



Bounds and Algorithms for Dither Screens

Craig Gotsman, Jan P. Allebach*
HP Israel Science Center**
HPL-96-28
February, 1996

**halftoning,
dithering**

A simple model for quantifying dither screen performance is proposed. Using this model, inherent limitations on the performance of dither screen halftoning are proved. An efficient method for generating quality dither screens, with good performance under our model, is presented.

Internal Accession Date Only

**School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285.*

Email: allebach@ecn.purdue.edu

***Technion City, Haifa 32000, Israel*

© Copyright Hewlett-Packard Company 1996

1 Introduction

A common technique for binary halftoning is the use of an image-independent dither “screen” of gray level thresholds. Given a gray level image, its halftone may be computed rapidly by comparing the gray level of each pixel against the corresponding screen threshold. We say that the dither screen *induces* the halftone. An n -pixel dither screen for m gray levels contains n/m thresholds of each gray level.

Much effort has been invested in the design of such screens (e.g. [Bay73, MP91, Uli93, AS79, Lin94]), such that the induced halftones will be artifact-free and visually pleasing. The hope is that the perceived halftone will be as close as possible to the input image, under some visual metric. A common way of measuring the quality of a dither screen is its performance on images of uniform gray level. Since a dither screen is image-independent, it is obvious that its performance will not be as good as “direct” image-dependent (but slower) halftoning techniques, e.g. error-diffusion [FS75] and search-based methods [PN92, MA92].

Denote by $I(g)$ the uniform gray scale image of level g , by D the dither screen, and by $H_D(g)$ the binary halftone corresponding to $I(g)$ induced by D . The construction of $H_D(g)$ may be viewed as an incremental process, where white dots are added to a black background as g is increased. The dither screen D is the matrix corresponding to the times during the process, at which the corresponding halftone pixel is first whitened. At any given moment during this process, the set of white dots already placed is the halftone $H_D(g)$ of some $I(g)$, implying that the set of white dots comprising $H_D(g_1)$ is a subset of $H_D(g_2)$, for any $g_2 > g_1$. This is a major constraint on the halftones, severely limiting $H_D(g_2)$, once $H_D(g_1)$ has been fixed. This contrasts with direct halftoning methods, in which $H(g_2)$ may be quite different from $H(g_1)$. On an intuitive level, this suggests that if the screen induces a superior halftone for $I(g)$, all the chances are that the halftone for other gray levels in the immediate vicinity of g will be “inferior” (in some sense).

In this paper, we address the question of what may be considered an “inferior” or “superior” halftone. Using our definitions, we present bounds on the performance of *any* dither screen.

Very briefly, the *halftone error* at level g , induced by dither screen D , is

$$\text{Err}_D(g) = \frac{1}{n} \|H_D(g) * f - I(g)\|^2$$

where f is a (low-pass) filter modeling the human visual system, $*$ is the convolution operator and $\|\cdot\|$ is the l_2 norm taken over all image pixels. Ideally, we would like $\text{Err}_D(g)$ to vanish for all g .

We show that for any dither screen D ,

$$\mathbf{Av}_{g \in \Gamma}[\text{Err}_D(g)] > c > 0$$

where c is a constant depending only on the filter f , and the average is taken over all gray levels g in the interval Γ .

As a simple example, consider a “random” dither screen R , where the thresholds are arranged randomly on the screen (see Figs. 2(b),(f) and 3(b) for halftones $H_R(g)$ induced by R). If f is a $k \times k$ box filter, it is easily shown that

$$\text{Err}_R(g) \approx \frac{1}{k^2} g(1 - g)$$

where g is normalized to the range $[0, 1]$, hence

$$\mathbf{Av}_{g \in [0,1]}[\text{Err}_R(g)] \approx \frac{1}{6k^2} .$$

A more carefully designed screen will reduce $\text{Err}(g)$ significantly, but our results imply that this cannot be achieved simultaneously at all gray levels.

