

# Non-independent Randomized Rounding and an Application to Digital Halftoning\*

Benjamin Doerr<sup>†‡</sup>

April 9, 2004

## Abstract

We investigate the problem to round a given  $[0, 1]$ -valued matrix to a  $0, 1$  matrix such that the rounding error with respect to  $2 \times 2$  boxes is small. Such roundings yield good solutions for the digital halftoning problem as shown by Asano et al. (SODA 2002). We present a randomized algorithm computing roundings with expected error at most 0.5463 per box, improving the 0.75 non-constructive bound of Asano et al. Our algorithm is the first one solving this problem fast enough for practical application, namely in linear time.

Of a broader interest might be our rounding scheme, which is a modification of randomized rounding. Instead of independently rounding the variables, we impose a number of suitable dependencies. Thus by equipping the rounding process with some of the problem information, we reduce the rounding error significantly compared to independent randomized rounding, which leads to an expected error of 0.82944 per box. Finally, we give a characterization of realizable dependencies.

**Key words:** Randomized rounding, discrepancy, digital halftoning.

## 1 Introduction

In this paper, we are concerned with rounding problems. In general form, these problems are of the following type: Given some numbers  $x_1, \dots, x_n$ , one is looking for roundings

---

\*Preliminary versions of this work appeared at ESA 2002 and SODA 2003.

<sup>†</sup>partially supported (associate member) by the graduate school ‘Effiziente Algorithmen und Multiskalenmethoden’, Deutsche Forschungsgemeinschaft

<sup>‡</sup>Mathematisches Seminar II, Christian-Albrechts-Universität zu Kiel, 24098 Kiel, Germany, bed@numerik.uni-kiel.de

$y_1, \dots, y_n$  such that some given error measures are small. By rounding we always mean that  $y_i = \lfloor x_i \rfloor$  or  $y_i = \lceil x_i \rceil$ . Since there are  $2^n$  possibilities, such rounding problems are good candidates for hard problems. In fact, even several restricted versions like the discrepancy problem are known to be *NP*-hard. An example related to the problems regarded in this paper can be found in Asano, Matsui and Tokuyama [3, 4].

On the other hand, there are cases that can be solved optimally in polynomial time. Knuth [17] for example has shown that there exists a rounding such that the errors  $|\sum_{i=1}^k (y_i - x_i)|$  and  $|\sum_{i=1}^k (y_{\pi(i)} - x_{\pi(i)})|$  for a fixed permutation  $\pi$  and all  $1 \leq k \leq n$  are at most  $\frac{n}{n+1}$ . Such roundings can be obtained by computing a maximum flow in a network. A recent generalization [7] shows this bound for arbitrary totally unimodular rounding problems.

## 1.1 The Digital Halftoning Problem: A Matrix Rounding Problem

The rounding problem regarded in this paper is motivated by an application from image processing. The *digital halftoning problem* is to convert a continuous-tone intensity image (each pixel may have an arbitrary ‘color’ on the white-to-black scale) into a binary image (only black and white dots are allowed). An intensity image can be represented by a  $[0, 1]$ -valued  $m \times n$  matrix  $A$ . Each entry  $a_{ij}$  corresponds to the brightness level of the pixel with coordinates  $(i, j)$ . Since many devices, e.g., laser printers, can only output white and black dots, we have to round  $A$  towards a  $0, 1$  matrix. Naturally, this has to be done in a way that the resulting image looks similar to the original one.

This notion of similarity is a crucial point. From the viewpoint of application, similarity is defined via the human visual system: A rounding is good, if an average human being can retrieve most of the original information from the rounded image. Using this ‘criterion’, several algorithms turned out to be useful: Floyd and Steinberg [11] proposed the error diffusion algorithm, that rounds the entries one by one and distributes the rounding error over neighboring not yet rounded entries. Lippel and Kurland [21] and Bayer [5] investigated the ordered dither algorithm, which partitions the image into fixed size submatrices and rounds each submatrix by comparing its entries with a threshold matrix of same size. One advantage of this approach is that it can be parallelized easily. Knuth [16] combined ideas of both approaches to get an algorithm called dot diffusion.

Though some work has been done in this direction, fewer understanding seems to be present on the theoretical side. In particular, a good mathematical formulation of similarity seems hard to find. Such a similarity measure is desirable for two reasons: Firstly, it allows to compare algorithms without extensive experimental testing. This is particularly interesting, since comparing different halftonings is a delicate issue. For example, it makes a huge difference whether the images are viewed on a computer screen or are printed on

a laser printer. Different printers can also give different impressions. Therefore, a more objective criterion would be very helpful. A second reason is that having a good criterion, one would have a clearer indication of how a digital halftoning algorithm should work. Thus developing good algorithms would be easier.

So far, the most widely accepted criterion for a good halftoning algorithm is that it has the ‘blue noise’ property (first detected by Ulichney [34], cf. also the surveys Ulichney [33] and Lau and Arce [19]). This refers not to a similarity measure comparing two images, but an analysis of how the algorithm performs on constant grey level areas. Thus, on the other hand, it gives no information on how changing intensities, in particular shapes, are reproduced. This work has been extended by Sullivan, Ray and Miller [32], who use a model based on the human visual system to compute good halftoning patterns for constant intensity areas.

Significant research has been done on “global” (regarding the whole image) similarity measures, see e.g. Lieberman and Allebach [20] and the references therein. However, in many cases the complexity of these measures makes a theoretical investigation almost impossible. Often even computing approximations efficiently is difficult.

On the international ACM/SIAM Symposium on Discrete Algorithms 2002, Asano et al. [2] presented a structurally much simpler similarity measure together with several theoretical results. Their experimental studies indicate that good digital halftonings have small error with respect to all  $2 \times 2$  subregions. This yields the following problem.

## 1.2 Problem Statement and Results

Let  $A \in [0, 1]^{m \times n}$  denote our input matrix. The set  $R_{ij} := \{i, i + 1\} \times \{j, j + 1\} = \{(i, j), (i + 1, j), (i, j + 1), (i + 1, j + 1)\}$  for some  $i \in [m - 1], j \in [n - 1]$ , is called a  $2 \times 2$  subregion (or box) in  $[m] \times [n]$ .<sup>1</sup> Denote by  $\mathcal{R}$  the set of all these boxes. We write  $A_{R_{ij}}$  for the  $2 \times 2$  matrix  $\begin{pmatrix} a_{i,j} & a_{i,j+1} \\ a_{i+1,j} & a_{i+1,j+1} \end{pmatrix}$  induced by  $R_{ij}$ . For any matrix  $A$  put  $\Sigma A := \sum_{i,j} a_{ij}$ , the sum of all its entries.

For a matrix  $B \in \{0, 1\}^{m \times n}$  — which is a rounding of  $A$  unless we have  $a_{ij}, b_{ij} \in \mathbb{Z}$  and  $a_{ij} \neq b_{ij}$  for some  $i, j$  — we define the rounding error of  $A$  with respect to  $B$  by

$$d_{\mathcal{R}}(A, B) := \sum_{R \in \mathcal{R}} |\Sigma A_R - \Sigma B_R|.$$

We usually omit the subscript  $\mathcal{R}$  when there is no danger of confusion.

Asano et al. [2] exhibited that roundings  $B$  such that  $d(A, B)$  is small, yield good digital halftonings. They showed that for any  $A$  an optimal rounding  $B^*$  satisfies  $d(A, B^*) \leq$

---

<sup>1</sup>For an arbitrary number  $r$  we denote by  $[r]$  the set of positive integers not exceeding  $r$ .

$0.75|\mathcal{R}|$ . They also gave a polynomial time algorithm computing a rounding  $B$  such that  $d(A, B) - d(A, B^*) \leq 0.5625|\mathcal{R}|$ . It is easy to see that there are matrices  $A$  such that all roundings (and in fact all integral matrices)  $B$  have  $d(A, B) \geq 0.5|\mathcal{R}|$ .

A major draw-back of the algorithm given in [2] is that it is not very practical, as it requires the solution of an integer linear program with totally unimodular constraint matrix. This leads to a run-time bound which is at least quadratic in the number  $nm$  of pixels. As pointed out in [2], this is too slow for a real world application.

In this paper, we present a randomized algorithm that runs in linear time. It may be implemented in parallel without problems. This divides the run-time by the number of processors available. The roundings computed by our algorithm have an expected error that exceeds the optimal one by at most  $0.3125|\mathcal{R}|$  (instead of  $0.5625|\mathcal{R}|$ ). They also satisfy the absolute bound  $E(d(A, B)) \leq 0.5463|\mathcal{R}|$ , beating the non-algorithmic bound of  $0.75|\mathcal{R}|$  of [2].

The distribution of the rounding error resulting from this algorithm is highly concentrated around the expected value. The probability that the error exceeds the expected one by more than  $\varepsilon|\mathcal{R}|$  is bounded by  $\exp(-\Omega(\varepsilon^2 \sqrt{|\mathcal{R}|}))$ . Our algorithm can also be derandomized. This yields a deterministic linear-time algorithm computing roundings with error guarantee at most the expected one of the randomized version.

For an experimental study of how well our algorithm performs in practise, we refer to Schnieder [25]. Some of these results are also contained in [9].

