Interleaving Schemes on Circulant Graphs with Two Offsets

Aleksandrs Slivkins
Department of Computer Science
Cornell University
Ithaca, NY 14853
slivkins@cs.cornell.edu

Jehoshua Bruck
Department of Electrical Engineering
California Institute of Technology
Mail Code 136-93, Pasadena, CA 91125
bruck@paradise.caltech.edu

August 29, 2003

Abstract

Interleaving is used for error-correcting on a bursty noisy channel. Given a graph G describing the topology of the channel, we label the vertices of G so that each label-set is sufficiently sparse. Interleaving scheme corrects for any error burst of size at most t; it is a labeling where the distance between any two vertices in the same label-set is at least t.

We consider interleaving schemes on infinite circulant graphs with two offsets 1 and d. In such graph the vertices are integers; edge ij exists if and only if $|i-j| \in \{1, d\}$. Our goal is to minimize the number of labels used.

Our constructions are covers of the graph by the minimal number of translates of some label-set S. We focus on minimizing the *index* of S, which is the inverse of its density rounded up. We establish lower bounds and prove that our constructions are optimal or almost optimal, both for the index of S and for the number of labels.

1 Introduction

Error-correcting codes work best when the errors are scattered. Since errors on noisy channels are often bursty, *interleaving* is used. The idea is to assign data points to a number of separate codes, so that the points assigned to the same code are less likely to be hit by the same error burst. The goal is to minimize the transmission overhead, which is proportional to the number of distinct codes. For a simple example, suppose we transmit a stream of bits using parity bits for error-correcting. Furthermore, suppose we know that error bursts are quite rare, but a single burst can damage up to three consecutive bits. So we split the bits into three sets as $\{123123\ldots\}$ and compute parity bits separately for each set.

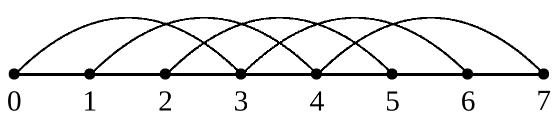
The way we interleave the codes largely depends on the topology of a noisy channel. Many noisy channels are one-dimensional, time being the only dimension. 2D noisy channels occur in optical recording [6], charged-coupled devices, 2D barcodes [5], and information hiding in digital images and video sequences. A holographic data storage system can be viewed as a 3D noisy channel [1].

Interleaving schemes. Most work on interleaving concentrated on 2D rectangular or circular error bursts (see [1] for references). The present paper takes after [1] in that it considers *arbitrary* error bursts of a given size t. In other words, our goal is to make sure that no error burst of size t or less contains two data points assigned to the same code.

Formally the topology of a noisy channel is given by a graph G on transmitted data points, so that two data points are likely to be hit by the same error burst if and only if they are close to each other in G. Error bursts are then modeled as connected subgraphs of G. Therefore we have the following labeling problem: given a graph G and an integer t, construct a labeling of G so that no connected subgraph of size t contains two vertices labeled the same, or, equivalently, the distance between any two vertices in a label-set is at least t. Such a labeling is called a *t-interleaving scheme,* where t is an *interleaving parameter*. The goal is to minimize *interleaving degree,* the number of distinct labels used. Note that for $t = 2$ it is just the graph-coloring problem.

History and the present scope. The history of the work on interleaving schemes is rather brief. Blaum et al. [1] introduced interleaving schemes and analyzed them on two- and three-dimensional arrays. The follow-up paper [2] generalized interleaving schemes to those with repetitions, where in any connected cluster of size t any label is repeated at most r times. Asymptotically optimal constructions on 2D arrays were presented for the case $r = 2$. Etzion and Vardy [4] considered the case $r > 2$.

In this paper we extend interleaving schemes beyond arrays. We consider a similar but substantially different topology, namely infinite circulant graphs G_d with two offsets $\{1, d\}$. The vertices of G_d are integers; an edge ij exists if and only if $|i - j| \in \{1, d\}$ (Fig. 1a). G_d is essentially a 2D-array of width d with a few extra edges (Fig. 1b). These 'extra edges', however, break the constructions from [1], thus making our problem interesting. For us the problem is more combinatorial than practical; we are especially interested in whether and when our constructions are *exactly* optimal.



(a) G_3 as a line

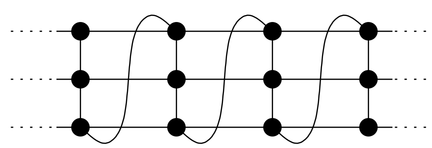

(b) G_3 as a 2D array with a few extra edges

Figure 1: G_3, the infinite circulant graph with two offsets $\{1, 3\}$.

Our approach and results. An interleaving scheme is a partition of the graph into label-sets. A set S can be a label-set if and only if the distance between any two points of S is at least t. Call such sets *t-sparse*. Call S *periodic* with a period p if it is that case that each n lies in S iff $n + p$ does. Say an interleaving scheme on G_d is *periodic* with a period p if each n is labeled the same as $n + p$.

We define the *density* of a periodic set S as the number of points within the period over the length of the period. We extend this definition to non-periodic sets as a limit $\frac{1}{2}s_n/n$ where s_n is the number of elements of S within the interval $[-n, n]$, whenever such limit exists. The *index* of S is the inverse of its density rounded up. Note that the index of S gives a lower bound on the number of copies of S needed to cover $\mathbb{Z}$. Accordingly, a lower bound on the index of a t-sparse set is a lower bound on the interleaving degree of a periodic interleaving scheme.

Our approach is to cover $\mathbb{Z}$ with copies of a periodic t-sparse set. First we find a t-sparse set with a minimal index, then we cover $\mathbb{Z}$ with a minimal number of copies thereof. The resulting interleaving degree is optimal or close to optimal. Most of our progress is on minimizing the index of a t-sparse set, which is itself an interesting combinatorial problem.

There are three cases which require separate constructions and lower bounds. Let $\delta = \lceil d/2 \rceil$.