We propose a method for constructing dither screens having low values of $\text{Err}(g)$. The method is based on efficient computational-geometric algorithms, and a greedy optimization procedure.

2 Dither Screen and Halftone Quality

How to quantify the fidelity of a halftoned version of a gray level image is a major open question. Even more difficult is to quantify the quality of a dither screen, which induces

a halftone on any given image. A common test for the latter is to examine the halftones induced on images of uniform grey level intensity. Bayer [Bay73] first formulated a criterion for an “optimal” dither screen. He postulated that the induced halftones for uniform grey level images should not contain too many low frequency components. This led him to design the popular Bayer dither screen, which we call B in the sequel.

The Bayer dither screen has symmetries which cause regular patterns in $H_B(g)$ (see Figs. 2(a),(e) and 3(a)). We propose a more sophisticated version of the Bayer argument, based on a simple visual perception model. For small values of g , $H(g)$ consists of a sparse sprinkle of white pixels on a black background. Since the distances between individual white pixels are large compared to the distance between the centers of adjacent image pixels, the fact that the pixels are actually grid points is not noticeable. Consequently, the white pixels are perceived as white “dots” on a continuous black background. The human eye resolves the individual dots, and a pleasing “texture” is one in which the dots are approximately equidistant, without significant “clusters” or “voids” [Uli93]. The number of dots in any given convex image region should be proportional to the area of the region. How to position dots in the plane with this property is the concern of classical combinatorial *discrepancy* theory [BC87]. Note that large values of g are dual to small values of g , in which the roles of the black and white pixels in $H(g)$ are reversed. For intermediate values of g , the number of black and white dots are of the same order of magnitude, and at approximately the same density. The fact that the “dots” are constrained to grid points and have finite area becomes a major problem. Now the eye does not resolve the individual pixels, but rather blends them together. This may be modeled mathematically by applying a low-pass filter to the halftone pattern. The halftone is considered good if its filtered version is similar to the input gray level image. Gotsman [Got93], Allebach and Analoui [AA92], Pappas and Neuhoff [PN92] and Mulligan and Ahumada [MA92] have used this idea to produce direct halftones for gray level images. The gray level at which to switch from the distance criterion to the average density criterion depends on the size of the support of the visual filter, which, in turn, depends on the viewing geometry.

3 The Dither Screen Generation Algorithm

As mentioned in the previous section, it is desirable to have a dither screen D inducing halftones $H_D(g)$ with a dot distribution as close to uniform as possible, for small values of g . Another way of saying this is that the dots are mutually as distant as possible from each other. Since the dither screen imposes the constraint that the n -dot pattern must be a subset of the $n + 1$ -dot pattern, this calls for an “incremental” algorithm, in which the dots are added one by one. We propose to generate dot patterns in a manner similar to that of the “maximal distance” sampling algorithm of Eldar et al [ELPZ94]. This is a pseudo-random procedure, which starts from a very small number of dots positioned randomly on the unit square. It then adds new dots *deterministically*. The next dot added is positioned at the point in the unit square most distant from all other previous points, namely, if P is the current point set, the next point is b , such that:

$$b = \arg \max_{q \in [0,1]^2} \min_{p \in P} \|p - q\| \quad .$$

This involves geometric computations, which may be done efficiently using the Voronoi diagram ([PS85], Chap. 8) of the dynamic point set P . In a nutshell, at each step, the center of the largest empty circle in $[0, 1]^2$ is added. This involves maintaining and updating the Voronoi diagram of P , and maintaining a priority queue of Voronoi vertices. Determining the n 'th point involves $O(\log n)$ computation, so it is very efficient.

As opposed to that of Eldar et al, our application imposes a constraint on the points involved, namely that they be grid points. Furthermore, in order that high-quality dot patterns be produced both for very high and very low densities, we attack the problem from both ends, alternating between adding white dots to a black background, and black dots to a white background. The exact procedure is described in Fig. 1. Using efficient geometric data structures and algorithms, we are able to generate the low and high thresholds of the dither screen very rapidly. This contrasts with the *void and cluster* algorithm [Uli93], which attempts to achieve a similar goal. That algorithm is slower, as the initial dot pattern is determined by an iterative process.