### 1.3 Non-independent Randomized Rounding

The key idea of our algorithm might also be of a broader interest. We develop a randomized rounding scheme where the individual roundings are not independent. The classical approach of randomized rounding due to Raghavan and Thompson [23] and Raghavan [22] is to round each variable independently with probability depending on the fractional part of its value. This allows to use Chernoff-type large deviation inequalities showing that a sum of independent random variables is highly concentrated around its expectation.

Randomized rounding has been applied to numerous combinatorial optimization problems that can be formulated as integer linear programs (cf. Srinivasan [28]). Though being very effective in the general case, a known difficulty with randomized rounding is how to use structural information about the underlying problem. Since only the solution of the linear relaxation is used, further information about the underlying problem cannot be exploited. One idea to overcome this is to use correlation among the events in the analysis of randomized rounding. This allows to strengthen the classical bounds for packing and covering problems as shown by Srinivasan [29].

In this paper, we try so use the structure in an earlier phase, namely in the design of the random experiment. This leads to randomized roundings where the variables are not rounded independently. There are few attempts to use non-independent roundings. A straight-forward one already appeared in the work of Raghavan and Thompson and is highly motivated by the structure of the problem. In many cases, the linear program under consideration contains constraints of the kind

$$x_{i_1} + \cdots + x_{i_k} = 1. \tag{1}$$

Typical examples are the vector selection problem (exactly one vector has to be chosen from each of the given sets, cf. Raghavan [22]) or multi-commodity flow problems (exactly one path has to be chosen from a non-integral mixture of paths adding up to the flow, cf. Raghavan and Thompson [23]). In these cases, the probability that an independent randomized rounding is feasible, i.e., that  $y_{i_1} + \cdots + y_{i_k} = 1$  can be arbitrarily close to  $\frac{1}{e}$ . If the number of equations of type (1) is large, this yields an exponentially small success probability.

The solution is to pick one of the variables  $x_{i_1}, \dots, x_{i_k}$  with probability given by its value in the relaxation, set this variable to one and all others to zero. Thus the rounding is not done independently, but subject to the constraint that exactly one variable receives the value of one. A recent work of Srinivasan [30] extend this to constraints where sums of variables are required to have a particular value other than one.

Another paper on non-independent roundings is Bertsimas, Teo and Vohra [6]. They use dependent roundings to give alternative proofs of integrality for several classical polyhedra. They also use ‘global’ dependencies imposing restrictions on the number of variables rounded up or down. For the integrality gap of MINSAT with clauses having at most  $k$  literals, this reduces the trivial upper bound of 2 to the sharp bound of  $2(1 - 2^{-k})$ .

Looking at these results on dependent randomized rounding, we feel that the option to design the random experiment in a way that it reflects the structure of the underlying problem has not been exploited sufficiently. In this paper we try to move a step forward into this direction. We impose dependencies that are not necessary in the sense of feasibility, but helpful in minimizing our objective function  $d(A, B)$ . For the rounding problem regarded in the present paper, this improves the bound of  $0.82944|\mathcal{R}|$  obtained by independent randomized rounding down to  $0.5463|\mathcal{R}|$ .

In addition to the randomized rounding condition for single variables, we impose dependencies of the type

$$P\left(\sum_{i \in I_k} y_i = \left\lfloor \sum_{i \in I_k} x_i \right\rfloor + 1\right) = \left\{ \sum_{i \in I_k} x_i \right\} \tag{2}$$

for some sets  $I_k$  (we write  $\{r\} := r - \lfloor r \rfloor$  to denote the fractional part of  $r$ ). Hence our roundings not only are randomized roundings concerning the single variables, but also with

respect to the sums  $\sum_{i \in I_k} x_i$ .<sup>2</sup> By choosing suitable dependencies, we obtain the above mentioned result.

We also regard the question what dependencies can be realized. More formally, we ask how the sets  $I_k$  have to be chosen, such that for all values of the input variables a randomized rounding satisfying (2) exists. Surprisingly, there is a simple characterization: Such roundings exist if and only if the hypergraph consisting of the sets  $I_k$  is totally unimodular. Of course, with this characterization at hand, the search for suitable dependency sets is much easier.

## 2 Independent Randomized Rounding

For a number  $x$  we write  $\lfloor x \rfloor$  for the largest integer not exceeding  $x$ ,  $\lceil x \rceil$  for the smallest being not less than  $x$ , and  $\{x\} := x - \lfloor x \rfloor$  for the fractional part of  $x$ . We say that some random variable  $X$  is a *randomized rounding* of  $x$  if  $\Pr(X = \lfloor x \rfloor + 1) = \{x\}$  and  $\Pr(X = \lfloor x \rfloor) = 1 - \{x\}$ . In particular, if  $x \in [0, 1]$ , we have  $\Pr(X = 1) = x$  and  $\Pr(X = 0) = 1 - x$ .

We first analyze what can be achieved with independent randomized rounding. We say that  $B$  is an *independent randomized rounding* of  $A$  if each entry  $b_{ij}$  is a randomized rounding of  $a_{ij}$  and all these roundings are mutually independent. Let us remark that we do not expect this to be a competitive approach for the digital halftoning problem. In fact, this idea was experimented with already in the fifties by Goodall [13] (see also Roberts [24]), and thus long before the seminal work of Raghavan and Thompson. At that time, this was known under the name *thresholding with white noise* or *random dither*.

We use the result below both to estimate the effect of adding dependencies to the rounding process and in the proofs of some of our later results. Note that the proof is different from typical randomized rounding applications: Since the boxes are small, using a large deviation bound makes no sense and one has to compute the expected error exactly.

**Theorem 1.** *Let  $A \in [0, 1]^{m \times n}$ ,  $B$  an independent randomized rounding of  $A$  and  $B^*$  an optimal rounding of  $A$  (that is,  $d(A, B^*)$  is minimal among all roundings of  $A$ ). Then*

$$\begin{aligned} E(d(A, B)) &\leq 0.82944|\mathcal{R}|, \\ E(d(A, B)) &\leq 0.75|\mathcal{R}| + d(A, B^*). \end{aligned}$$

*Proof.* By linearity of expectation,  $E(d(A, B)) = \sum_{R \in \mathcal{R}} E(d(A_R, B_R)) = \sum_{R \in \mathcal{R}} E(|\Sigma A_R - \Sigma B_R|)$ . Hence our analysis is reduced to the rounding problem of a single box  $R$ . The

---

<sup>2</sup>The dependencies (1) used in [23] actually are a special case of this type where it is also required that (in the notation of (2))  $\sum_{i \in I_k} x_i = 1$  holds.

expected rounding error of a box  $\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$  is

$$f(a_1, a_2, a_3, a_4) = \sum_{S \subseteq [4]} \prod_{i \in S} a_i \prod_{i \notin S} (1 - a_i) \left| |S| - \sum_{i \in [4]} a_i \right|.$$

Let  $a_1, \dots, a_4 \in [0, 1]$  such that  $f(a_1, a_2, a_3, a_4)$  is maximal. Assume  $a_1 < a_2$ . Let  $0 < \varepsilon \leq \frac{1}{2}(a_2 - a_1)$ . Then

$$f(a_1, a_2, a_3, a_4) < f(a_1 + \varepsilon, a_2 - \varepsilon, a_3, a_4) \quad (3)$$

is easily computed, contradicting the maximality of  $f(a_1, a_2, a_3, a_4)$ . Hence  $a_1 \geq a_2$ . We have  $f(a_1, a_2, a_3, a_4) = f(a_{\sigma(1)}, a_{\sigma(2)}, a_{\sigma(3)}, a_{\sigma(4)})$  for all  $a_1, \dots, a_4 \in [0, 1]$  and all permutations  $\sigma \in S_4$ . Thus we conclude that  $a_i = a_j$  for all  $i, j \in [4]$ . Finally, we only need to check that  $\tilde{f} : [0, 1] \rightarrow \mathbb{R}; a \mapsto f(a, a, a, a)$  never exceeds  $0.82944 = f(0.4)$ . This is not hard to see as  $f$  is piecewise a polynomial of degree 5.

For an arbitrary number  $x$  denote by  $d(x, \mathbb{Z}) := \min\{x - \lfloor x \rfloor, \lceil x \rceil - x\}$  its distance to the integers. Obviously,  $\sum_{R \in \mathcal{R}} d(\Sigma A_R, \mathbb{Z})$  is a lower bound for the optimal error  $\min_{B^*} d(A, B^*)$ . Analyzing  $\tilde{f} : [0, 1]^4 \rightarrow \mathbb{R}$  defined by  $\tilde{f}(a_1, a_2, a_3, a_4) = f(a_1, a_2, a_3, a_4) - d(a_1 + a_2 + a_3 + a_4, \mathbb{Z})$  in a similar manner as above yields that  $f$  is bounded from above by 0.75, and thus the second bound.  $\square$

The two bounds of Theorem 1 are tight as shown by matrices with all entries 0.4 and 0.5 respectively.

### 3 Non-independent Randomized Rounding

In this section we improve the previous bounds by adding some dependencies to the rounding process. We start with an elementary approach called *joint randomized rounding*, which reduces the chance that neighboring matrix entries are both rounded in the wrong way. This leads to a first improvement in Corollary 4. We then add further dependencies leading to our final bound.

#### 3.1 Joint Randomized Rounding

**Definition 2 (Joint Randomized Rounding).** *Let  $a_1, a_2 \in [0, 1]$ . We say that  $(b_1, b_2)$  is a joint randomized rounding of  $(a_1, a_2)$  if*

- (i) for all  $i \in [2]$ ,  $b_i$  is a randomized rounding of  $a_i$ ;

(ii)  $b_1 + b_2$  is a randomized rounding of  $a_1 + a_2$ .