- For $t \geq d - 1$ there is a simple unique optimal interleaving scheme and t-sparse set.
- For $t \leq \delta$ we use a *sphere-packing lower bound* similar to that of [1]. Our t-sparse sets and interleaving schemes are optimal for two infinite but sparse families of pairs (d, t), and $(1 + \frac{t}{d} + \frac{3}{t})$-

approximations otherwise. If $d > rt^2$ we construct a t-sparse set with the index at most 1 above optimum if t is even, and at most $t/2$ above optimum if t is odd.

Given that for 2D arrays the sphere-packing lower bound is tight [1], we investigated whether it remains tight in our setting. We proved that for odd $t \leq \delta$ it is not the case, unless $d \equiv \pm t \pmod{\lceil t^2/2 \rceil}$ (in which case there exists a simple optimal construction). We obtained a partial result for the case of even t.

- For $\delta < t \leq d-2$ the sphere-packing lower bound is sub-optimal; we derive our constructions and lower bounds by analyzing *triples* of consecutive elements of a t-sparse set. For each choice of (d, t) we construct a family of optimal t-sparse sets and an interleaving scheme that is optimal in most cases and is a $(1 + \frac{4}{t})$-approximation otherwise.

Further research. Two natural directions from interleaving schemes on circulant graphs with two offsets would be to interleaving schemes with repetitions and to circulant graphs with more than two offsets [3].

Organization of the paper. Section 2 analyzes the case $t > \delta$. Section 3 is on the case $t \leq \delta$: in Section 3.1 we derive the sphere-packing lower bound, Section 3.2 describes our construction, and Section 3.3 studies the "greedy" approach for constructing t-sparse sets. Finally, in Section 4 we investigate when the sphere-packing lower bound is exact.

Preliminaries. G_d stands for a circulant graph with two offsets $\{1, d\}$. We will talk interchangeably about subgraphs of G_d and subsets of $\mathbb{Z}$. We reserve d, t for the larger offset and the interleaving parameter, respectively. To simplify formulas we let $\delta = \lceil d/2 \rceil$, $\tau = \lceil t/2 \rceil$.

Define the distance $dist(u, v)$ between points u, v as the number of edges in a shortest uv-path in G_d. Let $dist(v) = dist(0, v)$. Define the distance $dist(S, v)$ between a set S and a point v as the minimal distance between v and $u \in S$. For an integer r and a set S define the r-*span* of S as the set of points at distance less than r from S.

By default, all numbers are integers and all sets are subsets of $\mathbb{Z}$. In particular, an interval $[a, b]$ actually denotes the set $[a, b] \cap \mathbb{Z}$. We will use 'iff' as a shorthand for 'if and only if'.

2 Case $t > \delta$: constructions and lower bounds

In this section we assume $t > \delta$. For $t > d - 2$ there is a simple unique optimal interleaving scheme and t-sparse set. For each choice of (d, t) such that $\delta < t \leq d - 2$ we construct a family of optimal t-sparse sets and an interleaving scheme that is optimal in most cases and is a $(1 + \frac{4}{t})$-approximation otherwise.

Call a point w *remote* if $dist(w) \geq t$. Define the *canonical representation* $\text{CAN}(v)$ of $v \geq 0$ as a pair (x, y) such that $v = xd + y$ and $-\delta < y \leq \delta$. Let v_{min} be the smallest positive remote point; let v_{max} be the largest point that is not remote. We will need the following simple facts.

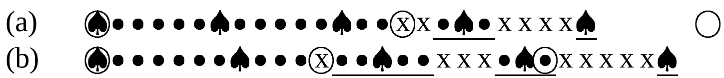
Fact 2.1 *(a)* $dist(v) = x + |y|$ *where* $(x, y) = \text{CAN}(v)$. *(b)* $v_{min} = (\tau - \delta)d + \delta$, $v_{max} = d(t-1)$.

Theorem 2.2 *Suppose $t > d - 2$. Then (a) the unique optimal periodic t-sparse set is $v_{min}\mathbb{Z}$, (b) the unique optimal interleaving scheme is a labeling where each vertex i is labeled $i \pmod{v_{min}}$.*

Proof: (a) The difference between any two elements of $v_{\text{min}}\mathbb{Z}$ is either v_{min} or at least $2v_{\text{min}} > v_{\text{max}}$, so this set is t-sparse (Fig 2a). It is a unique optimal periodic t-sparse set since the interval between any consecutive elements of a t-sparse set is at least v_{min}.

(b) It remains to prove that any other interleaving scheme is not optimal. Indeed, in any other interleaving scheme there is a label-set with two consecutive vertices u,v such that $|u - v| > v_{\min}$. Then the distance between any two points in the interval $[u + 1; u + v_{\min}]$ is less than t, so all points in this interval must be labeled distinctly, not using the label of u and v. This requires at least $v_{\min} + 1$ labels. □

Note that $v_{\min}\mathbb{Z}$ is t-sparse only if $t > d - 2$ since otherwise $dist(2v_{\min}) < t$ (Fig 2b).



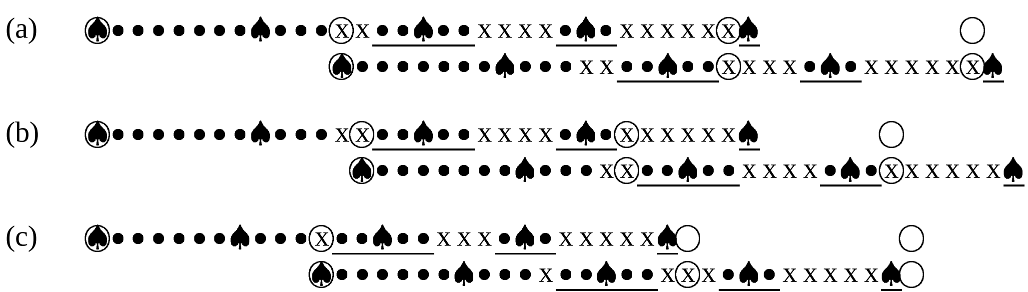
We show the t-span of $\{0\}$ for (a) $t > d-2$, (b) $t \le d-2$. We represent an r-span of $\{p\}$ as a string where consecutive characters correspond to consecutive integers as follows: ♠ is for $p + jd$, $j \in \mathbb{N}$, '•' for other elements of the r-span, and 'x' for other points. Vertices 0, $v_{\min}$, $2v_{\min}$ are encircled.

