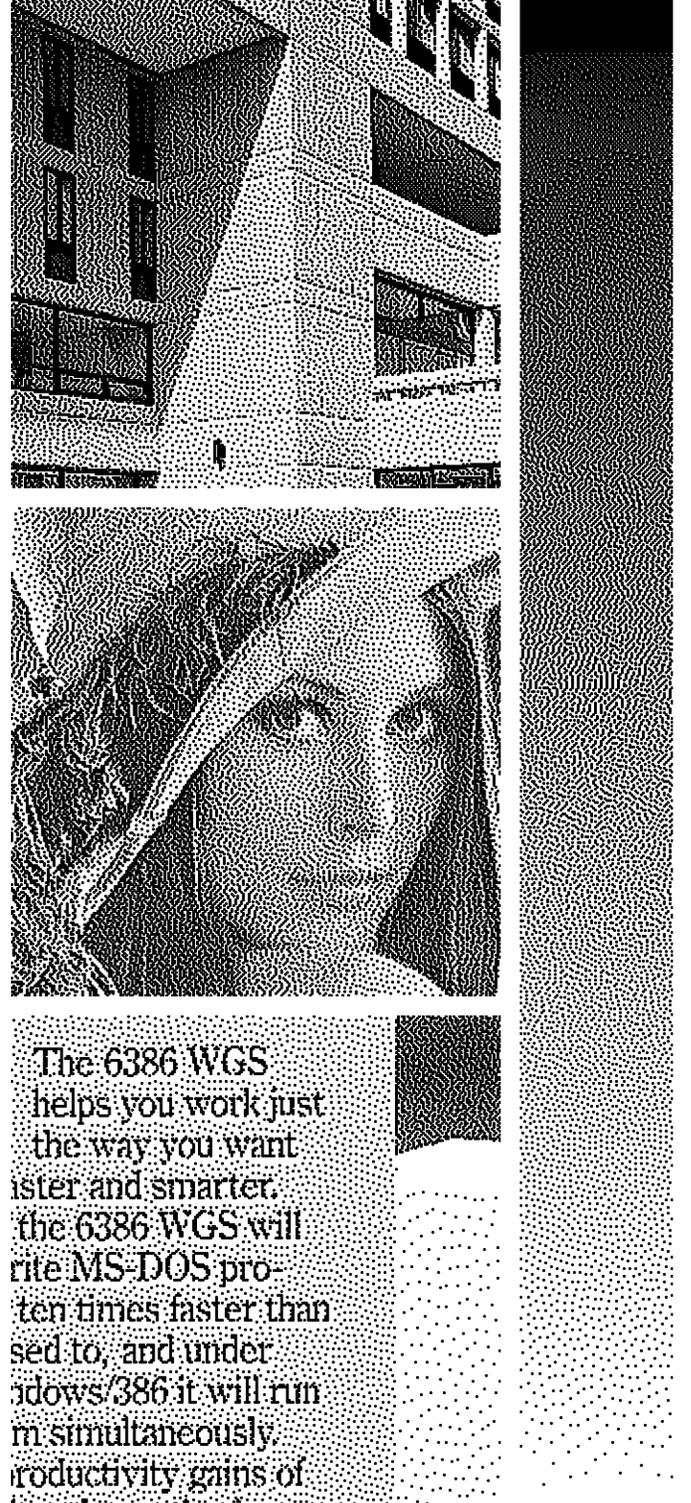


Fig. 15: Modified error diffusion at 300 dpi

(i.e.  $b_{m,n} = 0$ ), and can be as large as  $A_{m,n}^{\text{dot}}/A_{\text{pixel}}^{\text{pixel}} = \pi\rho^2/2$  when a dot is placed at the pixel  $(m,n)$  (i.e.  $b_{m,n} = 1$ ) and no previously placed dots overlap this pixel. Recall that  $\rho$  is the ratio of the actual dot radius to the ideal dot radius  $T/\sqrt{2}$ . Since  $\rho \geq 1$ ,  $\lambda_{m,n}$  can be greater than 1, in contrast to  $p_{m,n}$  of Eq. (16) which is always between 0 and 1. More importantly, Stucki's algorithm accounts only for the newly



Fig. 16: Classical screen (no microdither) at 400 dpi

placed ink, which may be outside the pixel boundaries, and does not account for previously placed ink, which may be inside the pixel boundaries. In other words, Stucki's algorithm accounts for ink at the wrong place. This effectively shifts the edges and, as we will see in the following examples, results in a loss of sharpness when compared to the proposed algorithm which accounts for the (actually present) ink within the pixel