It is clear that joint randomized roundings exist for all  $a_1, a_2 \in [0, 1]$ : Define  $(b_1, b_2)$  through

$$\begin{aligned}\Pr(b_1 = 1 \wedge b_2 = 0) &= a_1, \\ \Pr(b_1 = 0 \wedge b_2 = 1) &= a_2, \\ \Pr(b_1 = 0 \wedge b_2 = 0) &= 1 - a_1 - a_2\end{aligned}$$

in the case  $a_1 + a_2 \leq 1$ , and

$$\begin{aligned}\Pr(b_1 = 1 \wedge b_2 = 0) &= 1 - a_2, \\ \Pr(b_1 = 0 \wedge b_2 = 1) &= 1 - a_1, \\ \Pr(b_1 = 1 \wedge b_2 = 1) &= a_1 + a_2 - 1\end{aligned}$$

in the case  $a_1 + a_2 > 1$ . The next lemma shows that two joint randomized roundings are superior to four independent randomized roundings in terms of the rounding error.

**Lemma 3.** *Let  $A = (a_{11}, a_{12}, a_{21}, a_{22})$  be a box. Let  $(b_{11}, b_{12})$  be a joint randomized rounding of  $(a_{11}, a_{12})$ , and  $(b_{21}, b_{22})$  one of  $(a_{21}, a_{22})$  independent of the first. Put  $B = (b_{11}, b_{12}, b_{21}, b_{22})$ . Let  $B^*$  an optimal rounding of  $A$ . Then*

$$\begin{aligned}E(d(A, B)) &\leq \frac{16}{27} \leq 0.5926, \\ E(d(A, B)) &\leq 0.5 + d(A, B^*).\end{aligned}$$

*Proof.*  $b_{11} + b_{12}$  is a randomized rounding of  $a_{11} + a_{12}$ , and the same holds for the second row. Hence all we have to do is to bound the expected deviation of a sum of two independent randomized roundings from the sum of the original values. We show that  $g : [0, 1]^2 \rightarrow \mathbb{R}$  defined by

$$\begin{aligned}g(x, y) &:= xy(2 - x - y) + (x(1 - y) + (1 - x)y)|1 - x - y| \\ &\quad + (1 - x)(1 - y)(x + y)\end{aligned}$$

never exceeds 0.5926. As in the proof of Theorem 1 we find that  $g(x, y)$  is maximal for  $x = y$ . Maximizing  $x \mapsto g(x, x)$  is straightforward and yields maxima at  $(\frac{1}{3}, \frac{1}{3})$  and  $(\frac{2}{3}, \frac{2}{3})$ , both having objective value  $\frac{16}{27} \leq 0.5926$ . The second bound follows from maximizing  $\tilde{g}$  defined by  $\tilde{g}(x, y) := g(x, y) - d(x + y, \mathbb{Z})$ .  $\square$

The bounds in Lemma 3 are sharp as shown by matrices having  $a_{11} + a_{12} = a_{21} + a_{22} = \frac{1}{3}$  and  $a_{11} + a_{12} = a_{21} + a_{22} = \frac{1}{2}$  respectively. Plastering the grid with independent joint randomized roundings already yields a first improvement over independent randomized rounding:

**Corollary 4.** Let  $A \in [0, 1]^{m \times n}$ . Compute  $B \in \{0, 1\}^{m \times n}$  by independently obtaining  $(b_{i,2j-1}, b_{i,2j})$  from  $(a_{i,2j-1}, a_{i,2j})$  by joint randomized rounding for all  $i \in [m], j \in [\frac{n}{2}]$ , and also independently obtaining  $b_{in}$  from  $a_{in}$ ,  $i \in [m]$ , by usual randomized rounding, if  $n$  is odd. Then

$$\begin{aligned} E(d(A, B)) &\leq 0.7111|\mathcal{R}|, \\ E(d(A, B)) &\leq 0.625|\mathcal{R}| + d(A, B^*), \end{aligned}$$

where  $B^*$  shall be an optimal rounding of  $A$ .

*Proof.* At least half of the boxes, namely all  $R_{ij}$  such that  $j$  is odd, are rounded in the manner of Lemma 3. The remaining ones contain four independent randomized roundings. Thus

$$\begin{aligned} E(d(A, B)) &\leq \frac{1}{2} \frac{16}{27} |\mathcal{R}| + \frac{1}{2} 0.82944 |\mathcal{R}| \leq 0.7111 |\mathcal{R}|, \\ E(d(A, B)) - d(A, B^*) &\leq \frac{1}{2} 0.5 |\mathcal{R}| + \frac{1}{2} 0.75 |\mathcal{R}| \leq 0.625 |\mathcal{R}| \end{aligned}$$

by Theorem 1 and Lemma 3. □

### 3.2 Block Randomized Rounding

**Definition 5 (Block Randomized Rounding).** Let  $A = (a_{11}, a_{12}, a_{21}, a_{22})$  be a box. We call  $B = (b_{11}, b_{12}, b_{21}, b_{22})$  a block randomized rounding of  $A$  if

- (i) each single entry of  $B$  is a randomized rounding of the corresponding one of  $A$ , i. e.,  $\Pr(b_{ij} = 1) = a_{ij}$  and  $\Pr(b_{ij} = 0) = 1 - a_{ij}$  for all  $i, j \in [2]$ ;
- (ii) each pair of neighboring entries has the distribution of the corresponding joint randomized rounding, i. e., in addition to (i) we have for all  $(i, j), (i', j') \in [2] \times [2]$  such that either  $i \neq i'$  or  $j \neq j'$ ,

$$\begin{aligned} \Pr(b_{ij} + b_{i'j'} = \lfloor a_{ij} + a_{i'j'} \rfloor + 1) &= \{a_{ij} + a_{i'j'}\} \\ \Pr(b_{ij} + b_{i'j'} = \lfloor a_{ij} + a_{i'j'} \rfloor) &= 1 - \{a_{ij} + a_{i'j'}\}; \end{aligned}$$

- (iii) the box in total behaves like a randomized rounding, i. e., we have

$$\begin{aligned} \Pr(\Sigma B = \lfloor \Sigma A \rfloor + 1) &= \{\Sigma A\} \\ \Pr(\Sigma B = \lfloor \Sigma A \rfloor) &= 1 - \{\Sigma A\}. \end{aligned}$$

**Lemma 6.** Let  $B$  be a block randomized rounding of a  $2 \times 2$  matrix  $A$  and  $B^*$  an optimal rounding of  $A$ . Then

$$\begin{aligned} E(d(A, B)) &\leq 0.5, \\ E(d(A, B)) &\leq 0.125 + d(A, B^*). \end{aligned}$$

*Proof.* Direct consequence of item (iii) of the definition.  $\square$

Both bounds are sharp as demonstrated by matrices  $A$  such that  $\{\sum A\} = \frac{1}{2}$  and  $\{\sum A\} = \frac{1}{4}$  respectively. The interesting point is that block randomized roundings always exist. In fact, even more complicated roundings exist, as shown by the following definition and lemma.

**Definition 7 (Continuous Block Randomized Rounding).** *Let  $A \in [0, 1]^{m \times n}$ . We say that  $B$  is a continuous block randomized rounding of  $A$  if*

- (i) for all  $i \in [\frac{m}{2}], j \in [n - 1]$ ,  $R := \{2i - 1, 2i\} \times \{j, j + 1\}$ ,  $B|_R$  is a block randomized rounding of  $A|_R$ ;
- (ii) if  $m$  is odd, then for all  $j \in [n - 1]$ ,  $(b_{m,j}, b_{m,j+1})$  is a joint randomized rounding of  $(a_{m,j}, a_{m,j+1})$  as in Definition 2;
- (iii)  $b_{i,j}$  is mutually independent of all  $b_{i',j'}$  such that  $\lceil i/2 \rceil \neq \lceil i'/2 \rceil$ .

**Lemma 8.** *For any  $A \in [0, 1]^{m \times n}$ , a continuous block randomized rounding  $B$  of  $A$  exists and can be computed in linear time.*

*Proof.* We may assume that  $m$  is even: If not, add an extra row to  $A$  with all entries zero, compute  $B$  as continuous block randomized rounding of this matrix and note that  $B$  without the last row — which will be all zero — is a continuous block randomized rounding of  $A$ .

We propose the following randomized algorithm:

**Input:**  $A = (a_{ij}) \in [0, 1]^{m \times n}$ .

**Output:**  $B = (b_{ij}) \in \{0, 1\}^{m \times n}$ , a continuous block randomized rounding of  $A$ .

**For all** odd  $i \in [m]$  **do** {

Compute  $(b_{i,1}, b_{i+1,1})$  as a joint randomized rounding of  $(a_{i,1}, a_{i+1,1})$ .

For all  $j \in [n - 1]$  use the fact that  $(b_{i,j}, b_{i+1,j})$  is a joint randomized rounding of  $(a_{i,j}, a_{i+1,j})$  to compute  $(b_{i,j+1}, b_{i+1,j+1})$  such that  $B_{R_{ij}}$  is a block randomized rounding of  $A_{R_{ij}}$ .