Figure 2: $v_{\min}\mathbb{Z}$ is t-sparse iff $t > d - 2$.

For the rest of this section assume $\delta < t \le d-2$. Note that $v_{\min} \le v_{\max}/2$ iff $t \le d-2$. We will derive our constructions and lower bounds by analyzing *triples* of consecutive elements of a t-sparse set.

Say (w_1, w_2) is a *remote pair* (with a *sum* $w_1 + w_2$) if w_1, w_2 and $w_1 + w_2$ are positive and remote. Note that for a triple $x < y < z$ of consecutive elements of a t-sparse set, $(z - y, y - x)$ is a remote pair. Say a remote pair *induces* a periodic set $\{\{0, w_1\} + i(w_1 + w_2) : i \in \mathbb{Z}\}$ and an interleaving scheme which is a minimal covering of $\mathbb{Z}$ by copies of this set.

To simplify formulas, define γ as 0 if d is odd and 1 if d is even. Let $v_0 = v_{\min}$, $v_1 = v_0 + \gamma$. Let σ_i be the minimal sum of a remote pair $(v_i, \cdot)$ (Fig 3). Say a remote pair is *standard* if its sum is $\sigma_{\min} = \min(\sigma_0, \sigma_1)$. Now we can state the main theorem of this section.



Compute (a) σ_0 for $(d, t) = (8, 5)$, (b) σ_1 for $(d, t) = (8, 5)$, (c) σ_0 for $(d, t) = (7, 5)$.
In each example, the upper and lower lines are the t-spans of 0 and v_i respectively, in the notation of Fig. 2. We compute σ_i as the leftmost point that is remote in both lines. Points 0, v_i, σ_i and $\sigma_i + v_i$ are encircled.

Figure 3: Computing σ_0 and σ_1.

Theorem 2.3 *(a) The interleaving degree of any interleaving scheme and the index of any t-sparse set are at least $\sigma_{min}/2$, (b) Any standard remote pair induces a t-sparse set. (c) Our interleaving scheme (induced by standard remote pairs) is optimal if $\delta - \tau - \gamma$ is odd, or if both d and t are odd; it is a $(1 + \frac{4}{t})$-approximation otherwise.*

Proof: (a) First we claim that the minimal sum σ of a remote pair is $\sigma_{\min}$. Let (w_1, w_2), $w_1 \le w_2$ be a remote pair with a sum σ such that w_1 is minimal. If $w_1 = v_i$ then $s = \sigma_{\min}$ by definition of $\sigma_{\min}$. Else we can choose z so that $(w_1 - z, w_2 + z)$ is a remote pair, contradicting the minimality of w_1. Indeed, letting $\text{CAN}(w_1) = (x, y)$ we can define z as follows: if $dist(w_1) > t$ let $z = d$, else if $y > 0$ let $z = d + 1$, else let $z = d - 1$. Then $w_1 - z$ and $w_2 + z$ are remote by Fact 2.1. Claim proved.

Let S be a t-sparse set with a well-defined density ρ. Let $\{s_i : i \in \mathbb{Z}\}$ be the an increasing enumeration of S. For each i, $(s_{i+1} - s_i,\ s_{i+2} - s_{i+1})$ is a remote pair, so its sum $s_{i+2} - s_i$ is at least $\sigma_{\min}$. Then $s_n - s_{-n} \geq n\sigma_{\min}$ for any $n > 0$, so $\rho \leq 2/\sigma_{\min}$, which gives the required lower bound on the index of S.

To extend this bound to non-periodic interleaving schemes, let w_1, w_2, w_3 be three consecutive vertices labeled the same in an interleaving scheme. Then $(w_2 - w_1, w_3 - w_2)$ is a remote pair, so its sum $w_3 - w_1$ is at least $\sigma_{\min}$. Therefore, in the interval $[0; \sigma_{\min})$ at most two vertices can be marked by each label, which requires at least $\sigma_{\min}/2$ distinct labels.

(b) Let S be the set induced by a standard remote pair (w_1, w_2). For any $u, v \in S$, either

$$|u - v| \in \{0,\ w_1,\ w_2,\ \sigma_{\min},\ \sigma_{\min} + w_1,\ \sigma_{\min} + w_2\}$$

or else $|u - v| \geq 2\sigma_{\min}$, so $dist(v, u) \geq t$ since it follows from Lemma 2.5 that $2\sigma_{\min} > v_{\max}$.

Therefore, it remains to prove that $\sigma_{\min} + w_1$ and $\sigma_{\min} + w_2$ are remote. By Lemma 2.5 $\text{CAN}(\sigma_i) = (\cdot, y)$ where $|y| \leq t/2$. Now we claim that if $u = xd + y$ and v are positive remote and $|y| \leq t/2$ then $u + v$ is remote. Indeed, w.l.o.g. $v < v_{\max}$, so by Fact 2.4c

$$u + v = v_i + (x + \mu_1)d + (y + \mu_2)$$

where $i \in \{0, 1\}$ and $|\mu_2| \leq \mu_1$. Since $dist(u) = x + |y| \geq t$ it follows that $x \geq t/2 \geq |y|$, so $|y + \mu_2| \leq x + \mu_1$ and by Fact 2.4c $u + v$ is remote. Claim proved.