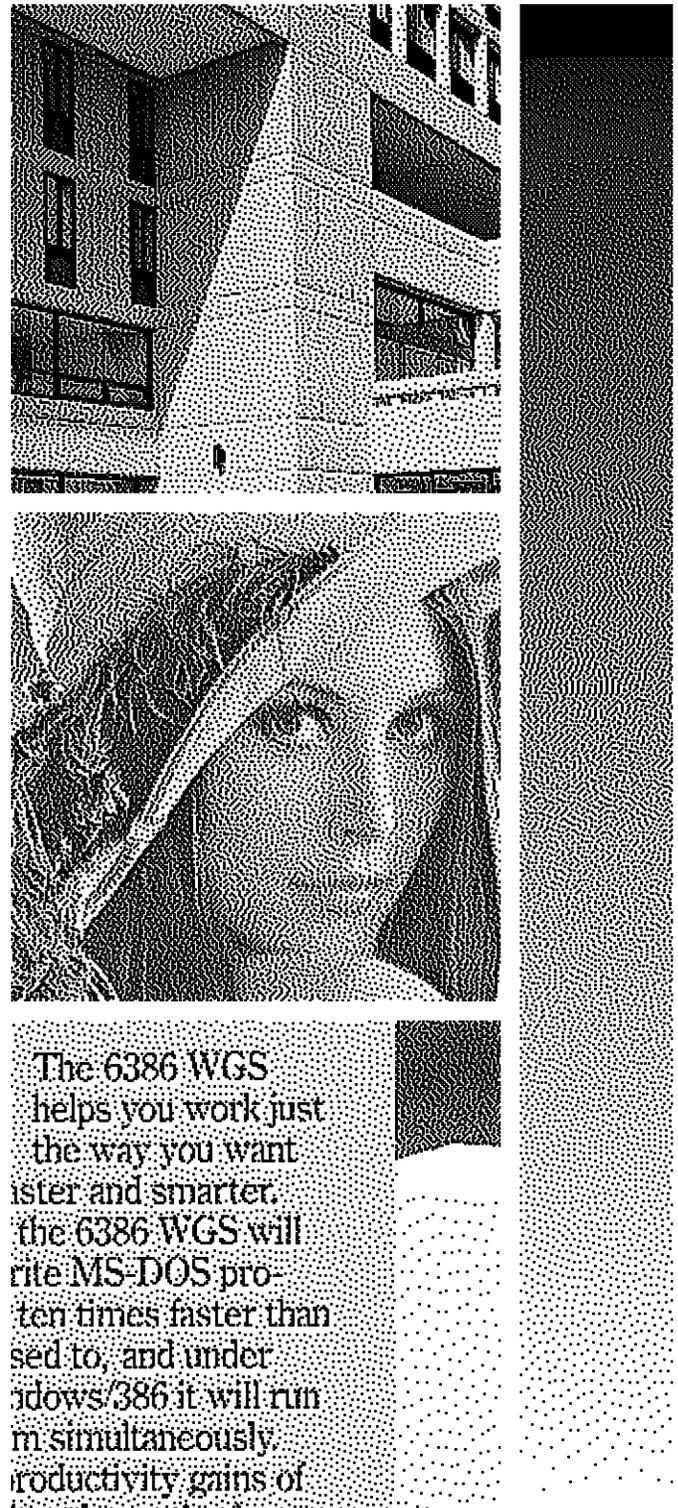


Fig. 17: Stucki's algorithm at 300 dpi

boundaries.

Notice that in Stucki's approach the error is computed only once, while in the proposed algorithm the "past" errors are updated every time a new dot overlaps the corresponding pixel. Thus, in Eqs. (13) and (15) the error depends on the "instant"  $(i, j)$ , while in Eqs. (17) and (19) it does not. This reduces the number of computations, an important consideration when

Stucki's work was first published, but less so now. Stucki's goal was to find an efficient way to account for dot overlap in error diffusion, while our goal is to obtain an accurate printer model that can be used to account for dot overlap and other distortions, not only in error diffusion, but in any halftoning technique (e.g. the least-squares approach [23], [24]), in order to produce the best quality pictures.

Figure 17 shows the results of Stucki's MECCA algorithm. A comparison with Figure 15 shows that the proposed algorithm produces sharper images than MECCA. The difference is particularly evident in the details of the "Lena" image (eyes, rim of the hat, feathers), in the concrete wall of the "Bank" image, and in the lettering of the "6386" image. We also observe that the proposed algorithm produces less objectionable low frequency artifacts, as can be seen in the "Ramp" image and in the smooth areas of the "Bank" and "Lena" images. Observe that MECCA tends to produce patterns with stronger diagonal streaks than the proposed algorithm. Both differences are a result of the different ways of accounting for the dot-overlap error in the two algorithms, as we discussed above. Finally, both algorithms account successfully for printer distortions.