Clearly, the algorithm is correct if the roundings can be computed efficiently in the inner loop. To simplify notation, we assume  $i = 1$  and  $j = 1$ . Let  $a_{11}, a_{12}, a_{21}, a_{22} \in [0, 1]$ . Assume that  $(b_{11}, b_{21})$  is a joint randomized rounding of  $(a_{11}, a_{21})$ . In the remainder of this proof we show that there is a rounding  $(b_{12}, b_{22})$  of  $(a_{12}, a_{22})$  such that  $(b_{11}, b_{12}, b_{21}, b_{22})$  is a block randomized rounding of  $(a_{11}, a_{12}, a_{21}, a_{22})$ .

Put  $s := a_{11} + a_{12} + a_{21} + a_{22}$ . We first show that it is enough to consider the case  $s \leq 2$ . Assume  $s > 2$ . Let  $a'_{ij} := 1 - a_{ij}$  for all  $i, j \in [2]$  as well as  $b'_{11} = 1 - b_{11}$  and  $b'_{21} = 1 - b_{21}$ .

Then  $a'_{11} + a'_{12} + a'_{21} + a'_{22} < 2$  and  $(b'_{11}, b'_{21})$  is a joint randomized rounding of  $(a'_{11}, a'_{21})$ . Let  $(b'_{12}, b'_{22})$  be such that  $(b'_{11}, b'_{12}, b'_{21}, b'_{22})$  is a block randomized rounding of the corresponding  $a'$  values. Put  $b_{12} := 1 - b'_{12}$  and  $b_{22} := 1 - b'_{22}$ . Now  $(b_{11}, b_{12}, b_{21}, b_{22})$  is easily shown to be a block randomized rounding of  $(a_{11}, a_{12}, a_{21}, a_{22})$ .

Hence from now on let  $s \leq 2$ . We distinguish a number of cases:

**Case 1:**  $s \leq 1$ . If  $b_{11} = 1$  or  $b_{21} = 1$ , then put  $b_{12} := 0$  and  $b_{22} = 0$ . Otherwise choose  $b_{12}$  and  $b_{22}$  with probabilities

$$\begin{aligned} \Pr(b_{12} = 1 \wedge b_{22} = 0) &= a_{12}/(1 - a_{11} - a_{21}) \\ \Pr(b_{12} = 0 \wedge b_{22} = 1) &= a_{22}/(1 - a_{11} - a_{21}) \\ \Pr(b_{12} = 0 \wedge b_{22} = 0) &= 1 - (a_{12} + a_{22})/(1 - a_{11} - a_{21}). \end{aligned}$$

We compute that  $(b_{11}, b_{12}, b_{21}, b_{22})$  is a block randomized rounding. Note that the ‘otherwise’ case occurs only if  $b_{11} + b_{21} = 0$ . Since  $(b_{11}, b_{21})$  is a joint randomized rounding of  $(a_{11}, a_{21})$ , this happens with probability  $1 - a_{11} - a_{21}$ . Hence the probability that  $b_{12}$  becomes 1 is this probability of  $1 - a_{11} - a_{21}$  times the probability that  $b_{12}$  becomes 1 in the ‘otherwise’-case, which is  $a_{12}/(1 - a_{11} - a_{21})$ . Hence  $\Pr(b_{12} = 1) = a_{12}$  as required. Similarly, the remaining probabilities are proven to be correct.

**Case 2:**  $1 < s \leq 2$  and  $a_{ij} + a_{i'j'} \leq 1$  for all  $i, i', j, j' \in [2]$  such that either  $i = i'$  or  $j = j'$ . We compute  $b_{12}$  and  $b_{22}$  according to the rule

**If**  $b_{11} + b_{21} = 0$   
**then** put  $b_{12} := 1$  with probability  $\frac{a_{12}}{(a_{12}+a_{22})}$ ,<sup>3</sup> put  $b_{22} := 1 - b_{12}$   
**else if**  $b_{11} = 1$   
**then**  $b_{12} := 0$ ; put  $b_{22} := 1$  with probability  $\frac{a_{22}(s-1)}{(a_{12}+a_{22})a_{11}}$   
**else if**  $b_{21} = 1$   
**then**  $b_{22} := 0$ ; put  $b_{12} := 1$  with probability  $\frac{a_{12}(s-1)}{(a_{12}+a_{22})a_{21}}$ .

Using the fact that  $(b_{11}, b_{21})$  is a joint randomized rounding of  $(a_{11}, a_{21})$  we compute exemplary

$$\begin{aligned} \Pr(b_{12} + b_{22} = 1) &= \Pr(b_{11} + b_{21} = 0) + \Pr(b_{11} = 1) \Pr(b_{22} = 1 \mid b_{11} = 1) \\ &\quad + \Pr(b_{21} = 1) \Pr(b_{22} = 1 \mid b_{21} = 1) \\ &= (1 - a_{11} - a_{21}) + a_{11} \cdot \frac{a_{22}(s-1)}{(a_{12}+a_{22})a_{11}} + a_{21} \cdot \frac{a_{12}(s-1)}{(a_{12}+a_{22})a_{21}} \\ &= a_{12} + a_{22}. \end{aligned}$$

**Remaining cases.** Let us call  $e = (i, j, i', j') \in [2]^4$  an edge if either  $i = i'$  or  $j = j'$ . We call  $e$  heavy, if  $a_{ij} + a_{i'j'} > 1$  holds. Exploiting symmetry, we continue this case distinction

<sup>3</sup>We use the phrase “put  $y := 1$  with probability  $p$ ” to describe these actions: Generate a number  $r \in [0, 1]$  uniformly at random. If  $r \leq p$ , put  $y := 1$ , else put  $y := 0$ .

treating separately three cases such that there is one heavy edge, as well as two cases such that there are two (intersecting) heavy edges. Neither of these cases is particularly difficult. In fact, the conditional probabilities arising are uniquely determined, hence working them out is an easy exercise.  $\square$

Unfortunately, we could not find any more concise way of computing these roundings. But this is more or less an aesthetic problem: Concerning computing times, a distinction into many simple cases is rather preferable. Assuming the remaining cases proven, we have given a linear time algorithm computing continuous block randomized roundings.

We shall first analyze the error of continuous block randomized roundings and then turn to the computational aspects of this approach. We use the following variant of the Chernoff inequality, which can be derived for example from Appendix A of Alon and Spencer [1].

**Lemma 9.** *Let  $X_1, \dots, X_r$  be mutually independent random variables such that each two values of a fixed  $X_i$  differ by no more than  $d$ . Let  $X = \sum_{i \in [r]} X_i$ . Then*

$$\Pr(X > E(X) + \varepsilon) < \exp(-2\varepsilon^2/(d^2r))$$

holds for all  $\varepsilon \geq 0$ .

**Theorem 10.** *Let  $B$  be a continuous block randomized rounding of  $A \in [0, 1]^{m \times n}$ . Then*

- (i) *The expected rounding error satisfies  $E(d(A, B)) \leq 0.5463|\mathcal{R}|$ .*
- (ii) *If  $B^*$  is an optimal rounding, then  $E(d(A, B)) - d(A, B^*) \leq 0.3125|\mathcal{R}|$ .*
- (iii) *For all  $\varepsilon > 0$ ,  $\Pr(d(A, B) > E(d(A, B)) + \varepsilon|\mathcal{R}|) < 3 \exp(-\frac{1}{6}\varepsilon^2(m-3))$ .*

*Proof.* At least a half of all boxes (those  $R_{ij}$  where  $i$  is odd) are block randomized roundings. The remaining boxes contain two independent joint randomized roundings. Thus Lemma 3 and 6 yield the bounds (i) and (ii):

$$\begin{aligned} E(d(A, B)) &\leq \frac{1}{2} \cdot \frac{16}{27}|\mathcal{R}| + \frac{1}{2} \cdot \frac{1}{2}|\mathcal{R}| = \frac{59}{108}|\mathcal{R}| \\ E(d(A, B)) - d(A, B^*) &\leq \frac{1}{2} \cdot \frac{1}{2}|\mathcal{R}| + \frac{1}{2} \cdot \frac{1}{8}|\mathcal{R}| = \frac{5}{16}|\mathcal{R}|. \end{aligned}$$

To prove the large deviation bound, let  $X_i = \sum_{j=1}^{n-1} d(A_{R_{ij}}, B_{R_{ij}})$  for  $i \in [m-1]$ . Then  $X_i$  is a non-negative random variable bounded by  $X_i \leq 2(n-1)$ . Note that  $X_i$  is mutually independent of all  $X_{i'}$  such that  $|i-i'| \geq 3$ . For  $k \in [3]$  let  $I_k = \{i \in [m-1] \mid i \equiv k \pmod{3}\}$  and  $Y_k = \sum_{i \in I_k} X_i$ . From Lemma 9 we conclude

$$\begin{aligned} &\Pr(d(A, B) > E(d(A, B)) + \varepsilon(n-1)(m-1)) \\ &\leq \Pr(\exists k \in [3] : Y_k > E(Y_k) + \varepsilon(n-1)|I_k|) \\ &< \sum_{k \in [3]} \exp\left(\frac{-2\varepsilon^2(n-1)^2|I_k|^2}{4(n-1)^2|I_k|}\right) \\ &\leq 3 \exp\left(-\frac{1}{6}\varepsilon^2(m-3)\right). \end{aligned}$$

□

In the proof we estimated the rounding error by the worst case rounding errors for the two cases of joint and block randomized rounding. Both worst cases cannot occur simultaneously in all corresponding  $2 \times 2$  boxes. We do not lose a lot, however. In his diploma thesis [25], Schnieder shows that for the matrix  $A$  with all entries  $\frac{4-\sqrt{10}}{6} \approx 0.1396$ , a continuous block randomized rounding  $B$  satisfies  $E(d(A, B)) = \frac{40\sqrt{10}-112}{27}|\mathcal{R}| \approx 0.536708|\mathcal{R}|$ . In fact, this is asymptotically the worst case, as a tedious analysis reveals.

