Chapter 1

Introduction

Digital halftoning refers to the process of converting a continuous-tone image or photograph into a pattern of black and white picture elements (Fig. 1.1) for reproduction by a binary display device such as an ink jet printer, which can only choose to print or not print dots. The human visual, acting like a low-pass filter, blurs these printed and not printed dots together to create the illusion of continuous shades of gray. Depending on the specific manner in which dots are distributed, a given display device can produce varying degrees of image fidelity with more or less graininess. According to the human visual system, randomly arranged and isolated dots, properly distributed, should produce images with the highest quality, maintaining sharp edges and other fine details. But at the same time, certain display and printing devices are incapable of reproducing isolated dots consistently from dot to dot and, consequently, introduce printing artifacts that greatly degrade those same details that the dot distribution is designed to preserve. For this reason, many printing devices produce periodic patterns of clustered dots, which are easier to produce consistently across the printed page. So in studying halftoning, the principal goal is to determine what is the optimal distribution of dots for that device and then to produce these patterns in a computationally efficient manner.

When analog halftoning was perfected in 1880, continuous-tone, monochrome photographs were reproduced as line drawings authored by highly skilled craftsmen, usually on scratch board. But with halftoning, newspapers and magazines could cheaply reproduce photographs in their publications, making photography a lucrative industry and resulting in a technical revolution in photographic equipment. In terms of photo-lithography, the early halftoning processes (Fig. 1.2) involved

Figure 1.1: Gray-scale image reproduced as an analog halftone.

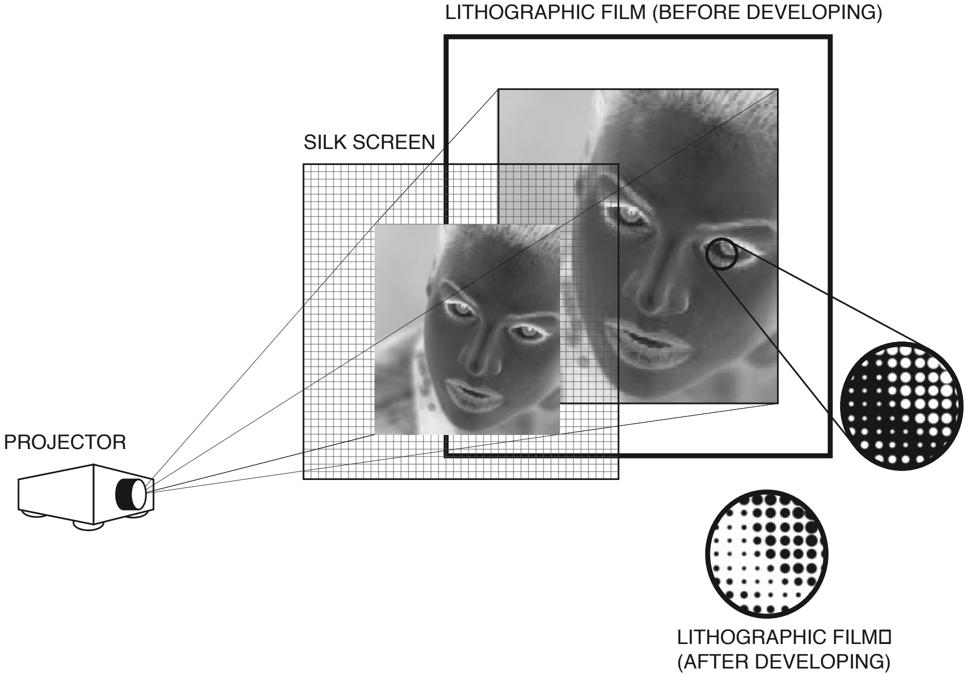


Figure 1.2: The binary halftone obtained by projecting the negative of the original continuous-tone image through a fine silk screen.

projecting light from the negative of a continuous-tone photograph through a mesh screen, such as finely woven silk, onto a photo-sensitive plate. Bright light, as it passes through a pin-hole opening in the silk screen, would form a large, round spot on the plate. Dim light would form a small spot. Light sensitive chemicals coating the plate would then form insoluble dots that varied in size according to the tones of the original photographs. After processing, the plate would have dots where ink was to be printed raised slightly above the rest of the plate.

Later versions of the halftoning process employed screens made of glass that were coated, on one side, by an opaque substance [116]. A mesh of parallel and equidistant lines were scratched in the opaque surface. A second mesh of parallel and equidistant lines were then scratched in the opaque surface running perpendicular to the original set. Screens would then differ in the number of lines per inch that had been scratched. While finer screens created better spatial resolutions (detail), the quality of the printing press would limit the fineness of the mesh.