(c) Consider a remote pair $(v_i, \sigma_i - v_i)$. By Lemma 2.5 and Fact 2.4b for each $j \leq \alpha/2$

$$(v_i + j(d + 1),\ \ \sigma_i - v_i - j(d + 1))$$

is a remote pair. Suppose d is odd or t is even. Then, since by Lemma. 2.5b $\sigma_1 \leq \sigma_0$ and $\sigma_1 - 2v_1 = \alpha_1(d + 1)$, there is a standard remote pair of the form (w, w) if α_1 is even, and $(w, w + d + 1)$ if α_1 is odd. Let S be the set induced by such a pair. If α_1 is even, S obviously extends to an optimal interleaving scheme. Now suppose α_1 is odd. With some arithmetic one can show that $g = \gcd(w, d + 1) = \gcd(t, d + 1)$. By Lemma 2.6 S extends to an interleaving scheme which is optimal if both d and t are odd (since then $\sigma_{\min} = 2w + d + 1$ is even and g is odd), and a $(1 + \frac{4}{t})$-approximation otherwise.

Finally consider the case when d is even and t is odd. Then $\alpha_0 = \alpha_1$, but $\sigma_1 = \sigma_0 + 1$, so we carry out a similar argument for $(v_0, \sigma_0 - v_0)$ and prove that there is a standard remote pair of the form $(w, w + 1)$ if α_0 is even and $(w, w + d + 2)$ if α_0 is odd. By Lemma 2.6 the former case extends to an optimal interleaving scheme, whereas the latter yields a $(1 + \frac{4}{t})$-approximation. □

With some more work we can strengthen Thm. 2.3c as follows: for the case when d is even and t is odd, there exists an interleaving scheme with interleaving degree one above optimal. To complete the proof of Thm. 2.3 we need some technical lemmas. We start with an easy fact.

Fact 2.4 *For $0 < v < v_{max}$ the following are equivalent:*

(a) v is remote.

(b) $\text{CAN}(v - v_{min}) = (\mu_1, \mu_2)$, where $-\mu_1 \leq \mu_2 \leq \mu_1 + \gamma$.

(c) For some $i \in \{0, 1\}$ $v = v_i + \mu_1 d + \mu_2$, where $|\mu_2| \leq \mu_1$.

Lemma 2.5 *(a) $v_{min} + v_i = \alpha d + \beta$ where $\alpha = 2(t - \delta) + 1$ and $\beta \in \{0, 1\}$, specifically $\beta = 1$ if d is odd and $\beta = i$ if d is even. (b) $\sigma_i - v_i = v_{min} + \alpha_i(d + 1) + \gamma$, where $\alpha_i = \delta - \tau - i$ if d and t are even and $\alpha_i = \delta - \tau$ otherwise.*

Proof: Part (a) is an easy computation. For part (b), let $(\alpha_{i\sigma},\ \beta_{i\sigma})$ be the canonical representation of $\sigma - v_{\min} - v_i$, let $(\alpha_i,\ \beta_i)$ be the same for $\sigma = \sigma_i$. Let

$$W_i = \{\sigma > v_i |\ -\alpha_{i\sigma} \le \beta_{i\sigma} \le \alpha_{i\sigma} + \gamma \text{ and } -\delta < \beta + \beta_{i\sigma} \le \delta\}$$

We claim that σ_i is the smallest remote element of W_i. By Fact 2.4b we only need to show that $\beta + \beta_i \le \delta$. Suppose not. Then $\beta = 1$ and $\beta_i = \delta$. Therefore $\sigma = \sigma_i - 1$ is remote since $\sigma \ge v_{\min}$ and $\sigma \equiv \delta \pmod{d}$, and $\sigma - v_i$ is remote by Fact 2.4b. So (v_i, σ) is a remote pair, which contradicts the minimality of σ_i. Claim proved.

Therefore $\alpha_i = \min\{x|\ \varphi(x) \ge t\}$ where $\varphi(x)$ is the maximal value of $dist(\sigma)$ given that $\sigma \in W_i$ and $\alpha_{i\sigma} = x$. For $\sigma \in W_i$ it is the case that $dist(\sigma) = (\alpha + \alpha_{i\sigma}) + |\beta + \beta_{i\sigma}|$, which is maximized, for a fixed $\alpha_{i\sigma}$, only if $\beta_{i\sigma} = \alpha_{i\sigma} + \gamma$. Thus $\varphi(x) = 2x + \alpha + \beta + \gamma$, and part (b) follows easily. □

Finally, the following lemma extends remote pairs to interleaving schemes. We omit the proof since it is (essentially) a special case of Lemma 3.2.

Lemma 2.6 *Let S be the set induced by a remote pair (w_1, w_2). Let $g = \gcd(w_1, w_2)$. Then the smallest number of copies of S required to cover $\mathbb{Z}$ is $g\lceil (w_1 + w_2)/(2g)\rceil$, which is at most g plus the index of S.* □

3 Case $t \le \delta$: constructions and lower bounds

3.1 The sphere-packing lower bound

Recall that the r-span of a set S is the set of points at distance less than r from S. Similar to [1], define a *t-sphere* $S_t = S_t(p)$ centered at a vertex p as the τ-span of $\{p\}$ for odd t and of $\{p, p+d\}$ for even t. Note that for the purposes of this section one could also define $S_{2\tau}(p)$ as the τ-span of $\{p, p+1\}$.

To compute the size of a t-sphere, consider G_d as a two-dimensional $d \times \infty$ mesh with "extra edges" between $(0, n)$ and $(d-1, n+1)$ for all n (Fig. 1b). It is easy to see that for $t \le \delta$ a t-sphere centered at (δ, n) is exactly the same in G_d as in the 2D mesh, since the t-sphere simply does not reach the "extra edges". Therefore by [1] the size of any t-sphere is $\lceil t^2/2 \rceil$. Now we can state the *sphere-packing lower bound*.

Theorem 3.1 *The interleaving degree of any interleaving scheme and the index of any t-sparse set are at least $|S_t|$. Moreover, a t-sparse set whose index is* exactly *$|S_t|$ can be extended to an interleaving scheme with the same interleaving degree.*

Proof: First we claim that if $dist(p,q) \ge t$ then the t-spheres centered at p and q are disjoint. Assume $S_t(p)$ and $S_t(q)$ intersect at w. If t is odd then $dist(p,w) \le \tau - 1$ and $dist(q,w) \le \tau - 1$, so by the triangle inequality $dist(p,q) < t$. Now suppose t is even. Then either $dist(p,w) \le \tau - 1$ or $dist(p+d, w) \le \tau - 1$, same for q. Therefore by the triangle inequality $dist(p,q) < t$ unless $dist(p,w) = dist(q,w) = \tau$. In the latter case, however, $dist(p+d, w) = dist(q+d, w) = \tau - 1$, so there exists a path from $p+d$ to $q+d$ with less than t vertices. Shifting this path by $-d$ produces a pq-path of the same length. Claim proved.