For the same reason, the relative bound is not sharp. Again the loss is not too large: Let  $A$  be such that  $a_{ij} = 0.4$ , if  $i \equiv 1, 2 \pmod{4}$ , and  $a_{ij} = 0.6$ , if  $i \equiv 3, 4 \pmod{4}$ . Then  $E(d(A, B)) = 0.3|\mathcal{R}| + d(A, B^*)$ , where  $B^*$  is an optimal rounding of  $A$ .

The large deviation bound (iii) displays one disadvantage of non-independent randomized rounding. Due to the dependencies, the rounding error cannot be written as sum of  $\Omega(|\mathcal{R}|)$  independent random variables. One might think that this problem could be overcome by replacing the Chernoff bound by martingale inequalities. The following example shows that the problem of the errors  $d(A_{R_{ij}}, B_{R_{ij}})$  not being independent is ‘real’:

Let  $A \in \{0, \frac{1}{2}\}^{4 \times n}$  such that  $a_{ij} = 0$  if and only if  $i \in \{2, 3\}$  and  $j$  is even. Let  $B$  be a continuous block randomized rounding of  $A$ . Then  $B$  is completely determined by  $b_{11}$  and  $b_{41}$ . In particular, we have  $d(A, B) = \frac{2}{3}|\mathcal{R}|$ , if  $b_{11} = b_{41}$ , and  $d(A, B) = \frac{1}{3}|\mathcal{R}|$ , if  $b_{11} \neq b_{41}$ . Note that both events occur with probability  $\frac{1}{2}$  each.

### 3.3 Algorithmic Properties of Continuous Block Randomized Rounding

Computing matrix roundings with the randomized approach above has some noteworthy advantages.

First of all, it is fast. Computing  $(b_{i,j+1}, b_{i+1,j+1})$  in Lemma 8 takes constant time. Therefore the whole matrix rounding can be done in linear time. The problem of computing time can be further addressed with parallel computing (and this is actually an issue when discussing digital halftoning algorithms). Since the roundings of the  $2 \times n$  double-rows are independent, it is no problem to assign them to different processors. From the viewpoint of application to digital halftoning, these points are crucial improvements over the algorithm of [2], that roughly has quadratic time complexity (ignoring a polylogarithmic factor). As stated in [2], this is too slow for high-resolution images.

Our algorithm can be derandomized with the method of conditional probabilities. Since the arising conditional probabilities can be computed efficiently, we do not need pessimistic estimators. The computation of the conditional probabilities can be arranged in a way that

the resulting algorithm has linear time complexity. We refer to Srivastav [31] for a recent survey on derandomization issues.

Another advantage is diversity. Suppose that we do not want to find a good rounding with respect to the  $2 \times 2$  boxes, but with respect to  $3 \times 3$  boxes (or with respect to both). We currently have no hint whether the error with respect to these sets is a better measure for the visual quality of the resulting digital halftoning, but it seems plausible to try this experimentally. Hence we need an algorithm computing such roundings.

The algorithm of Asano et al. seems not to work very well for this problem. The roundings computed by their algorithm may have error up to  $(1.5 - 9\varepsilon)|\mathcal{R}|$  with respect to the  $3 \times 3$  boxes. This is shown by a matrix  $A$  with entries  $\frac{1}{6} - \varepsilon$  only, which may be rounded to the all-zero matrix. We are also pessimistic that their approach in general can be extended to  $3 \times 3$  boxes or other larger structures.

Contrary to these difficulties, non-independent rounding does very well: We may even use a similar rounding as before: Assuming  $m, n$  even for simplicity, compute  $B$  from  $A$  by letting  $B_{R_{ij}}$  be a block randomized rounding of  $A_{R_{ij}}$  independently for all odd  $i \in [m-1]$  and odd  $j \in [n-1]$ . We call such an  $A$  an independent block randomized rounding. Now each  $3 \times 3$  box contains exactly one block randomized rounding, two joint randomized roundings and one single randomized rounding. Since the four values of the block randomized rounding in total behave like a single randomized rounding, and so do the two values of each joint randomized rounding, the expected error of a  $3 \times 3$  box is just given by Theorem 1. We have

**Theorem 11.** *With respect to the family  $\mathcal{R}_3$  of  $3 \times 3$  boxes, an independent block randomized rounding  $B$  of  $A$  has expected error  $d_{\mathcal{R}_3}(A, B) \leq 0.82944|\mathcal{R}_3|$ .*

We presented Theorem 11 to demonstrate how easily the ideas of this section can be applied to other rounding problems as well. There are different non-independent randomized roundings achieving slightly better bounds. The reader is encouraged to try to find some on his own. The following characterization might be helpful for this purpose.

## 4 Proof of the Characterization

In this section we put the results presented before into a more general framework and prove the characterization proposed in the introduction. As should be clear by now, we are looking for randomized roundings that are also ‘good’ with respect to some sets (i.e. sums) of variables. It is convenient to describe such structures through hypergraphs: A *hypergraph* is a pair  $\mathcal{H} = (V, \mathcal{E})$ , such that  $V$  is a finite set and  $\mathcal{E} \subseteq 2^V$ . The elements of  $V$  are called *vertices*, those of  $\mathcal{E}$  *hyperedges* or *edges* for short. In our setting,  $V$  will always be some

index set the for the variables and  $\mathcal{E}$  contains those sets of variables where we want our roundings to be ‘good’ on.

**Definition 12.** Let  $\mathcal{H} = (I, \mathcal{E})$  be a hypergraph. Let  $x_i \in \mathbb{R}$  for  $i \in I$ . Let  $y_i, i \in I$ , be random variables. For  $E \subseteq I$  put  $x_E := \sum_{i \in E} x_i$  and  $y_E := \sum_{i \in E} y_i$ . We call  $y_E$  a randomized rounding of  $x_E$ , if

$$\begin{aligned} \Pr(y_E = \lfloor x_E \rfloor + 1) &= \{x_E\} \\ \Pr(y_E = \lfloor x_E \rfloor) &= 1 - \{x_E\} \end{aligned}$$

holds. We say that  $(y_i)_{i \in I}$  is a randomized rounding of  $(x_i)_{i \in I}$  with respect to the dependency hypergraph  $\mathcal{H}$  (or a randomized  $\mathcal{H}$ -rounding) if  $y_E$  is a randomized rounding of  $x_E$  for all  $E \in \mathcal{E}$  and  $y_i$  is a randomized rounding of  $x_i$  for all  $i \in I$ .

In this language, continuous block randomized roundings are randomized roundings with respect to the hypergraph  $\mathcal{H} = ([m] \times [n], \mathcal{E})$  and

$$\begin{aligned} \mathcal{E} &= \{R_{ij} \mid i \in [m-1] \text{ odd}, j \in [n-1]\} \\ &\cup \{\{(i, j), (i, j+1)\} \mid i \in [m], j \in [n-1]\} \\ &\cup \{\{(i, j), (i+1, j)\} \mid i \in [m-1] \text{ odd}, j \in [n]\}. \end{aligned}$$

A hypergraph is said to be *totally unimodular* if its incidence matrix is totally unimodular. Recall that a matrix is totally unimodular if each square submatrix has determinant  $-1$ ,  $0$  or  $1$ . There is a rather complicated recursive characterization for totally unimodular matrices and hypergraphs due to Seymour [26, 27]. For our purposes, an easier, though fundamental, result is sufficient:

**Theorem 13 (Ghouila-Houri [12]).** A hypergraph  $\mathcal{H} = (V, \mathcal{E})$  is totally unimodular if and only if for each subset  $V_0$  of vertices there is a partition  $V_1 \dot{\cup} V_2 = V_0$  such that any hyperedge  $E$  satisfies  $||E \cap V_1| - |E \cap V_2|| \leq 1$ .

Let us give some intuition to this result. For  $V_0 \subseteq V$ ,  $\mathcal{H}|_{V_0} = (V_0, \{E \cap V_0 \mid E \in \mathcal{E}\})$  is an induced subhypergraph of  $\mathcal{H}$ . We call a hypergraph  $\mathcal{H} = (V, \mathcal{E})$  perfectly balanced, if there is a partition  $V_1 \dot{\cup} V_2 = V$  of its vertex set such that  $||E \cap V_1| - |E \cap V_2|| \leq 1$  holds for all  $E \in \mathcal{E}$ , i.e., apart from single vertices of odd-cardinality hyperedges each hyperedge contains the same number of vertices in  $V_1$  and  $V_2$ . In this language, the Ghouila-Houri theorem states that  $\mathcal{H}$  is totally unimodular if and only if each induced subhypergraph is perfectly balanced. Both from the definition and (easier) from Ghouila-Houri’s characterization one can deduce that  $\mathcal{H}$  is totally unimodular if and only if  $(V, \mathcal{E} \cup \{\{v\} \mid v \in V\})$  is totally unimodular. A second result we use is the well-known theorem of Hoffman and Kruskal:

**Theorem 14 (Hoffman, Kruskal [14]).** Let  $A \in \mathbb{R}^{m \times n}$  be a totally unimodular matrix. Let  $b, b' \in \mathbb{Z}^m$  and  $c, c' \in \mathbb{Z}^n$ . Then

$$\{x \in \mathbb{R}^n \mid b \leq Ax \leq b', c \leq x \leq c'\}$$

is an integral polyhedron.