Later still, the glass plate mesh was replaced altogether with

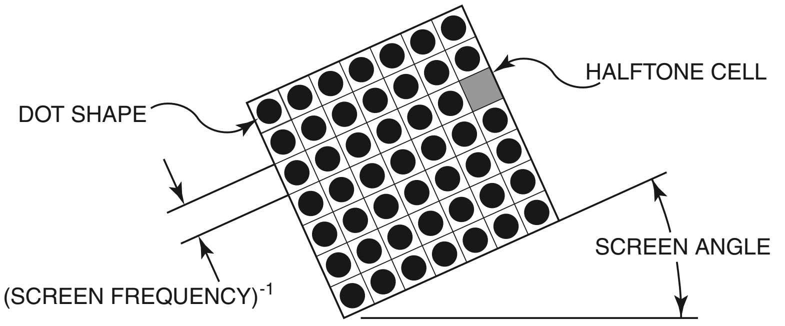


Figure 1.3: The screen frequency, dot shape, and screen angle for an analog halftone pattern.

a flexible piece of processed film, placed directly in contact with the unexposed lithographic film [28]. This *contact screen* had direct control of the dot structure (Fig. 1.3) being able to control the screen frequency (the number of lines per inch), the dot shape (the shape of dots as they increase in size from light to dark), and the screen angle (the orientation of lines relative to the positive horizontal axis).

1.1 AM Digital Halftoning

Today, printing is a far more advanced process with the introduction of non-impact printing technologies and the emergence of desktop publishing. Brought on by advancements in the digital computer [28], the photo-mechanical screening process introduced in 1880 has, in many instances, been replaced by digital imagesetters. In some instances, printing is no longer binary as continuous-tone dye-sublimation printers are now readily available but due to their speed and material requirements (special papers and inks), have not reached the wide-spread acceptance of color ink jet or electro-photographic (laser) printers, although dye-sublimation has made a major come back with a consumer market for 4 inch by 6 inch photograph printers.

In these digital printers, the halftoning process of projecting a continuous-tone original through a halftone screen has been replaced with a raster image processor (RIP) that converts each pixel of the original image from an intermediate tone directly into a binary dot based upon a pixel-by-pixel comparison of the original image with an array of thresholds (Fig. 1.4). Pixels of the original with intensities greater than their corresponding threshold are turned "on" (printed)

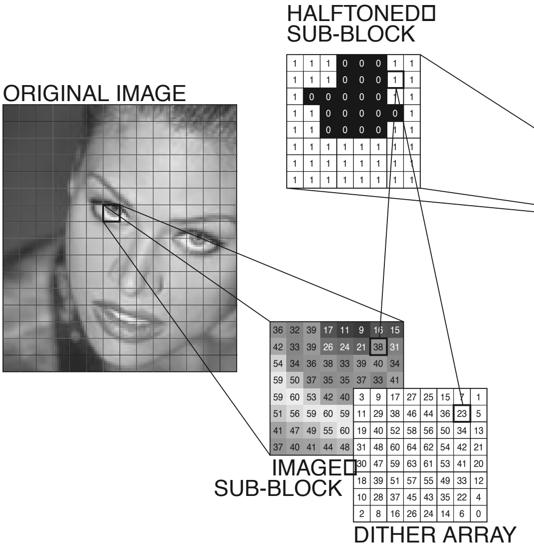


Figure 1.4: Digital AM halftoning.

in the final halftoned image while pixels less than their corresponding thresholds are turned "off". For large images, the threshold array is tiled end-to-end until all pixels of the original have a corresponding threshold.

When first introduced, RIPs imitated the halftone patterns of contact screens by employing clustered-dot ordered dithering, where the threshold array is small (8×8, 12×12, or 16×16) and is composed of consecutive thresholds arranged along a spiral path radiating outward from the array's center (Fig. 1.5). These arrangements of thresholds result in a single cluster of "on" pixels centered within each tile or *cell*, forming a regular grid of round dots that vary in size according to tone. These techniques are commonly referred to as *amplitude modulated* or AM digital halftoning due to their modulating of the size of printed dots. Like contact screens, the resulting patterns vary in their screen frequency, dot shape, and screen angle.