For values of g in the midtone area, the maximal-distance procedure breaks down. Because of

the grid constraint, there are too many equivalent candidates at each step, and we essentially have to choose randomly between them. This degrades the quality of the dither screen. For these values of g , it is best to try and minimize Err_{MD} directly. This is done by a gradient-descent procedure, again advancing simultaneously from both ends of the gray level scale.

Figs. 2(c),(g) and 3(c) show some results of these procedures.

4 A Bound on Dither Screen Quality

This section shows that all dither screens are inherently limited in their capability to produce high quality halftones. Given a dither screen D , the induced halftone $H_D(g)$ might be good for individual values of g , but this usually is at the expense of the induced halftones at adjacent values of g . In other words, $H_D(g)$ cannot simultaneously be good for all values of g in a large interval.

Definition 1 Denote by $\delta(i)$ the real unit n -vector of zeros with 1 at the i 'th coordinate. An n pixel dither screen, denoted by D , is a permutation π_D of $(1, \dots, n)$. G is the set of m normalized gray levels, $G = \{0, \frac{1}{m}, \frac{2}{m}, \dots, 1\}$. D induces a halftone of $I(g)$, $H_D : G \rightarrow \{0, 1\}^n$, by

$$H_D(g) = \sum_{k=1}^{gn} \delta(\pi_D(k)) \quad (1)$$

Assume m divides n . As g is incremented from $\frac{i}{m}$ to $\frac{i+1}{m}$, $\frac{n}{m}$ vectors (dots) are added to the induced halftone.

Theorem 1 Let D be an n -pixel dither screen on m gray levels, and f a unit sum n -filter. Define

$$\text{Err}_D(g) = \frac{1}{n} \|H_D(g) * f - I(g)\|^2$$

where $I(g) = (g, \dots, g)$ and $\|\cdot\|$ is the l_2 norm. Let $\Gamma = \{g_1, \dots, g_2\}$ and $|\Gamma| = m(g_2 - g_1) + 1$.

Then

$$\mathbf{A} \mathbf{v}_{g \in \Gamma} [\text{Err}_D(g)] \geq \frac{1}{4} \left(1 - \frac{1}{|\Gamma|}\right) \sum_{\omega \neq 0} |\hat{f}(\omega)|^2 \quad .$$

```

/*
* Generate maximal-distance dither screen.
*  $n$  is the number of screen pixels.  $m$  is the number of gray levels.
*  $gthresh$  is a level threshold such that  $gthresh \leq m/2$ .
* The procedure deals with levels  $g$  in the range  $g \in [0, gthresh]$ 
* and  $g \in [m - gthresh, m]$ .
*
*  $random\_subset(S,r)$  produces a random subset of  $S$  of size  $r$ .
*/

const r=10;

maximal_dist( $n, m, gthresh$ )
begin
   $G := \{1, \dots, n\}$ ;
   $B := random\_subset(G,r)$ ;
   $G := G - B$ ;
   $W := random\_subset(G,r)$ ;
   $G := G - W$ ;
   $thresh := 0$ ;
  while ( $thresh \leq gthresh$ ) do
    begin
       $b := \max_{p \in G} \min_{q \in B} \|p - q\|$ ;
       $D(b) := m - thresh$ ;
       $B := B \cup \{b\}$ ;
       $G := G - \{b\}$ ;
       $w := \max_{p \in G} \min_{q \in W} \|p - q\|$ ;
       $D(w) := thresh$ ;
       $W := W \cup \{w\}$ ;
       $G := G - \{w\}$ ;
      if ( $mod(|B|, n/m) == 0$ )  $thresh++$ ;
    end;
  return D;
end;

```

Figure 1: The maximal-distance screen generation algorithm.