In summary, Stucki's algorithm is more efficient computationally, while the proposed algorithm produces images that are sharper and have more pleasant halftone patterns than Stucki's method.

## V. CONCLUSIONS

We presented printer models that can be used with halftoning techniques to produce high quality images using standard laser printers. We proposed a general framework for printer models, and found a specific model for laser printers. The circular dot-overlap model is very simple but captures the most important printer distortions. It is characterized by one parameter that can be easily adapted to specific printers.

We also presented a new halftoning technique that relies on the printer models we developed. It is an extension of error diffusion and exploits the printer distortions to increase both the spatial and gray-scale resolution of the printed images. Our experiments indicate that the modified error diffusion algorithm offers a substantial improvement over conventional halftoning techniques. It provides high quality reproductions with reasonable complexity.

## ACKNOWLEDGEMENTS

We wish to thank Tom Michel and Richard Drechsler for their help with the laser printers and postscript.

## REFERENCES

- [1] D. L. Neuhoff and T. N. Pappas, "Perceptual coding of images for halftone display," *IEEE Trans. Image Processing*, vol. 3, no. 7, pp. 341–354, July 1994.
- [2] —, "Perceptual coding of images for halftone display," in *Proc. ICASSP-91*, (Toronto, Canada), May 1991, pp. 2797–2800, vol. 4.
- [3] CCITT Draft Recommendation T.82, ISO/IEC Draft International Standard 11544, WG9-S1R5.1., "Coded representation of picture and audio information – progressive bi-level image compression," April 3, 1992.
- [4] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. ICASSP-89*, (Glasgow, Scotland), May 1989, pp. 1945–1948, vol. 3.
- [5] G. Wallace, W. B. Pennebaker, and J. L. Mitchell, "JPEG technical specification, revision 5.2," May 10, 1990.
- [6] P. G. Roetling and T. M. Holladay, "Tone reproduction and screen design for pictorial electrographic printing," *Journal of Applied Phot. Eng.*, vol. 15, no. 4, pp. 179–182, 1979.
- [7] H. Sonnenberg, "Laser-scanning parameters and latitudes in laser xerography," *Applied Optics*, vol. 21, no. 10, pp. 1745–1751, May 1982.
- [8] R. W. Floyd and L. Steinberg, "An adaptive algorithm for spatial grey scale," in *Proc. SID*, 1976, pp. 75–77, vol. 17/2.
- [9] T. N. Pappas and D. L. Neuhoff, "Model-based halftoning," in *Proc. SPIE, Human Vision, Visual Proc., and Digital Display II* (San Jose, CA), Feb. 1991, pp. 244–255, vol. 1453.
- [10] R. Ulichney, *Digital Halftoning*. Cambridge, MA: The MIT Press, 1987.
- [11] A. C. Naiman, "CRT spatial non-linearities and luminance linearization," in *Raster Imaging and Digital Typography* (R. Morris, Ed.). Cambridge, England: Cambridge Univ. Press, Oct. 1991, proceedings of the RIDT91 Int. Conf., (Boston, MA).
- [12] A. C. Naiman and W. Makous, "Spatial non-linearities of grayscale CRT pixels," in *Proc. SPIE, Human Vision, Visual Proc., and Digital Display III*, (San Jose, CA), Feb. 1992, pp. 41–56, vol. 1666.
- [13] P. Stucki, "MECCA – a multiple-error correcting computation algorithm for bilevel image hardcopy reproduction," IBM Research Laboratory, Zurich, Switzerland, Research Report RZ1060, 1981.
- [14] —, "Advances in digital image processing for document reproduction," in *VLSI Engineering*, T. L. Kunii, Ed. Tokyo: Springer-Verlag, 1984, pp. 256–302.
- [15] D. Anastassiou, "Error diffusion coding for A/D conversion," *IEEE Trans. Circuits Syst.*, vol. CAS-36, no. 9, pp. 1175–1186, Sept. 1989.
- [16] J. P. Allebach, "Binary display of images when spot size exceeds step size," *Applied Optics*, vol. 19, no. 15, pp. 2513–2519, Aug. 1980.
- [17] R. W. Pryor, G. M. Cinque, and A. Rubinstein, "Bilevel displays – a new approach," in *Proc. SID*, 1978, pp. 127–131, vol. 19.
- [18] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1989.
- [19] J. P. Allebach, "Visual model-based algorithms for halftoning images," in *Proc. SPIE, Image Quality*, 1981, pp. 151–158, vol. 310.
- [20] A. Zakhori, S. Lin, and F. Eskafi, "A new class of B/W and color halftoning algorithms," in *Proc. ICASSP-91*, (Toronto, Canada), May 1991, pp. 2801–2804, vol. 4.
- [21] M. Analoui and J. P. Allebach, "Model based halftoning using direct binary search," in *Proc. SPIE, Human Vision, Visual Proc., and Digital Display III* (San Jose, CA), Feb. 1992, pp. 96–108, vol. 1666.
- [22] J. B. Mulligan and A. J. Ahumada, Jr., "Principled halftoning based on models of human vision," in *Proc. SPIE, Human Vision, Visual Proc., and Digital Display III* (San Jose, CA), Feb. 1992, pp. 109–121, vol. 1666.
- [23] T. N. Pappas and D. L. Neuhoff, "Least-squares model-based halftoning," in *Proc. SPIE, Human Vision, Visual Proc., and Digital Display III* (San Jose, CA), Feb. 1992, pp. 165–176, vol. 1666.
- [24] D. L. Neuhoff, T. N. Pappas, and N. Seshadri, "One-dimensional least-squares model-based halftoning," in *Proc. ICASSP-92*, San Francisco, CA, Mar. 1992, pp. 189–192, vol. 3.
- [25] F. W. Campbell, J. J. Kulikowski, and J. Levinson, "The effect of orientation on the visual resolution of gratings," *J. Physiology (London)*, vol. 187, pp. 427–436, 1966.
- [26] W. F. Schreiber, *Fundamentals of Electronic Imaging Systems*. Springer-Verlag, 1986.
- [27] L. G. Roberts, "Picture coding using pseudo-random noise," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 145–154, 1962.
- [28] J. P. Allebach, "Reconstruction of continuous-tone from halftone by projections onto convex sets," in *Proc. 1988 Int. Conf. Adv. in Comm. and Control Sys.* (Baton Rouge), Oct. 1988, pp. 469–478, vol. 1.
- [29] C.-K. Dong, "Measurement of printer parameters for model-based halftoning," B.S./M.S. Thesis, MIT, May 1992.
- [30] T. N. Pappas, C.-K. Dong, and D. L. Neuhoff, "Measurement of printer parameters for model-based halftoning," *Journal of Electronic Imaging*, vol. 2, no. 3, pp. 193–204, July 1993.
- [31] J. F. Jarvis, C. N. Judice, and W. H. Ninke, "A survey of techniques for the display of continuous-tone pictures on bilevel displays," *Comp. Graphics and Image Proc.*, vol. 5, pp. 13–40, 1976.
- [32] C. J. Rosenberg, "Measurement-based verification of an electrophotographic printer dot model for halftone algorithm correction," in *IS&T's 8th Int. Cong. Adv. Non-Impact Printing Techn.* (Williamsburg, VA), Oct. 25–30, 1992, pp. 286–291.