1.2 FM Digital Halftoning

Due to freedoms afforded by digital printers, the idea of printing isolated pixels in an effort to minimize halftone visibility (the visibility of the individual dots to a human viewer) emerged as an alternative to clustered-dot dithering. By maintaining the size of printed dots for all

3	9	17	27	25	15	7	1
11	29	38	46	44	36	23	5
19	40	52	58	56	50	34	13
31	48	60	64	62	54	42	21
30	47	59	63	61	53	41	20
18	39	51	57	55	49	33	12
10	28	37	45	43	35	22	4
2	8	16	26	24	14	6	0

Figure 1.5: An 8×8 dither array (65 gray-levels).

gray-levels as individual pixels, new dispersed-dot halftoning techniques varied, according to tone, the spacing between printed dots, earning the name *frequency modulated* or FM halftoning. Early FM halftoning techniques were proposed by Bayer [11] and Bryngdahl [15] and produced an ordered arrangement of isolated dots. These techniques, like AM halftoning schemes, quantized each pixel independently of its neighbors (point process) according to a dither array but with consecutive thresholds dispersed as much as possible. The problem associated with these early FM techniques is that, as in the case of Bayer's dither array (Fig. 1.6), resulting halftoned images (Fig. 1.7) suffered from a periodic structure that added an unnatural appearance [123].

For a far better approach to FM halftoning, Floyd and Steinberg [40] proposed the revolutionary error-diffusion algorithm (shown in Fig. 1.8 and covered in Chapter 5), an adaptive technique that quantized each pixel according to a statistical analysis of an input pixel and its neighbors (neighborhood process), leading to a stochastic arrangement of printed dots. While this neighborhood process had higher computational complexity, the resulting patterns had apparent spatial resolutions much higher than those achieved by clustered dots (Fig. 1.9); furthermore, as a stochastic patterning of dots, the patterns eliminated the occurrence of the moiré that was produced by the superimposing of two or more regular patterns.

By using FM halftoning schemes, printers maximize their apparent spatial resolution and are relieved of the strict tolerances on screen angles and screen registration. They can also use more and more colors to produce larger color gamuts (the set of achievable colors that can be produced by the printer) [76]. Notably, though, with its associated advantages, FM halftoning has, with few exceptions, only been employed

0	58	14	54	3	57	13	53
32	16	46	30	35	19	45	29
8	48	4	62	11	51	7	61
40	24	36	20	43	27	39	23
2	56	12	52	1	59	15	55
34	18	44	28	33	17	47	31
10	50	6	60	9	49	5	63
42	26	38	22	41	25	37	21

Figure 1.6: An 8×8 Bayer's dither array (65 gray-levels).

in ink jet printers. The problem is the increased scrutiny placed on the printer's ability to print small, isolated dots.

Noting Fig. 1.10, the ideal display produces dots that completely cover the sample area associated with a given pixel without overlapping neighboring pixels' sample areas. By printing all pixels, perfect black can be obtained. In a real printing device, individual printed dots are round and in order to produce perfect black, must be large enough as to cover the entire sample area (Fig. 1.11). By overlapping neighboring sample areas, though, the resulting tone is darker than the fraction of all pixels that are printed. Assuming that dots are printed consistently (small variation in size and shape from printed dot to printed dot), this distortion in tone can be corrected by adjusting the intensity level of the input image before halftoning. The amount of compensation depends on the arrangement of printed dots with dispersed-dot (FM) patterns requiring greater degrees of correction than clustered (AM). Ink jet printers are such a device that prints (approximately) round dots that overlap neighboring pixels (Fig. 1.12). Being able to compensate for distortions introduced by the printing process for any arrangement of dots, ink jet printers can enjoy the benefits associated with FM halftoning.

In the electro-photographic printing process, the size and shape of dots varies greatly from printed dot to printed dot (Fig. 1.13) and this variation can only be minimized when dots are grouped together to form clusters. Images printed using isolated dots tend to show severe tonal distortion and exhibit a great deal of variation in tone across the printed page, as illustrated in Fig. 1.14. For this reason, unreliable printing devices such as lithographic presses and laser printers continue to use AM halftoning schemes.

Figure 1.7: Gray-scale image halftoned using Bayer's dither printed at 150 dpi.

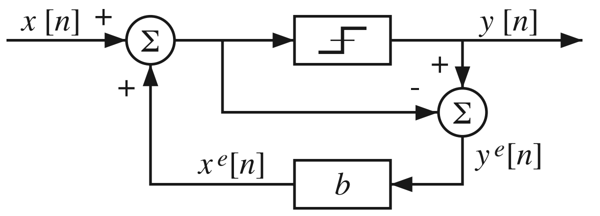


Figure 1.8: The error-diffusion algorithm.