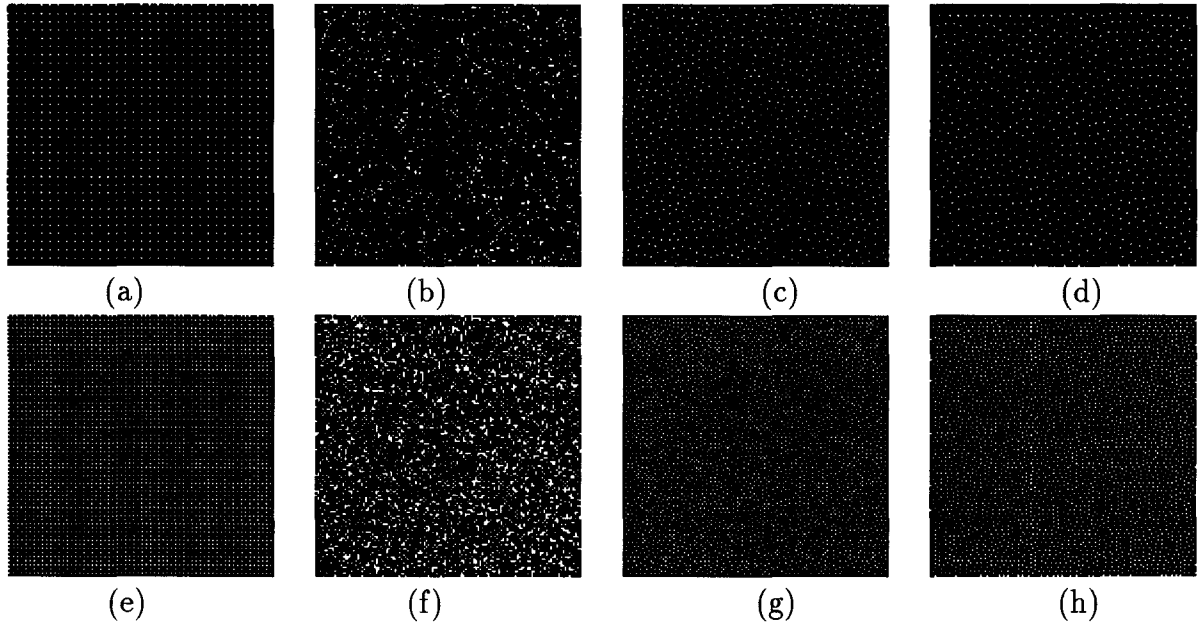


Figure 2: Halftones for uniform gray level images $g_1 = 15/255$ and $g_2 = 50/255$ induced by dither screens (a) Bayer $H_B(g_1)$ (b) Random $H_R(g_1)$ (c) Maximal distance $H_{MD}(g_1)$. (d) Floyd-Steinberg error diffusion $H_{FS}(g_1)$ for comparison. (e) $H_B(g_2)$, (f) $H_R(g_2)$, (g) $H_{MD}(g_2)$, (h) $H_{FS}(g_2)$. Note that $H_R(g_1) \subset H_R(g_2)$, $H_B(g_1) \subset H_B(g_2)$ and $H_{MD}(g_1) \subset H_{MD}(g_2)$, but a similar relation does not hold between $H_{FS}(g_1)$ and $H_{FS}(g_2)$.