Let S be a t-sparse set of minimal index and density ρ. Since the sets $S_t(p)$, $p \in S$ are pairwise disjoint, the density of their union U is $\rho|S_t| \le 1$, so the index of S is at least $|S_t|$. Now suppose it is exactly $|S_t|$. Then the density of U is 1. Since S is periodic, U is periodic, too, so $U = \mathbb{Z}$. Partition U as follows: let $U_i = \{v_i(p) : p \in S\}$, where $v_i(p)$ is the i-th vertex of $S_t(p)$ from the left. Then the sets U_i are translates of S, hence they are t-sparse. Label all points of U_i with i to get an optimal interleaving scheme.

To extend the lower bound to non-periodic interleaving schemes, note that the distance between any two points in a t-sphere is less than t, so all points of a t-sphere must be labeled differently in an interleaving scheme. □

We used the fact that the distance between any two points in a t-sphere is less than t. It is an open question whether there exist larger sets with this property. Finding such sets would be nice given a $\Omega(t)$ gap between (the general case of) our construction and the sphere-packing lower bound.

3.2 Two-Offset Construction

We will construct t-sparse sets that reach or almost reach the sphere-packing lower bound, and extend them efficiently to interleaving schemes. All things being equal, we prefer t-sparse sets with a simple structure, since they are "nicer", easier to implement and to reason about.

In this section define $q, r \in \mathbb{Z}$ by $d = (q+1)t + r$, $-1 \le r \le t-2$. Let $S_0 = \{0, t, 2t, \ldots, qt\}$. Say a set S with a period p is *two-offset* if $S \cap [0,p) = S_0$. The following lemma extends two-offset sets to interleaving schemes.

Lemma 3.2 *Let S be a two-offset set with a period p. Let $\phi = p/(q+1)$, $g = \gcd(t,p)$. Then the smallest number of copies of S required to cover $\mathbb{Z}$ is $g\lceil \phi/g \rceil$, which is at most g plus the index of S.*
Proof: Let's try to cover the interval $X = [0,p)$. For each integer i let

$$A_i = \{(i + jt) \bmod p : j \in \mathbb{Z}\}.$$

From elementary number theory, the sets $A_0 \ldots A_{g-1}$ form a disjoint partition of X, so the size of each A_i is p/g. Now, each copy of S intersects with exactly one A_i, the size of intersection being $q+1$. Therefore, one needs at least $N = \left\lceil \frac{|A_i|}{q+1} \right\rceil$ copies to cover one of the sets A_i, and at least gN copies to cover all of them. Conversely, to cover $\mathbb{Z}$ by gN copies of S we need the sets $i + j(q+1)t + S$ where $0 \le i < g$ and $0 \le j < N$. Finally, it is easy to see that $gN \le g + \lceil \phi \rceil$, where $\lceil \phi \rceil$ is the index of S.

Define the *two-offset construction* as the two-offset set with a period

$$p_0 = \begin{cases} d\tau - \tau, & \text{if } (t \text{ even and } r = -1) \text{ or } (t \text{ odd and } r = 0,1) \\ d\tau + \tau, & \text{otherwise.} \end{cases}$$

Our results about the two-offset construction are summarized in the following theorem.

Theorem 3.3 *(a) The two-offset construction is t-sparse. (b) Its index achieves the sphere-packing lower bound iff t is even and $d \equiv \pm 1 \pmod t$. If $d > rt^2$ the index is at most 1 above optimum if t is even, and at most τ above optimum if t is odd. Otherwise it is a $(1 + \frac{t}{d} + \frac{1}{t})$-approximation. (c) The interleaving scheme induced by the two-offset construction is optimal iff t is even and $d \equiv \pm 1 \pmod t$. Otherwise it is a $(1 + \frac{t}{d} + \frac{3}{t})$-approximation.*
Proof: Say a set is S_0-remote if all its points are at distance at least t from S_0. The two right-most points of the t-span of S_0 are $p_1 = d(t-1) + qt < dt$ and $p_2 = p_1 - t$. Note that $dt < 2p_0$ unless t is even and $r = -1$, in which case $p_2 < 2p_0 < p_1 < 2p_0 + t$. Therefore the set $jp_0 + S_0$ is S_0-remote for all $j \ge 2$. To prove (a) it remains to show that $p_0 + S_0$ is S_0-remote.

For $0 \le i \le t$ and $0 \le j \le q$ let $v_{ij} = id + jt$ and define the intervals $B_{ij} = v_{ij} + (i - t, t - i)$. Then B_{ij} is the part of the t-span of v_{0j} that lies in $[id - t, id + d + t]$. It is easy to see that the t-span of S_0 is equal to the union of the sets B_{ij} (Fig. 4).

Now let $j < q$. Define the *overlap* between two integer intervals as the size of their intersection if they intersect, and the negated number of points between them if they don't. Then $x_{ij} = t - 2i - 1$ is the overlap between B_{ij} and $B_{i,j+1}$, and $y_i = t - 2i - r - 2$ is the overlap between B_{iq} and $B_{i+1,0}$.

Partition the interval $[id, id + d)$ into intervals $I_{ij} = [v_{ij}, v_{i,j+1})$ and $J_i = [v_{iq}, v_{i+1,0})$. Say an interval is *free* if it contains a point at distance at least t from S_0. Then the intervals I_{ij} are free iff $x_{ij} < 0$ iff $i \ge \tau$, and the intervals J_i are free iff $y_i < 0$ iff $i \geq \lfloor \frac{t-r}{2} \rfloor$. It is easy to verify that all elements of $p_0 + S_0$ lie in the intervals I_{ij} and J_i that are free. Therefore $p_0 + S_0$ is S_0-remote, which completes part (a).