1.3 AM-FM Hybrids

Now as printers began achieving print resolutions greater than 1,200 dpi, the limits of FM halftoning were also being reached, just as AM's was in 1990 before the introduction of low-cost color. Researchers, therefore, begin to look at AM-FM hybrids which produce dot clusters that vary, according to tone, in both their size and spacing [6, 66, 75, 94, 110, 125, 126, 128] (Fig. 1.15). When considering the reproduction of monochrome images, AM-FM hybrids are, in general, capable of producing patterns with lower visibility (higher spatial resolution) compared to AM and, if stochastic, do so without a periodic structure adding an artificial texture to the printed image. With some amount of clustering, these halftones are easier to print reliably and with little variation in the resulting tone. In the reproduction of color, stochastic hybrids also maintain the same freedom from periodic moiré associated with FM halftoning.

1.3.1 Why *Modern* Digital Halftoning?

Since its publishing in 1987, Ulichney's *Digital Halftoning* [122] has proven to be the single most unifying book about halftoning research. While a large portion of that book is devoted to AM halftoning, its most significant contribution is its description of error-diffusion as a generator of *blue-noise*, visually pleasing halftone patterns composed of randomly placed, isolated dots. Ulichney used the term "blue" as a reference to the spectral content of the random patterns as being composed exclusively of high-frequency spectral components just as blue light is composed exclusively of the high-frequency spectral components of white light. Ulichney further showed that the average distance between same colored dots was directly related to the gray-level of the image before halftoning. In summary, Ulichney's work tells us when

Figure 1.9: Gray-scale image halftoned using Floyd's and Steinberg's error-diffusion algorithm.

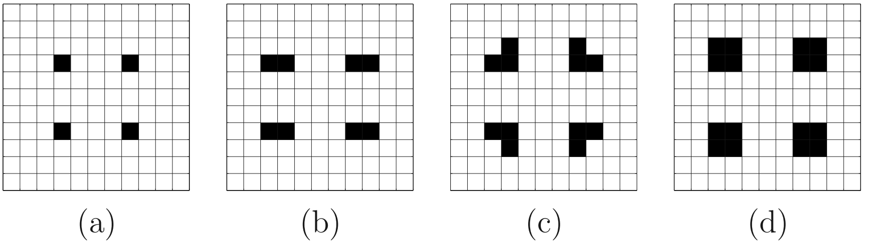


Figure 1.10: Clusters of (a) one, (b) two, (c) three, and (d) four printed dots from an ideal printer with the solid lines indicating the border between neighboring output pixels.

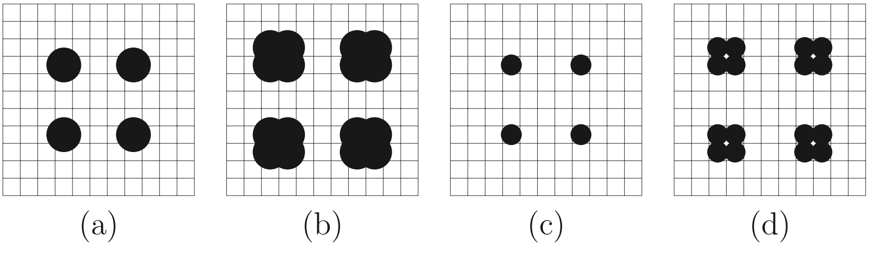


Figure 1.11: Clusters of (a,c) one and (b,d) four printed round dots where the dots of (a) and (b) cover the entire sample area while the dots of (c) and (d) do not cover the corners.

one FM halftone is better, visually, than another. It is thanks to the improved image quality produced by blue-noise that ink jet printers are so prevalent today as these printers have directly benefited from the improved spatial resolution afforded by randomly placed, isolated dots.

In the decade that followed the publishing of *Digital Halftoning*, many technological advancements were documented directly addressing the creation of blue-noise. The most notable of which may be introduction of blue-noise dither arrays that convert the pixels of continuous-tone images to binary through a pixel-wise comparison of the original pixel with a corresponding pixel in the dither array. A process identical to AM halftoning with the exception that these blue-noise dither arrays are typically much larger (128×128 or 256×256) as seen in Fig. 1.16. Being a point process operation, blue-noise dither array halfton-

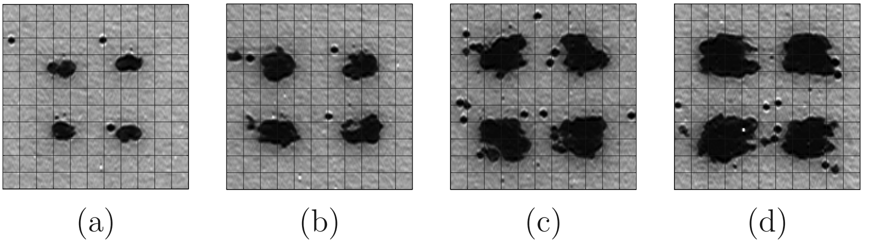


Figure 1.12: Clusters of (a) one, (b) two, (c) three, and (d) four printed dots printed on ink jet printer.