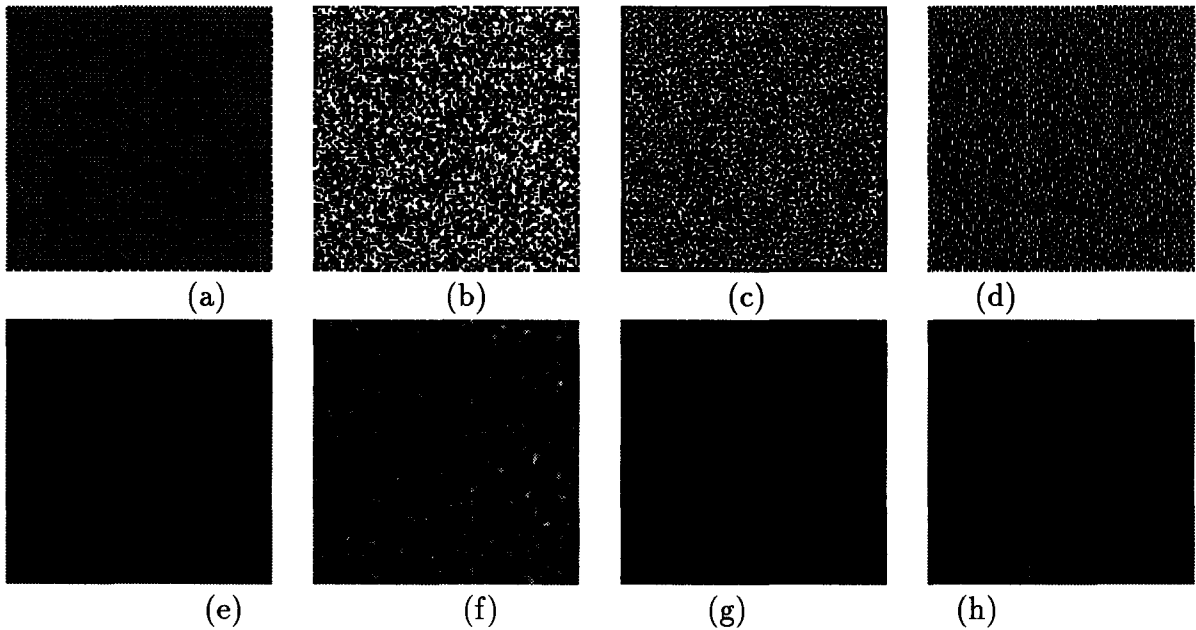


Figure 3: Halftones for uniform images $g = 100/255$ induced by dither screens. (a) Bayer $H_B(g)$ (b) Random $H_R(g)$ (c) Maximal distance $H_{MD}(g)$. (d) Floyd-Steinberg error diffusion $H_{FS}(g)$ for comparison. The halftones, filtered by a 3×3 box filter, are in (e) $\text{Err}_B(g) = 0.65 \times 10^{-2}$, (f) $\text{Err}_R(g) = 2.64 \times 10^{-2}$ and (g) $\text{Err}_{MD}(g) = 0.57 \times 10^{-2}$ (h) $\text{Err}_{FS}(g) = 0.41 \times 10^{-2}$.

Proof: Define the constants of the discrete Fourier transform such that $\widehat{f}(0) = \mathbf{E}[f]$. Then the convolution theorem is

$$(\widehat{f * g}) = n(\widehat{f} \cdot \widehat{g}) \quad (2)$$

and Parseval's theorem

$$\langle f, g \rangle = n \langle \widehat{f}, \widehat{g} \rangle \quad (3)$$

$\langle \cdot, \cdot \rangle$ is the inner product. Note that $\|f\|^2 = \langle f, f \rangle$. Since $(H_D(\widehat{g}) * f)(0) = \mathbf{E}[H_D(g) * f] = \mathbf{E}[H_D(g)] = \mathbf{E}[I(g)] = g$, transferring to the Fourier domain by applying (3) and then (2) implies that the error is obtained as the sum of squares of the energy in all non-zero frequencies:

$$\begin{aligned} \text{Err}_D(g) &= \frac{1}{n} \|H_D(g) * f - I(g)\|^2 \\ &= \|H_D(g) * \widehat{f} - I(g)\|^2 \\ &= \sum_{\omega \neq 0} |(H_D(\widehat{g}) * f)(\omega)|^2 \\ &= n^2 \sum_{\omega \neq 0} |\widehat{H}_D(g)(\omega) \cdot \widehat{f}(\omega)|^2 \\ &= n^2 \sum_{\omega \neq 0} |\widehat{H}_D(g)(\omega)|^2 \cdot |\widehat{f}(\omega)|^2 \end{aligned}$$