Hoffman and Kruskal also showed a converse result: An integral matrix  $A$  is totally unimodular if the polyhedron  $\{x \in \mathbb{R}^n \mid x \geq 0, Ax \leq b\}$  is integral for all  $b \in \mathbb{Z}^m$ . But we shall not need that much understanding of integral polyhedra. Basically, it suffices to know the above theorem and the fact that any bounded polyhedron is the convex hull of its extremal points. Thus for any  $x \in P$  there are  $k \leq n + 1$ ,  $w^{(1)}, \dots, w^{(k)} \in \text{ex}(P)$ ,  $\lambda_1, \dots, \lambda_k \in [0, 1]$  such that  $\sum_{\ell \in [k]} \lambda_\ell = 1$  and  $x = \sum_{\ell \in [k]} \lambda_\ell w^{(\ell)}$ . This is the well-known Carathéodory theorem (see Eckhoff [10] for a survey).

**Theorem 15.** *Let  $\mathcal{H} = (I, \mathcal{E})$  be a hypergraph. The following two properties are equivalent:*

- (i) *For all  $(x_i)_{i \in I}$  there is a randomized  $\mathcal{H}$ -rounding  $(y_i)_{i \in I}$ .*
- (ii)  *$\mathcal{H}$  is totally unimodular.*

*Proof.* Since the vertex set of  $\mathcal{H}$  is not relevant, we may conveniently assume that  $I = [n]$ . Let  $\mathcal{H}$  be totally unimodular and  $x_1, \dots, x_n \in \mathbb{R}$ . Without loss of generality (cf. the remark following Theorem 13), we may assume that  $\{j\} \in \mathcal{E}$  for all  $j \in [n]$ . Let  $A$  be an incidence matrix of  $\mathcal{H}$ , i.e., fix an enumeration  $\mathcal{E} = \{E_1, \dots, E_m\}$  of the hyperedges and define  $A = (a_{ij})$  by  $a_{ij} = 1$ , if  $j \in E_i$ , and  $a_{ij} = 0$  otherwise. Put  $P = \{w \in \mathbb{R}^n \mid \lfloor Ax \rfloor \leq Aw \leq \lceil Ax \rceil\}$ . By definition,  $x \in P$ . Since the  $n \times n$  identity matrix is a submatrix of  $A$ ,  $P$  is bounded.

Thus,  $x$  is a convex combination of the extremal point of  $P$ : There are  $w^{(1)}, \dots, w^{(k)} \in \text{ex}(P)$ ,  $\lambda_1, \dots, \lambda_k \in [0, 1]$  such that  $\sum_{\ell \in [k]} \lambda_\ell = 1$  and  $x = \sum_{\ell \in [k]} \lambda_\ell w^{(\ell)}$ . By the theorem of Hoffman and Kruskal, all  $w^{(\ell)}$  are integral, and in our case we even have  $w_j^{(\ell)} \in \{\lfloor x_j \rfloor, \lceil x_j \rceil\}$ . Define the random variables  $y_1, \dots, y_n$  by putting  $y = w^{(\ell)}$  with probability  $\lambda_\ell$ .

We compute that  $y_1, \dots, y_n$  is a randomized rounding of  $x_1, \dots, x_n$  with respect to  $\mathcal{H}$ : Let  $E = E_i \in \mathcal{E}$ . If  $x_E$  is integral, then  $(Aw^{(\ell)})_i = x_E$  for all  $\ell \in [k]$  by definition of  $P$ . Hence  $y_E = (Ay)_i$  trivially is a randomized rounding of  $x_E$ . Assume that  $x_E$  is non-integral. We have  $x_E = \sum_{\ell \in [k]} \lambda_\ell (Aw^{(\ell)})_i$ . Put  $L_E^+ = \{\ell \in [k] \mid (Aw^{(\ell)})_i = \lceil x_E \rceil\}$ . Since  $(Aw^{(\ell)})_i \in \{\lfloor x_E \rfloor, \lceil x_E \rceil\}$  for all  $\ell \in [k]$ , we conclude that  $\{x_E\} = \sum_{\ell \in L_E^+} \lambda_\ell$ . Thus  $\Pr(y_E = \lceil x_E \rceil) = \sum_{\ell \in L_E^+} \lambda_\ell = \{x_E\}$ . As  $y_E$  takes only the two values  $\lfloor x_E \rfloor$  and  $\lceil x_E \rceil$ , we also have  $\Pr(y_E = \lfloor x_E \rfloor) = 1 - \{x_E\}$ . This shows that  $y_1, \dots, y_n$  is a randomized  $\mathcal{H}$ -rounding of  $x_1, \dots, x_n$ .

Now assume that  $\mathcal{H}$  is not totally unimodular. By the theorem of Ghouila-Houri [12] some induced subhypergraph of  $\mathcal{H}$  has discrepancy at least two, i.e., there is a  $V_0 \subseteq [n]$  such that for any 2-partition  $V_1 \dot{\cup} V_2 = V_0$  there is an  $E \in \mathcal{E}$  such that  $||E \cap V_1| - |E \cap V_2|| \geq 2$ . Put  $x_j = 0$  for all  $j \in [n] \setminus V_0$  and  $x_j = \frac{1}{2}$  for all  $j \in V_0$ . Let  $y_1, \dots, y_n$  be a randomized  $\mathcal{H}$ -rounding of  $x_1, \dots, x_n$ . Then  $y_j = 0$  holds with probability one for all  $j \in [n] \setminus V_0$  by definition of randomized  $\mathcal{H}$ -rounding. For  $j \in V_0$ ,  $y_j$  may take the values 0 and 1 only. Let  $\tilde{y}$  be a possible outcome of the underlying random experiment, that is, an image of the random variable  $y$  having probability greater zero. Let  $V_1 = \{j \in V_0 \mid \tilde{y}_j = 0\}$  and

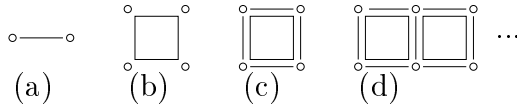


Figure 1: Sample diagrams

$V_2 = \{j \in V_0 \mid \tilde{y}_j = 1\}$ . Let  $E \in \mathcal{E}$  such that  $\left| |E \cap V_1| - |E \cap V_2| \right| \geq 2$ . Then  $\tilde{y}_E = |E \cap V_2|$ , whereas  $x_E = \frac{1}{2}|E \cap V_0|$ . Thus we have  $|x_E - \tilde{y}_E| \geq 1$ . Since  $|x_E - y_E| \geq 1$  holds with non-zero probability,  $y_E$  is not a randomized rounding of  $x_E$ . This is a contradiction to our assumption that  $y_1, \dots, y_n$  is a randomized  $\mathcal{H}$ -rounding of  $x_1, \dots, x_n$ .  $\square$

The use of linear programming ideas to deal with dependencies resembles the work of Koller and Megiddo [18] and Karger and Koller [15]. However, these results are of a completely different nature. There, the aim is to cope with dependencies, since they can lead to smaller sample spaces, whereas we try to use dependencies. There, linear programming ideas are used to transform a given sample space satisfying some dependencies into a smaller one where the dependencies still hold, whereas we construct a sample space directly from a set of given dependencies. Nevertheless, all results indicate that linear programming is a useful framework when dealing with probabilistic constraints

## 5 Alternative Randomized $\mathcal{H}$ -roundings

In this section we investigate what can be achieved by using different randomized  $\mathcal{H}$ -roundings for the approximation problem of Asano et al. [2]. We omit the proofs for reasons of space, but let us stress that these results heavily depend on Theorem 15: It allows to decide whether a randomized  $\mathcal{H}$ -rounding exists or not by simple checking whether  $\mathcal{H}$  is totally unimodular or not. Thus we may postpone the work of designing a rounding algorithm (and the proof of Lemma 8 indicates that this might be quite tedious) until we found a dependency hypergraph  $\mathcal{H}$  yielding suitable bounds.

We denote the hypergraphs regarded through simple diagrams. We assume that all variables are randomized roundings. If two variables form a joint randomized rounding, we denote this by connecting the respective nodes by an edge (Figure 1(a)). If four variables are rounded together (i.e., they form a hyperedges in  $\mathcal{H}$ ), we connect them by a square (Figure 1(b)). Thus a single block randomized rounding is depicted by the diagram in Figure 1(c).

We shall always assume that the grid is plastered with the pattern, and further, that nodes not connected through a series of hyperedges are rounded independently. Hence Figure 1(c) also denotes the independent block randomized rounding investigated in Theorem 11. Repeated overlapping patterns will be indicated through (ellipsis) dots. Figure 1(d) thus describes the hypergraph used in Section 3.