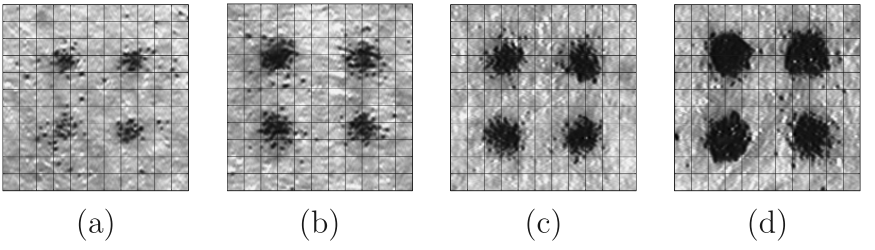
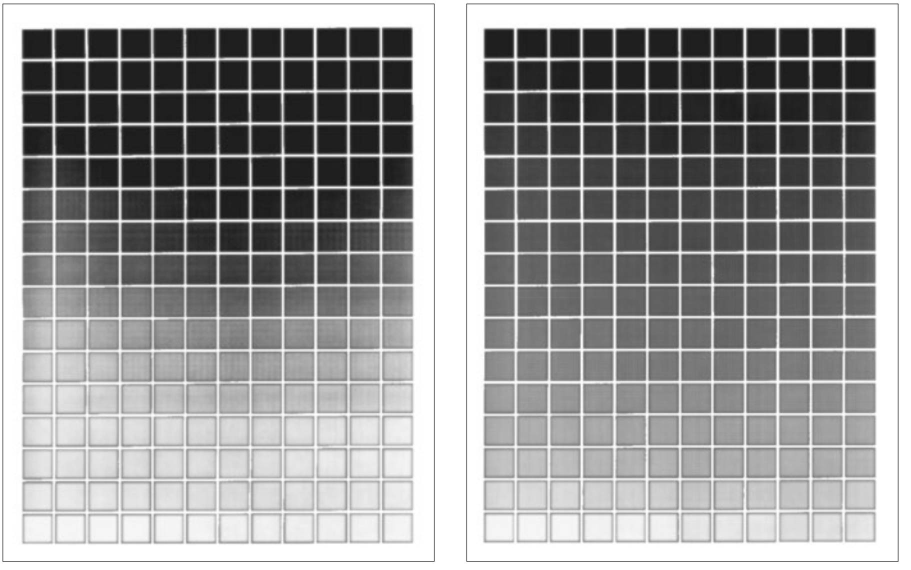


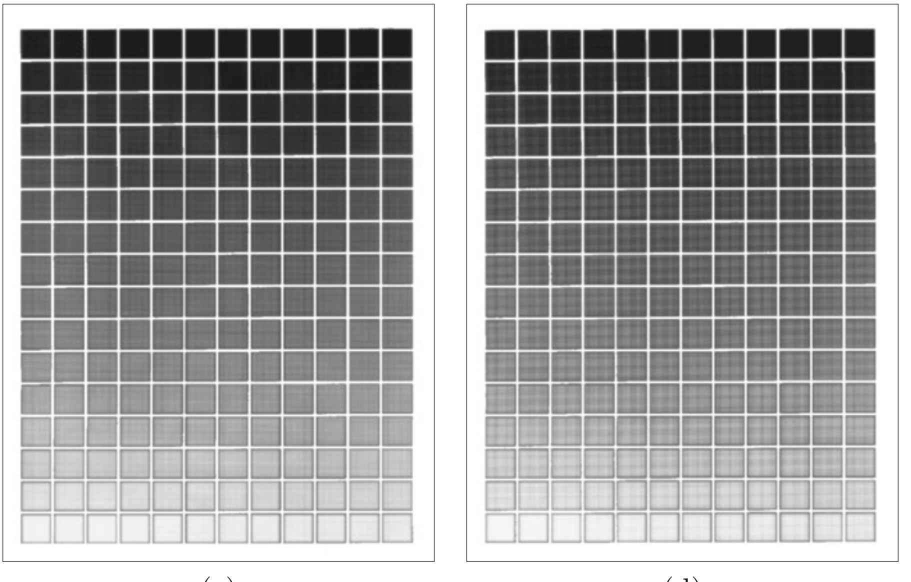
Figure 1.13: Clusters of (a) one, (b) two, (c) three, and (d) four printed dots printed on a laser printer.

ing is far less computationally complex than error-diffusion, making it a viable technique for commercial printing applications whose image sizes are orders of magnitude larger than those associated with desktop printers.