Hence

$$\begin{aligned} \sum_{g \in \Gamma} \text{Err}_D(g) &= \sum_{g \in \Gamma} \left[n^2 \sum_{\omega \neq 0} |\widehat{f}(\omega)|^2 |\widehat{H}_D(g)(\omega)|^2 \right] \\ &= n^2 \sum_{\omega \neq 0} \left[|\widehat{f}(\omega)|^2 \left[\sum_{g \in \Gamma} |\widehat{H}_D(g)(\omega)|^2 \right] \right] \end{aligned}$$

The inner sum may be bound using definition (1):

$$\begin{aligned} \sum_{g \in \Gamma} |\widehat{H}_D(g)(\omega)|^2 &= \sum_{g \in \Gamma} \left| \sum_{k=1}^{gn} \widehat{\delta}(\pi_D(k)) \right|^2 \\ &= \sum_{g \in \Gamma} \left| \frac{1}{n} \sum_{k=1}^{gn} \exp(i\omega\pi_D(k)) \right|^2 \\ &\geq \frac{|\Gamma| - 1}{4n^2} \end{aligned}$$

The last inequality is by Lemma 1 in the Appendix. From this follows:

$$\begin{aligned}
\mathbf{A} \mathbf{v}_{g \in \Gamma} [\text{Err}_D(g)] &= \frac{1}{|\Gamma|} \sum_{g \in \Gamma} \text{Err}_D(g) \\
&\geq \frac{n^2}{|\Gamma|} \cdot \frac{|\Gamma| - 1}{4n^2} \sum_{\omega \neq 0} |\hat{f}(\omega)|^2 \\
&= \frac{1}{4} \left(1 - \frac{1}{|\Gamma|}\right) \sum_{\omega \neq 0} |\hat{f}(\omega)|^2
\end{aligned}$$

■

5 Experimental Results

It is interesting to see how some dither screens compare in their performance to the lower bound of Theorem 1, and if direct halftones of uniform intensity images beat them. Figs. 4-6 summarize this comparison.

The Bayer dither matrix [Bay73] is an example of a screen designed such that the 2×2 box filter produces zero error for gray levels $g = \{\frac{0}{255}, \frac{64}{255}, \frac{128}{255}, \frac{192}{255}\}$ (see Fig. 4). For this filter, and the entire interval $\Gamma = \{0, \frac{1}{255}, \frac{2}{255}, \dots, 1\}$, the Bayer dither matrix gives an average error of 1.05×10^{-2} , and the maximal-distance screen 1.96×10^{-2} , compared to the lower bound of 0.10×10^{-2} . A random dither matrix gives the average value 4.11×10^{-2} , which is *much* larger. Note that the Bayer dither screen is not as good when different filters are used, e.g. the 3×3 box filter (Fig. 5) or the truncated Gaussian (Fig. 6). Direct halftoning algorithms, such as Floyd-Steinberg error-diffusion [FS75], yield error values less than those of the dither screen induced halftones, as can be seen in Figs. 4-6. The average values for all methods, and the lower bound, are summarized in Table 1.

6 Discussion

The maximal distance algorithm for dither screen design could conceivably be extended to perform *direct* halftoning of grey level images. This is a current research avenue.