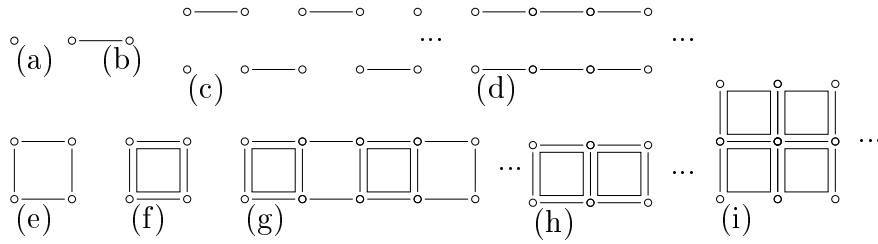


Figure 2: Totally unimodular hypergraphs

	absolute	relative
(a)	0.8295	0.75
(b)	0.7111	0.625
(c)	0.75	0.5926
(d)	0.5926	0.5
(e)	0.6518	0.5625
(f)	0.6287	0.4688
(g)	0.5695	0.4063
(h)	0.5493	0.3125
(i)	0.5617	0.4166

Figure 3: Error bounds of the  $\mathcal{H}$ -roundings depicted in Fig. 2

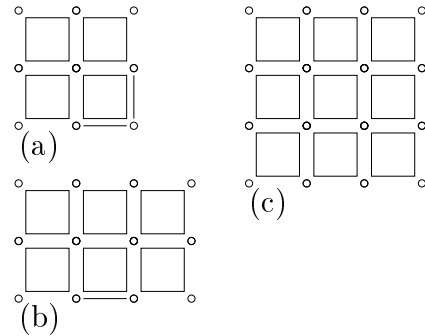


Figure 4: Hypergraphs that are not totally unimodular

In Figure 2 we show some totally unimodular hypergraphs. In Figure 3 we give upper bounds for the errors inflicted by applying the corresponding  $\mathcal{H}$ -roundings to Asano's approximation problem. The middle column contains an upper bound for the expected absolute error  $E(d(A, B))$ , the right-side one an upper bound for the expected loss  $E(d(A, B)) - d(A, B^*)$  compared to an optimal rounding  $B^*$ . These upper bound are derived using the methods of Section 3. They are either sharp or close to the optimum.

Figure 3 shows several interesting facts. We see that already simple dependencies yield a significant improvement over the independent case (a). Nevertheless, the right choice of dependencies is important: While (b) and (c) look almost identical — plastering the grid with independent joint randomized roundings — the resulting error bounds are not the same. Better bounds can be achieved by larger structures than just independent joint randomized roundings. On the other hand, structures larger than (h) either yield worse bounds as (i), or cannot be realized, since the corresponding hypergraph is not totally unimodular. Some examples of the latter type are shown in Figure 4. As Figure 4(c) shows, the dependencies that all boxes are already randomized roundings are not realizable. Thus a non-independent randomized rounding having  $E(d(A, B)) \leq \frac{1}{2}|\mathcal{R}|$  seems not to exist. Currently we do not know the truth in between our upper bound of  $0.5463|\mathcal{R}|$  and the lower one of  $\frac{1}{2}|\mathcal{R}|$ .

There is some caution necessary in rows (e), (g), and (i). These three patterns generate  $\mathcal{R}$ -boxes that are not sums of independent randomized roundings. In (e) the error inflicted in the box containing the four in a cycle manner connected nodes depends on the precise way this rounding is generated. For our estimate we assumed that the non-intersecting joint randomized roundings neither display a positive nor a negative correlation. This is justified by the fact that a rounding exists in which the errors are in fact likely to cancel (namely (f)). An analogous statement holds for (g) and (h), but not for (i). Here Figure 4 tells us further restrictions on the random process are not possible. Therefore our upper bound here assumes that we have a bad correlation among the horizontally adjacent nodes that do not form a joint randomized rounding.

Note that such negative correlation can indeed occur: Assume three nodes 1, 2 and 3 such that 1 and 2 as well as 2 and 3 form a joint randomized rounding. Let  $x_1 = 1 - x_2 = x_3$ . Then any non-independent randomized rounding with respect to these dependencies has  $y_1 = y_3$ . Thus the error  $|(x_1 + x_3) - (y_1 + y_3)|$  has not the distribution of two independent randomized roundings, but a worse one, namely twice the one of a single randomized rounding.

## 6 Summary and Outlook

This paper describes a new approach in randomized rounding. By imposing suitable dependencies, we improve the expected rounding error significantly. For a problem arising in digital halftoning, this improves previous algorithms both according to run-time and rounding error. In particular, we presented the first algorithm solving this problem fast enough for practical application, namely in linear time.

On the methodological side, this paper shows that non-independent randomized rounding can be very effective if one succeeds in finding the right dependencies. To this end, we give a complete characterization of realizable dependencies.

From this work, some open questions arise:

- What is the true approximability of the matrix rounding problem suggested by Asano et al., i.e., close the gap between the lower bound of  $0.5|\mathcal{R}|$  and the upper one stemming from our methods.
- For which further rounding problems is it possible translate part of the structure of the problem into suitable dependencies for the random experiment?
- Engineering aspects: How do the roundings computed by our algorithm look to the human eye? Asano et al. proposed the idea that low error roundings with respect to  $2 \times 2$  boxes look nice, but of course they could check this only with their algorithm.

Therefore, theoretically, our roundings might even look worse. Coming from the other end, one might (and actually should) investigate other error measures. Using the ideas of this paper, one may also design rounding algorithms that optimize different error measures stemming from larger boxes or even completely different geometric structures. It could well be that such error measure are better suited to distinguish good halftonings from worse ones.

## Acknowledgments

I would like to thank Tetsuo Asano, Naoki Katoh, Koji Obokata and Takeshi Tokuyama for providing me with a preprint of their SODA 2002 paper and some useful discussion on the topic. I am also grateful that Henning Schnieder gave his kind permission to use some of his numerical results and sample images in the appendix of this paper.

## References

- [1] N. Alon and J. H. Spencer. *The Probabilistic Method*. John Wiley & Sons, Inc., 2nd edition, 2000.
- [2] T. Asano, N. Katoh, K. Obokata, and T. Tokuyama. Matrix rounding under the  $L_p$ -discrepancy measure and its application to digital halftoning. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 896–904, 2002.
- [3] T. Asano, T. Matsui, and T. Tokuyama. On the complexities of the optimal rounding problems of sequences and matrices. In *Algorithm Theory—SWAT 2000 (Bergen)*, volume 1851 of *Lecture Notes in Computer Science*, pages 476–489, Berlin, 2000. Springer.
- [4] T. Asano, T. Matsui, and T. Tokuyama. Optimal roundings of sequences and matrices. *Nordic J. Comput.*, 7:241–256, 2000.
- [5] B. E. Bayer. An optimum method for two-level rendition of continuous-tone pictures. In *Conference Record, IEEE International Conference on Communications*, volume 1, pages (26–11)–(26–15). IEEE, 1973.
- [6] D. Bertsimas, C. Teo, and R. Vohra. On dependent randomized rounding algorithms. *Oper. Res. Lett.*, 24:105–114, 1999.
- [7] B. Doerr. Lattice approximation and linear discrepancy of totally unimodular matrices. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 119–125, 2001. To appear in *Combinatorica*.

- [8] B. Doerr. Non-independent randomized rounding. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 506–507, 2003.
- [9] B. Doerr and H. Schnieder. Non-independent randomized rounding and an application to digital halftoning. In R. Möhring and R. Raman, editors, *Proceedings of ESA 2002*, volume 2461 of *Lecture Notes in Computer Science*, pages 399–410, Berlin–Heidelberg, 2002. Springer Verlag.
- [10] J. Eckhoff. Helly, Radon, and Carathéodory type theorems. In *Handbook of Convex Geometry, Vol. A, B*, pages 389–448. North-Holland, Amsterdam, 1993.
- [11] R. W. Floyd and L. Steinberg. An adaptive algorithm for spatial grey scale. In *SID 75 Digest*, pages 36–37. Society for Information Display, 1975.
- [12] A. Ghouila-Houri. Caractérisation des matrices totalement unimodulaires. *C. R. Acad. Sci. Paris*, 254:1192–1194, 1962.
- [13] W. M. Goodall. Television by pulse code modulation. *Bell Syst. Tech. J.*, 30:33–49, 1951.
- [14] A. J. Hoffman and J. B. Kruskal. Integral boundary points of convex polyhedra. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 223–246. 1956.
- [15] D. R. Karger and D. Koller. (De)randomized construction of small sample spaces in NC. *J. Comput. System Sci.*, 55:402–413, 1997.
- [16] D. E. Knuth. Digital halftones by dot diffusion. *ACM Trans. Graphics*, 6:245–273, 1987.
- [17] D. E. Knuth. Two-way rounding. *SIAM J. Discrete Math.*, 8:281–290, 1995.
- [18] D. Koller and N. Megiddo. Constructing small sample spaces satisfying given constraints. *SIAM J. Discrete Math.*, 7:260–274, 1994.
- [19] D. L. Lau and G. R. Acre. *Modern Digital Halftoning*, volume 8 of *Signal Processing and Communication*. Marcel Dekker, Inc., New York, Basel, 2001.
- [20] D. J. Lieberman and J. P. Allebach. Efficient model based halftoning using direct binary search. In *Proceedings of the 1997 IEEE International Conference on Image Processing*, volume 1, pages 775–778. IEEE Comput. Soc, 1997.
- [21] B. Lippel and M. Kurland. The effect of dither on luminance quantization of pictures. In *IEEE Trans. Commun. Tech. COM-19*, pages 879–888, 1971.
- [22] P. Raghavan. Probabilistic construction of deterministic algorithms: Approximating packing integer programs. *J. Comput. Syst. Sci.*, 37:130–143, 1988.