The index of the two-offset construction is, for $p_0 = d\tau \pm \tau$,

$$\left\lceil \frac{p_0}{q+1} \right\rceil = t\tau + \psi = \begin{cases} |S_t| + \tau - 1 + \psi, & t \text{ is odd} \\ |S_t| + \psi, & t \text{ is even} \end{cases} \quad (1)$$

where $\psi = \lceil \tau(r \pm 1)/(q+1) \rceil$. Thus the sphere-packing lower bound is achieved iff t is even $r = \pm 1$. Note that if $d > rt^2$ and $r \neq \pm 1$ then $\psi = 1$. This completes part (b). The approximation guarantee in part (c) follows from Lemma 3.2. □

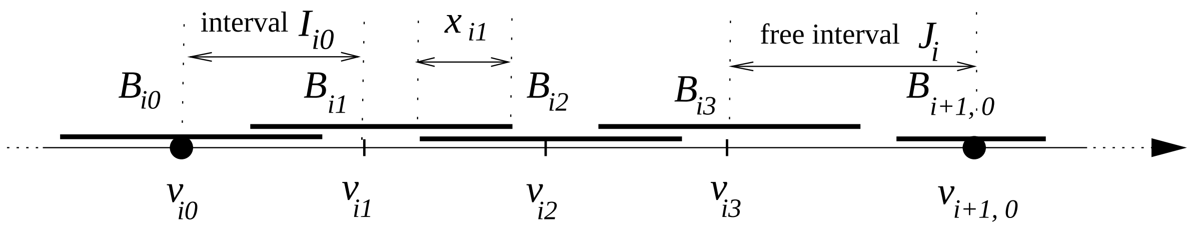


Figure 4: Span of S_0 as the union of the sets B_{ij}

3.3 Greedy approach

A natural way to construct t-sparse sets is the following *greedy* algorithm. Start with an empty set S and $j = 0$. For each consecutive j, insert j into S iff $S \cup \{j\}$ is t-sparse. Since this decision depends only on the *header* $(S - j) \cap [-dt, 0]$ of S, and there are only finitely many possible headers, the construction is periodic starting from some m (i.e. for some p and all $n \geq m$ it is the case that $n \in S$ iff $n + p \in S$). The algorithm stops as soon as the period is detected. The *greedy construction* is the set obtained by replicating the (smallest) period in both directions.

Obviously, the greedy construction is t-sparse. In this construction, each element in is as close as possible to the smaller elements, which makes one hope that it is dense enough. However, it may be the case that if we make some intervals larger, some subsequent intervals can be made shorter, thus increasing the overall density.

Theorem 3.4 *The greedy construction is two-offset iff* $d \equiv 0, \pm 1 \pmod{t}$*, in which case it coincides with the two-offset construction.*

Proof: The greedy algorithm starts out with an empty set, then proceeds to $S = S_0$. Let w be the next number inserted into S. It is easy to see that for $r \leq 1$ we have $w = p_0$, so the 'if' direction follows. Now assume $r \geq 2$. For the converse it suffices to show that $w + t$ is not S_0-remote. Let $\eta = \lfloor \frac{t-r}{2} \rfloor$ and recall the definitions of I_{ij} and J_i from the proof of Thm. 3.3. If $\eta < \tau - 1$ then $w \in I_{\eta q}$, so $w + t$ is not S_0-remote since it lies in the interval $I_{\eta+1,0}$ which is not free. Else it is the case that $\eta = \tau - 1$, $r = 2$ and t is even, so $w + t$ is again not S_0-remote because the only two S_0-remote points in J_η are w and $w + 1$, and the only S_0-remote point in $I_{\tau 1}$ is $w + t + 1$. □

If $d \not\equiv 0, \pm 1 \pmod{t}$, the greedy construction is quite ugly. Computer searches show that the periods are rather long and lack apparent structure. In particular, it is not clear when exactly the periodicity starts.

Note that for $t \geq d-1$ the greedy construction is the unique optimal t-sparse set from Thm. 2.2, and for $\delta < t \leq d-2$ it is the set induced by a remote pair $(v_0, \sigma_0 - v_0)$, which is optimal for odd d and optimal or close to optimal for even d (see Thm. 2.3).

4 More on the sphere-packing lower bound

In this section we assume $t \leq \delta$ and investigate when the sphere-packing lower bound is exact. We solve this question for odd t and give a partial result for even t.

Say a set is *s-optimal* if it is t-sparse and its index reaches the sphere-packing lower bound. Define *the even construction* as the set $|S_t|\mathbb{Z}$ where S_t is a t-sphere defined in Section 3.1. Obviously, among all sets $w\mathbb{Z}, w \in \mathbb{Z}$ only the even construction can be s-optimal. Now we are ready to state the main theorem of this section.

Theorem 4.1 *(a) For odd t, s-optimal constructions exist iff $d \equiv \pm t \pmod{|S_t|}$, in which case the even construction is s-optimal. (b) For even t, the even construction is s-optimal only if $d \equiv \pm 1 \pmod{t}$.*

The 'if'-direction of Thm. 4.1a is easy. It can be derived from one of the constructions in [1], but for the sake of completeness we will prove it here. Let $s = |S_t|$. We claim that if t is odd and $d \equiv t \pmod{s}$ then the even construction is t-sparse. Suppose not, then there are points $p > q$ such that $p \equiv q \pmod{s}$ and $dist(p,q) < t$. Let $p - q = id + j$, $-\delta < j \leq \delta$. Then s divides $id + j$, hence $it + j$. Since $it + j < 2s$, it is equal to s. Therefore, by a simple computation, $t = 2i \pm 1 = \pm(2j - 1)$, so $dist(p,q) = i + |j| = t$, contradiction. Claim proved. For $d \equiv -t \pmod{s}$ the proof is similar.