This paper provides theoretical bounds on the performance of all dither screens. It is not surprising that such bounds exist, considering the extremely constrained nature of the induced

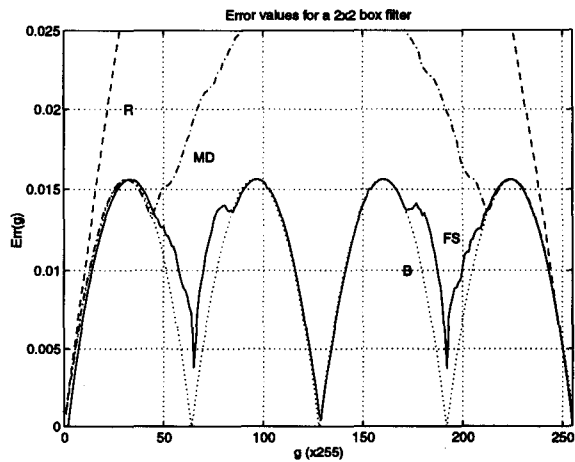


Figure 4: Error values for dither screens produced by random (R), Bayer (B) and maximal distance (MD) dither screens using a 2x2 box visual filter. The graph for Floyd-Steinberg error-diffusion (FS) is provided for comparison.

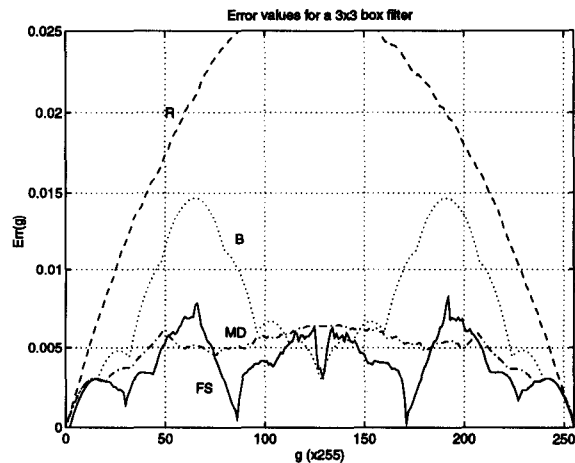


Figure 5: Error values for dither screens produced by random (R), Bayer (B) and maximal distance (MD) dither screens using a 3x3 box visual filter. The graph for Floyd-Steinberg error-diffusion (FS) is provided for comparison.

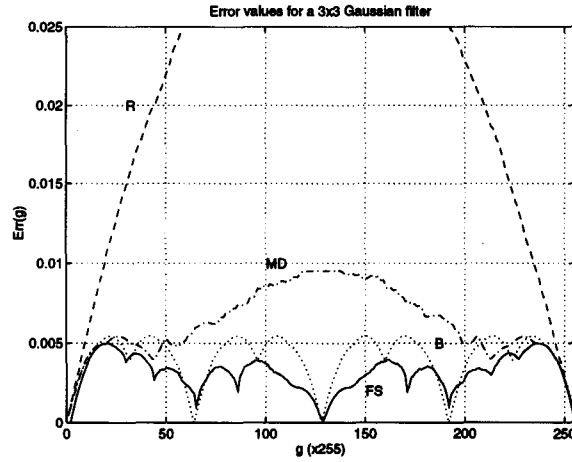


Figure 6: Error values for dither screens produced by random (R), Bayer (B) and maximal distance (MD) dither screens using a 3x3 discrete Gaussian filter. The graph for Floyd-Steinberg error-diffusion (FS) is provided for comparison.

| Method | | R | B | MD | FS | LB |
|--------------|--|------|------|------|------|------|
| Filter | | | | | | |
| 2x2 box | | 4.11 | 1.05 | 1.96 | 1.15 | 0.10 |
| 3x3 box | | 1.82 | 0.78 | 0.48 | 0.40 | 0.04 |
| 3x3 Gaussian | | 2.30 | 0.41 | 0.63 | 0.31 | 0.05 |

Table 1: Comparison of average error values ($\times 10^2$) over the entire 256 gray levels for various filters and dither screens. Included for comparison are the value for Floyd-Steinberg error diffusion and the lower bound of Theorem 1.

halftones. However, the lower bound of Theorem 1 seems to be rather weak, far from tight. We believe it can be improved significantly, perhaps using discrepancy theory, or exponential sums.