Now the error-diffusion algorithm, itself, has also undergone some major improvements over the past two decade [58]. One of the more profound improvements has been the introduction of threshold modulation, where the threshold used to quantize an input pixel to either one or zero is varied in a given fashion. An early approach to threshold modulation was proposed by Ulichney [122], who suggested adding white-noise to the threshold in order to break up worm patterns and periodic textures (Fig. 1.17). This approach was later shown by Knox [55] to be equivalent to adding low-level white-noise to the original image before halftoning. Another approach to threshold perturbation was later proposed by Eschbach and Knox [34], who suggested varying the threshold by a scalar multiple of the current input pixel. This oper-



(a) (b)



(c) (d)

Figure 1.14: Gray-scale ramps produced by a laser printer using (a) FM halftoning and (b-d) AM halftoning with halftone cells of size 8×8, 12×12, and 16×16 pixels respectively.

Figure 1.15: Gray-scale image halftoned using an AM-FM hybrid.

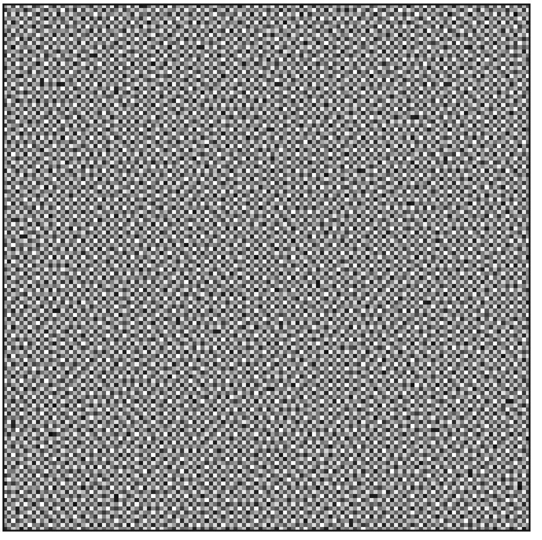


Figure 1.16: A 128 × 128 blue-noise mask.

ation sharpened the resulting halftone pattern, thereby eliminating the need for applying an edge sharpening filter to the input image prior to halftoning.

A completely different approach to halftoning, proposed since Ulichney's book, is Analoui's and Allebach's [5] direct binary search (DBS) algorithm. This algorithm iteratively swaps and toggles the binary halftone pattern according to the error between the original continuous-tone image and the binary halftone image as defined by a model of the human visual system. Typically, the human visual system is modeled as a low-pass filter, and by employing it into the halftoning algorithm, causes same colored dots to be spread as far apart as possible, creating a power spectrum that models blue-noise. While far more computationally complex than even error-diffusion, the DBS algorithm's complexity has greatly diminished since first being proposed and may one day challenge error-diffusion.

Looking back at error-diffusion, another innovation or particular importance to halftoning is Pappas' and Neuhoff's [96] model-based halftoning, which employs error-diffusion in the traditional sense but with the printed dot modeled by a round circle, Fig. 1.11, that overlaps neighboring pixels. This phenomenon of overlapping neighboring pixels, in the printed output, is generally referred to as dot-gain and is typically corrected for by changing the input intensity levels prior to halftoning. Pappas and Neuhoff, by employing a dot model into