- [23] P. Raghavan and C. D. Thompson. Randomized rounding: A technique for provably good algorithms and algorithmic proofs. *Combinatorica*, 7:365–374, 1987.
- [24] L. G. Roberts. Picture coding using pseudo-random noise. *IRE Trans. Infor. Theory*, IT-8:145–154, 1962.
- [25] H. Schnieder. *Digital Halftoning through Non-independent Randomized Rounding*. Diploma Thesis, Christian–Albrechts–Universität zu Kiel, 2003.
- [26] P. D. Seymour. Decomposition of regular matroids. *J. Combin. Theory Ser. B*, 28:305–359, 1980.
- [27] P. D. Seymour. Applications of the regular matroid decomposition. In *Matroid theory (Szeged, 1982)*, pages 345–357. North-Holland, Amsterdam, 1985.
- [28] A. Srinivasan. Approximation algorithms via randomized rounding: A survey. In M. Karonski, editor, *Lectures on approximation and randomized algorithms. Proceedings of the Berlin-Poznan summer school*, Advanced Topics in Mathematics, pages 9–71, Warsaw, 1999. Polish Scientific Publishers.
- [29] A. Srinivasan. Improved approximation guarantees for packing and covering integer programs. *SIAM J. Comput.*, 29:648–670, 1999.
- [30] A. Srinivasan. Distributions on level-sets with applications to approximations algorithms. In *Proc. 41th Ann. IEEE Symp. on Foundations of Computer Science*, pages 588–597, 2001.
- [31] A. Srivastav. Derandomization in combinatorial optimization. In S. Rajasekaran, P. M. Pardalos, J. H. Reif, and J. D. P. Rolim, editors, *Handbook of Randomized Computing*, chapter 18. Kluwer Academic Publishers, 2001.
- [32] J. R. Sullivan, L. A. Ray, and R. Miller. Design of minimum visual modulation halftone patterns. *IEEE Transactions on Systems, Man, and Cybernetics*, 21:33–38, 1991.
- [33] R. Ulichney. *Digital Halftoning*. MIT Press, 1987.
- [34] R. Ulichney. Dithering with blue noise. *Proc. IEEE*, 76:56–79, 1988.

## Appendix

This paper is intended to be of theoretical nature. Since one of the referees felt that readers from a more applied background would severely miss images and experimental data, we do our best to provide some in this appendix.

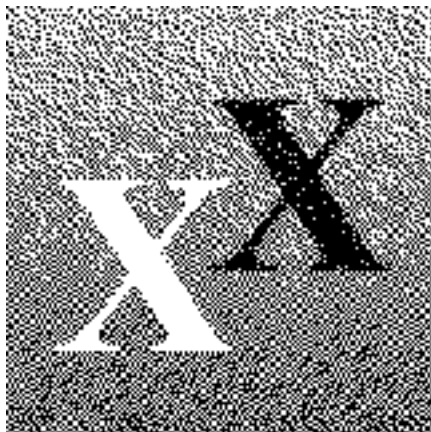
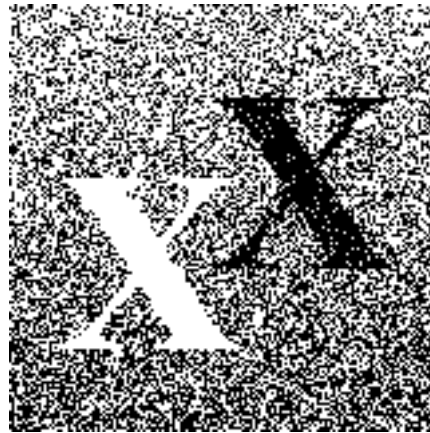
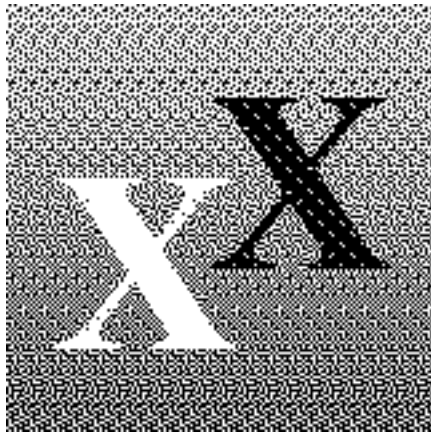
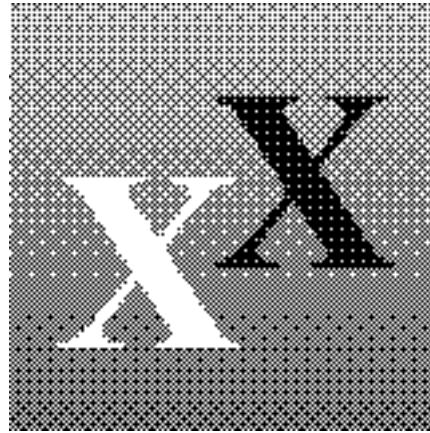
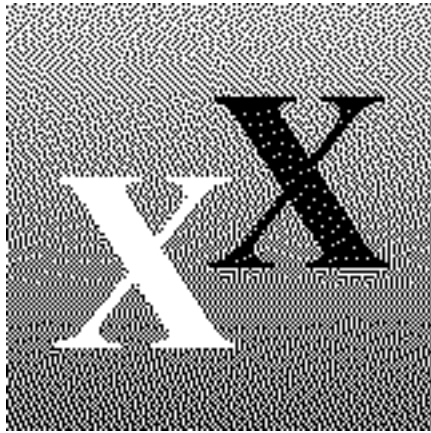


Figure 5: Halftoning results of different standard algorithms and ours:

Top left: Error diffusion.

Top right: Ordered dither.

Middle left: Dot diffusion.

Middle right: Random dither (randomized rounding).

Bottom left: Non-independent randomized rounding.

We applied the three classical algorithms mentioned in the introduction together with independent randomized rounding and our algorithm to several images. All image data used 1 byte per pixel resulting in an integer value between 0 and 255. We used two types of input data: Real-world images taken with a digital camera (not shown in this paper) and artificial images produced with a commercial imaging software. Naturally, the first type is more suitable to estimate how well the algorithm performs in real-world applications, whereas the second is better suited to demonstrate the particular strengths and weaknesses of an algorithm.

The images displayed in Fig. 5 are obtained from halftoning an artificially generated image of size of  $160 \times 160$  pixels with the standard algorithms of error diffusion, ordered dither, dot diffusion, random dither (independent randomized rounding) and our non-independent randomized rounding algorithm. In all cases, the standard form of the algorithms has been used. It is known that some algorithms produce better outputs if the inputs are subject to some preprocessing. Since such data is mainly available for the algorithms of error diffusion and ordered dither, we believe that an unbiased comparison might be easier without any preprocessings. As indicated in the first paragraph, the intention of this appendix is merely to give the interested reader an idea of how our ideas could work in practise rather than giving evidence that they are competitive with previous algorithms.

Fig. 5 shows that the randomized nature of our algorithm has a positive effect on the generation of unwanted structures and grains. Unwanted structures include all kinds of regular patterns like snakes, crosses or labyrinths that attract objectionable attention. In particular error diffusion and ordered dither algorithm tend to produce those.

Grains emerge, if in dark (respectively light) parts of the picture two or more white (respectively black) pixels touch and thus build a recognizable block. Randomized rounding is very vulnerable to this problem, which is why it is not used in practice for digital halftoning. On the other end we find error diffusion, which hardly produces any grains. It seems that algorithms that are good concerning graininess tend to produce unwanted structures and vice versa. In this sense, non-independent randomized rounding could be a fair compromise: Being by far less grainy than independent randomized rounding on the one hand, it is unlikely to produce unwanted structures on the other.

We also computed the actual errors generated by the algorithms on several input images

Method	Total	$3 \times 3$	$2 \times 2$
Error Diffusion	0.00027	0.45	0.35
Ordered Dither	0.00052	0.65	0.34
Dot Diffusion	0.00024	0.65	0.43
Randomized Rounding	0.00056	0.96	0.64
Non-independent RR	0.00034	0.60	0.40

Table 1: Experimental average errors of the algorithms

(on roughly 300,000 pixels). The averages values are given in Table 1. The error measures regarded are the following.

**Total error:** The absolute difference in the average intensity of input and output image  $\frac{1}{mn}|\sum_{i,j}(a_{ij} - b_{ij})|$ .

$2 \times 2$  **error:** The average error in the  $2 \times 2$  boxes, that is,  $\frac{1}{|\mathcal{R}|}d(A, B)$ , where  $d(A, B)$  is the error measure proposed by Asano et al.

$3 \times 3$  **error:** The average error in the  $3 \times 3$  boxes.

As can be seen, all algorithms change the average intensity only minimally. Concerning the errors in the  $2 \times 2$  and  $3 \times 3$  boxes, the case is more interesting. Whereas for the  $2 \times 2$  boxes error diffusion is slightly and non-independent randomized rounding is significantly worse than the best performer ordered dither, things change for  $3 \times 3$  boxes. Here error diffusion takes the lead, followed by non-independent randomized rounding. Together with the visual quality of the output images, this poses the questions whether larger boxes might yield a better similarity measure. The fact that error diffusion performs relatively good also asks for theoretical bounds for the errors generated with respect to the  $2 \times 2$  boxes. Here the inherently sequential nature of this algorithm seems to impose some difficulties.