7 References

- [AA92] J.P. Allebach and M. Analoui. Model-based halftoning by direct binary search. In *Proceedings of the 1992 SPIE/IS&T Symposium on Electronic Imaging Science and Technology*, 1992.
- [AS79] J.P. Allebach and R.N. Stradling. Computer-aided design of dither signals for binary display of images. *Applied Optics*, 18:2708–2713, 1979.
- [Bay73] B.E. Bayer. An optimum method for two-level rendition of continuous-tone pictures. In *Proceedings of IEEE Intl. Conf. on Comm.*, pages 26–11 – 26–15, 1973.
- [BC87] J. Beck and W. Chen. *Irregularities of Distribution*. Cambridge Univ. Press, Cambridge, 1987.
- [ELPZ94] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Zeevi. The farthest-point strategy for progressive image sampling. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem*, 1994.
- [FS75] R. Floyd and L. Steinberg. An adaptive algorithm for spatial gray scale. In *SID Symposium*, pages 36–37, 1975.
- [Got93] C. Gotsman. Halftoning image sequences. *The Visual Computer*, 9:255–266, 1993.
- [Lin94] Q. Lin. Improving halftone uniformity and tonal response. In *IS&T Tenth Congress on Advances in Non-Impact Printing Technologies*, pages 377–380, New Orleans, Louisiana, Oct. 1994.
- [MA92] J.B. Mulligan and A.J. Ahumada. Principled halftoning based on human vision models. In *Proceedings of the 1992 SPIE/IS&T Symposium on Electronic Imaging Science and Technology*, 1992.
- [MP91] T. Mitsa and K. Parker. Digital halftoning using a blue noise mask. In *Proceedings of SPIE, Image Proc. Algorithms and Techniques II*, volume 1452, pages 47–56, 1991.

- [PN92] T.N. Pappas and D.L. Neuhoff. Least-squares model-based halftoning. In *Proceedings of the 1992 SPIE/IS&T Symposium on Electronic Imaging Science and Technology*, 1992.
- [PS85] F.P. Preparata and M.I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, 1985.
- [Uli93] R. Ulichney. The void-and-cluster method for dither array generation. In *Proceedings of the 1993 SPIE/IS&T Symposium on Electronic Imaging Science and Technology*, pages 265–270, 1993.

Appendix

Lemma 1 *Let v_1, \dots, v_m be unit complex vectors. Then*

$$\sum_{i=1}^m \left| \sum_{k=1}^i v_k \right|^2 \geq \frac{m}{4}$$

Proof: Denote $w_i = \sum_{k=1}^i v_k$. For any $i \geq 1$,

$$\begin{aligned} |w_i|^2 + |w_{i+1}|^2 &= |w_i|^2 + |w_i + v_{i+1}|^2 \\ &\geq |w_i|^2 + (|w_i| - |v_{i+1}|)^2 \\ &= |w_i|^2 + (|w_i| - 1)^2 \\ &\geq 1/2 \end{aligned}$$

The first inequality is the triangle inequality, and the second is due to the fact that the real function $f(x) = x^2 + (1 - x)^2$ is minimized by $f(1/2) = 1/2$.

Hence, for any even m ,

$$\begin{aligned} \sum_{i=1}^m |w_i|^2 &= \sum_{i=1}^{m/2} (|w_{2i-1}|^2 + |w_{2i}|^2) \\ &\geq \sum_{i=1}^{m/2} 1/2 \\ &= \frac{m}{4} \end{aligned}$$

For odd m , the first term of the sum is 1, and the previous argument applies for the next $m - 1$ terms, so the theorem still holds. ■

Note: If the v_k are random unit complex vectors (this corresponds to the case of the random dither screen), the value of $|\sum_{k=1}^i v_k|^2$ is i on the average, hence, for large m $\sum_{i=1}^m |\sum_{k=1}^i v_k|^2 = \theta(m^2)$.