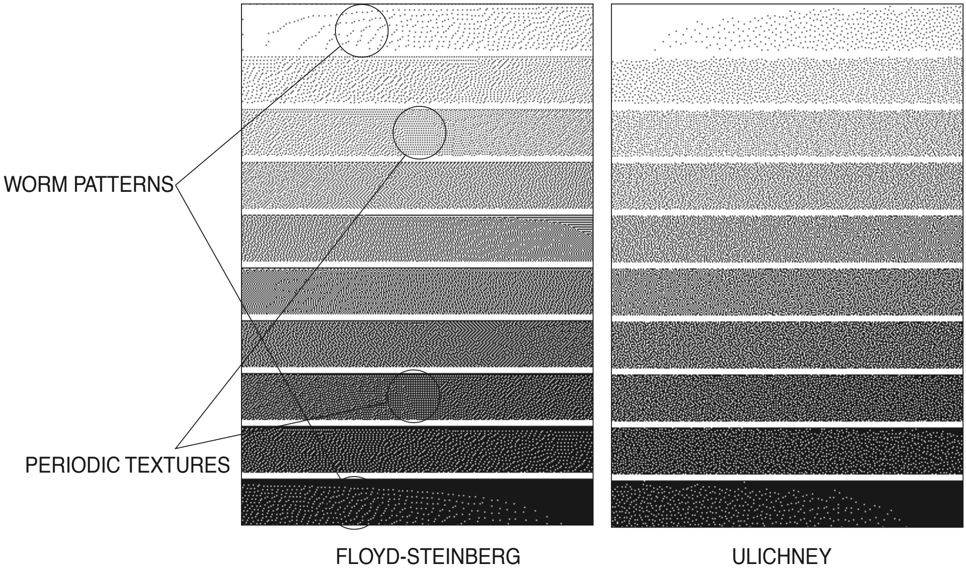


Figure 1.17: A comparison of Floyd's and Steinberg's original error-diffusion algorithm with Ulichney's threshold modulated error-diffusion algorithm.

the halftoning algorithm, directly addressed the problems of printer distortion, where the printer's ability to print isolated dots reliably (with little variation in the size and/or shape of dots from printed dot to printed dot) affects and sometimes dictates which arrangements of dots, whether dispersed as in FM or clustered as in AM, can or cannot be used. This brings about another very important aspect of halftoning that has surfaced as a direct result of the blue-noise model, that is that even with ten years of technological improvements, blue-noise has made little inroads into electro-photographic printing or digital offset printing.

The reason is clear being that blue-noise and FM halftones in general are expensive, if not impossible, to produce consistently across a page, from page to page, or from day to day or even hour to hour. Figure 1.18 illustrates the resulting variation in tone across a page for an error-diffused halftone representing gray-level $\frac{7}{10}$ produced by a laser printer set at 1,200 dpi. In this figure, the average variation in tone along the vertical axis is plot along side a picture of the page, and the average variation in tone along horizontal is plotted below the picture. In order to be reasonably efficient, a process like Pappas' and Neuhoff's requires that the statistical characteristics of the printed dot do not change either over time or across the page or according to the location of other dots. So a real challenge has come to face researchers as they try to improve the image quality in these types of printers. What they are finding is that halftoning algorithms that cluster same color pixels together, in a random fashion, hold the key by creating patterns that are easier to produce consistently from page to page and by creating color halftone patterns without moiré.

When first published in 2001, the goal of this book was to introduce, in a concise and unifying framework, these modern approaches to digital halftoning. The new model introduced was that of *green-noise*, a statistical model that described the spatial and spectral characteristics of visually pleasing dither patterns composed of a random arrangement of clustered dots that vary with gray-level in both their spacing apart and their size and shape. The term "green" referred to the mid-frequency only content of corresponding halftone patterns as green light is the mid-frequency component of white. What made green-noise a unifying framework is that it was tunable, describing the ideal characteristics for a range of cluster sizes where small clusters were reserved for reliable printing devices that showed small variation in the size and shape of printed dots and large clusters were reserved for unreliable printing devices.

In addition to describing the traits that would determine when one green-noise halftone pattern was better than another, we addressed the problem of computational efficiency by introducing the green-noise mask [63], a variation on blue-noise dither arrays for producing green-noise. Like the green-noise model, these new masks were also tunable, making them applicable to a wide range of printing devices. Having blue-noise as a limiting case of green-noise, the green-noise mask construction procedure described could even be applied to constructing the previously invented blue-noise arrays. As a last component to green-noise, the first edition of this book addressed green-noise's application to color printing where the primary task is to reproduce color as accurately as possible without introducing visible moiré patterns. While color printers were once very expensive and therefore rare, blue-noise has made color printing common place in both homes and offices. For any halftoning research to be noteworthy, it must address color printing, and we showed how green-noise was applied in color printers without introducing moiré.

Now in presenting a second edition of this book, we wanted to broaden the range of topics addressed by including a new chapter focus-