The rest of the proof of Thm. 4.1 is quite technical; we split it into multiple lemmas. For notational convenience we partition a t-sphere $S_t(p)$ into *stations* (clusters around the points of the form $p + kd$, $k \in \mathbb{Z}$), which can be grouped into the *left branch*, the *right branch*, and (for odd t) the *central station* (Fig. 5a). The notation for picturing t-spheres is summarized in Fig. 5b. In the rest of this section, let S be a (general) s-optimal set; let p be a (general) element of S. We will make a heavy use of the fact that the t-spheres centered in S form a partition of $\mathbb{Z}$.

♠xxxxxx•♠•xxxx••♠••xx•••(♠)•••xx••♠••xxxx•♠•xxxxxx♠
{ Left Branch } { central } { Right Branch }

♠xxxxxx•♠•xxxx••(♠)••xxx••♠••xxxx•♠•xxxxxx♠
{ Left Branch } { Right Branch }

(a) t-spheres for $(d,t) = (8,7)$ (above) and $(d,t) = (8,6)$ (below). Centers are encircled.
(b) We represent a t-sphere $S_t = S_t(p)$ by a string where consecutive characters correspond to consecutive numbers. We use ♠ for the points $p + kd \in S_t$, $k \in \mathbb{Z}$, • for other points of S_t, and 'x' for points not in S_t. Stations are underlined.

Figure 5: *Stations* and *branches* of a t-sphere.

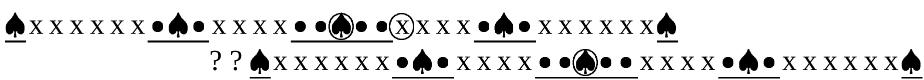
4.1 Case of odd t

Let $L_S = S - (\tau - 1)d$, $R_S = S + (\tau - 1)d$. So $p \in L_S$ (resp. $p \in R_S$) iff p is the leftmost (resp. rightmost) point of some t-sphere centered in S.

Lemma 4.2 *$p+\tau$ is in L_S or R_S.*

Proof: Since the t-spheres centered in S partition $\mathbb{Z}$, $p+\tau \in S_t(q)$ for some $q \in S$, $q \neq p$. Suppose $p+\tau$ lies in the left branch of $S_t(q)$ but is not the leftmost element thereof (see Fig. 6). Then, letting $p_1 = p-d+\tau-1$, $p_2 = p_1+1$, it is easy to see that p_1-1 lies in $S_t(p)$, p_2+1 lies in $S_t(q)$, whereas p_1 and p_2 lie in neither. Thus, p_1 and p_2 are covered by some other t-sphere(s) centered in S. How can that be? In a t-sphere all stations except the leftmost and the rightmost ones have length ≥ 3. Thus, p_1 and p_2 are the leftmost or the rightmost points of some t-spheres centered in S. If p_1 or p_2 is the leftmost point of such a t-sphere S', then S' intersects $S_t(p)$ at $p+\tau-1$, contradiction. So p_1 and p_2 are the rightmost elements of t-spheres $S_t(q_1), S_t(q_2)$ where $q_1, q_2 \in S$. Then $q_1+1 = q_2$, contradiction.

So if $p+\tau$ lies in the left branch of $S_t(q)$, $p+\tau$ must be its leftmost element. Else $p+\tau$ lies in the right branch of $S_t(q)$ or in its central station. Then by a similar proof $p+\tau$ must be the rightmost element of $S_t(q)$. □



- Upper row: $S_t(p)$; p and $p+\tau$ are encircled.
- Lower row: $S_t(q)$; $p-d+\tau-1,\ p-d+\tau$ are labeled by '?'; q is encircled.

Figure 6: For the proof of Lemma 4.2: $S_t(p)$ and $S_t(q)$ for $(d,t) = (8,5)$

It follows that $p+t \notin S$. Indeed, if $p+t \in S$ then, since the t-spheres centered in S are disjoint, $S_t(p+t)$ is the only t-sphere centered in S that contains $p+\tau$. But $p+\tau$ is the inner point of $S_t(p+t)$, contradicting Lemma 4.2. Claim proved. In particular, the two-offset construction from Section 3.2 can not be s-optimal since it starts with $\{0, t, 2t, \dots\}$.

Lemma 4.3 *In both (a) and (b) exactly one of the two statements is true (see Fig. 7):*

a) $p+\tau \in L_S$ and $p-d+\tau-1 \in R_S$
   $p+\tau \in R_S$ and $p+d+\tau-1 \in L_S$

b) $p-\tau \in L_S$ and $p-d-\tau+1 \in R_S$
   $p-\tau \in R_S$ and $p+d-\tau+1 \in L_S$

Proof: By Lemma 4.2, either $p+\tau \in L_S$ or $p+\tau \in R_S$. Suppose $p+\tau \in L_S$ and let $p' = p-d+\tau-1$. Since p' is neither in $S_t(p)$ nor in the t-sphere containing $p+\tau$, it is an element of some other t-sphere $T = S_t(q)$, $q \in S$. Since p' is the leftmost element of some station of T, $p+\tau = p'+d+1 \in T$, unless p' is the rightmost element of T.

Case $p+\tau \in R_S$ is solved similarly. Part (b) follows from part (a) by symmetry. □

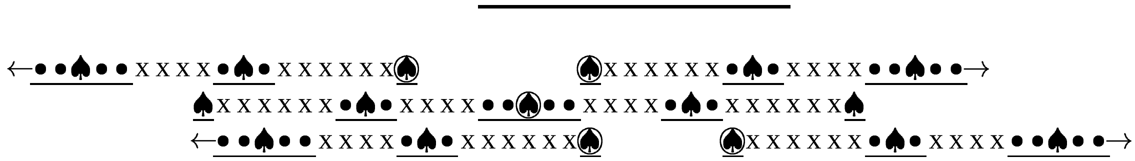
Lemma 4.4 *(a) Exactly one of the following is true (Fig. 8a):*

$$p+\tau,\ p+d-\tau+1 \in L_S \quad \textit{and} \quad p-\tau,\ p-d+\tau-1 \in R_S \tag{2}$$

$$p+\tau,\ p-d-\tau+1 \in R_S \quad \textit{and} \quad p-\tau,\ p+d+\tau-1 \in L_S \tag{3}$$

(b) If (2) then $p+d-t \in S$, if (3) then $p+d+t \in S$ (Fig. 8b).