PLACE  
PHOTO  
HERE

**Thrasylvoulos N. Pappas** (M'87) received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1979, 1982, and 1987, respectively.

He has been with the Signal Processing Research Department at AT&T Bell Laboratories, Murray Hill, NJ, since 1987. His research interests are in image processing and computer vision. Recently, he has been working on image halftoning and compression based on models of display devices and the

human visual system.

PLACE  
PHOTO  
HERE

**David L. Neuhoff** (S'72-M'74-SM'83-F'94) was born in Rockville Centre, New York on August 18, 1948. He received a B.S.E. from Cornell University in 1970 and an M.S. and Ph.D. in Electrical Engineering from Stanford University in 1972 and 1974, respectively.

In 1974, he joined the University of Michigan, Ann Arbor, Michigan, where he is currently a Professor of Electrical Engineering and Computer Science. From 1984-1989 he was Associate Chairman of the Systems Science and Engineering Division of the Department. From Sept. 1989 through June 1990 he was on leave at AT&T Bell Laboratories, Murray Hill, NJ, in the Signal Processing Dept.. His research and teaching interests are in communications and information theory, especially data compression, quantization, Shannon theory, coding for magnetic recording, halftoning and constrained coding.

Dr. Neuhoff is a member of Eta Kappa Nu, Tau Beta Pi and Sigma Xi, and a Senior Member of the IEEE. He is an Associate Editor for Source Coding for the IEEE Transactions on Information Theory. He served on the Board of Governors of the IEEE Information Theory Society, 1988-90; he was Chairman of the IEEE Southeastern Michigan Chapter of Division I in 1978; and he was Co-Chairman of the 1986 IEEE International Symposium on Information Theory in Ann Arbor.