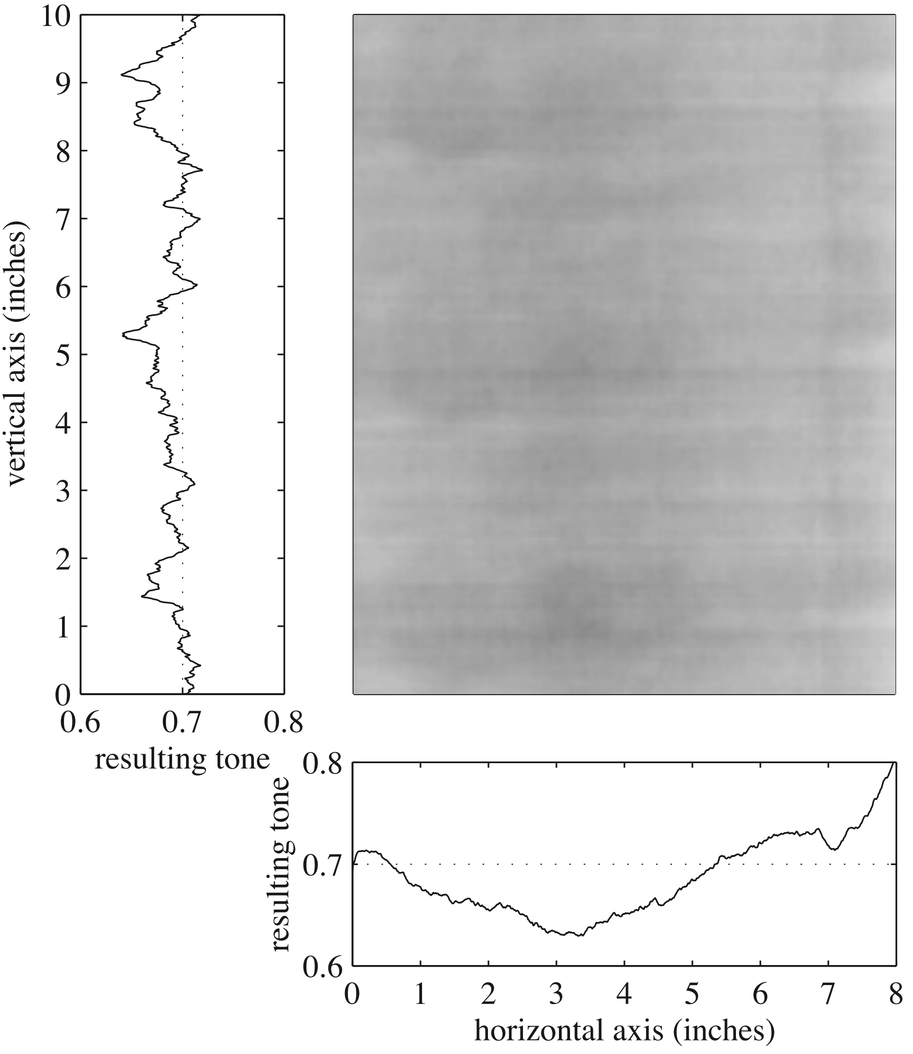


Figure 1.18: The resulting printed page produced by a laser printer at 1,200 dpi where the gray-level $\frac{7}{10}$ is produced using error-diffusion. The average variations in tone along the horizontal axis and the vertical axis are also shown, plotted alongside and below the picture of the printed page.

ing on the prior art of AM dither array construction, defining in greater detail the benefits and challenges associated with irrational screen angles. We have also included a thorough review of the original blue-noise model for hexagonal sampling grids, where we show that the original model was incorrect in its assumptions about distributing minority pixels in areas close to 50% gray. Specifically, it was determined that near 50% gray, the sampling grid was restricting the placement of printed dots, which was injecting periodic textures into the dot pattern. So by introducing a minimum amount of clustering, a halftone can be produced that accepts a certain amount of low-frequency graininess in exchange for reducing the periodic texture produced by the underlying sampling grid. As such, it is possible to produce visually pleasing dither patterns on hexagonal sampling grids which have a smaller maximum spatial frequency compared to the equivalent rectangular grid.

In the area of color halftoning, this edition greatly expands the prior discussion of superimposing component green-noise dither patterns by focusing on the problem of minimizing stochastic moiré, the aperiodic equivalent of the moiré produced by superimposing AM halftones. In particular while it is generally accepted that superimposing FM halftones results in a certain amount of color noise, it is the superposition of green-noise dither patterns that results in significantly worse amounts of this noise as the frequency content of stochastic moiré is directly related to the frequency content of the component dither patterns. Given this relationship, stochastic moiré can be minimized by properly varying the coarseness of the component patterns across colors.

Now in a problem closely related to color halftoning, a chapter has been added that focuses on the process of multi-toning and the creation of patterns composed of more than two gray-levels. Specifically, we address this new area by means of a blue-noise multi-tone model describing, not only the optimal distribution of multi-tone pixels, but also the optimal distributions of the various intensity levels between the component inks. In particular, we look at the change in halftone texture across gray-levels as the concentrations of component inks vary choosing the concentrations that minimize discontinuities in texture as being optimal. And finally, this book includes an introduction to the lenticular printing problem where multiple images are spatially multiplexed onto the back of a plastic lens array such that, when viewed through the lenses, only one component image is visible depending upon the angle at which the lens arrayed is viewed.