Proof: (a) By Lemma 4.3, there are four possible cases: (2), (3), $p \pm \tau \in L_S$, and $p \pm \tau \in R_S$. If $p \pm \tau \in L_S$ then by Lemma 4.3 $p_\pm \in S$, where $p_\pm = p - d\tau \pm (\tau-1)$. Thus $p_+ - p_- = t-1$, contradiction. The case $p \pm \tau \in R_S$ is ruled out similarly.

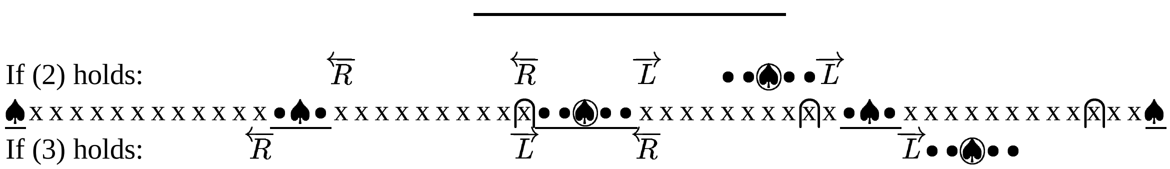


- 1st row: $p+\tau \in L_S$ and $p-d+\tau-1 \in R_S$ are labeled.
- 2nd row: $S_t(p)$ is shown; p is labeled .
- 3rd row: $p+\tau \in R_S$ and $p+d+\tau-1 \in L_S$ are labeled.

Figure 7: The two options in Lemma 4.3, for $(d,t) = (8,5)$

(b) Suppose (2) holds. Let $T = S_t(q)$, $q \in S$, be the t-sphere containing $p' = p + d - \tau$ (Fig. 8c). Say p' is contained in the station W of T. Let W_L, W_R be the stations of T immediately to the left and immediately to the right from W. Suppose W is not the *central* station of T. Then either W_L or W_R is wider than W. Since p' is the rightmost point of W, either $p - \tau \in W_L$ or $p' + d \in W_R$ (Fig. 8c), so at least one of these points lie in T. However, we claim that both points belong to other t-spheres centered in S. Indeed, by Lemma 4.4a $p - \tau \in R_S$. By the same lemma $p' + 1 \in L_S$ is the leftmost point of some t-sphere S' centered in S, so $p' + d \in S'$. Claim proved. Thus, W is the central station of T, so $q = p + d - t$.

If (3), we let $T = S_t(q)$, $q \in S$, be the t-sphere containing $p' = p + d + \tau$. Then by a similar argument $q = p + d + t$. □



(a) Middle row: $S_t(p)$ is shown, p is encircled. Vertices $p \pm \tau$ and $p \pm d \pm (\tau - 1)$ are labeled by $\overrightarrow{L}$ (resp. by $\overleftarrow{R}$) when they are in L_S (resp. in R_S).

(b) Upper row: $q = p + d - t$ is encircled; the central station of $S_t(q)$ is shown. Same for $q = p + d + t$ in the lower row.

(c) Middle row: $p' - d$, p' and $p' + d$ are labeled by $\bigcap$, where $p' = p + d - \tau$.

Figure 8: The two options in Lemma 4.4 for $(d,t) = (14,5)$

Now we can complete the proof of the main theorem.

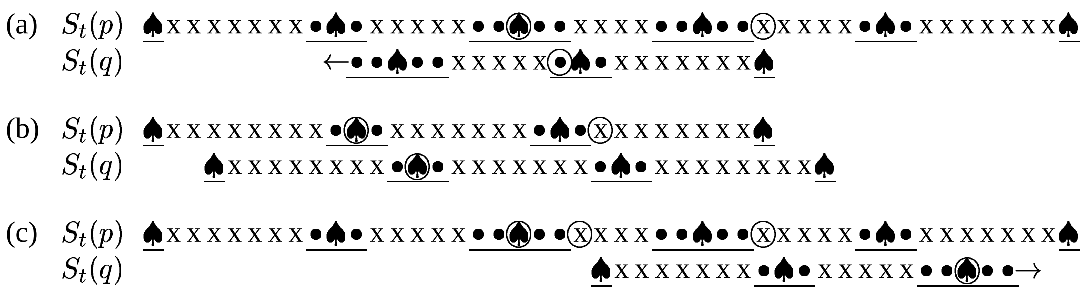
Proof of Thm. 4.1a: Let S be an s-optimal t-sparse set. Take any $p \in S$. If (2) then by Lemma 4.4b $q = p + d - t \in S$. Now we apply Lemma 4.4a to q. Either (2) or (3) must hold for q. Since $q + \tau = p + d - \tau + 1 \in L_S$, (2) does. So we apply Lemma 4.4b again: $q + d - t \in S$. In the same fashion, $p + k(d - t) \in S$ for any $k \in \mathbb{N}$. Since this holds for any $p \in S$, S is periodic with a (not necessarily smallest) period $d - t$. Since S is s-optimal, the density of S is $w/(d - t) = 1/|S_t|$, where w is the number of points of S within one period. Thus, $|S_t|$ divides $d - t$. If (3) holds for p, then by a similar argument $|S_t|$ divides $d + t$. □

4.2 Case of even t

Lemma 4.5 *At least one of* $p + \tau(d + 1)$, $p - \tau(d - 1)$ *is in* S.

Proof: The long t-spheres centered S partition $\mathbb{Z}$. In particular, $p' = p + d + \tau$ is an element of some t-sphere $T = S_t(q)$, $q \in S$. Clearly, p' is a leftmost element of some station of T. Which station? If p' is in the right branch of T then either $p + \tau - 1$ is in both T and $S_t(p)$ (Fig. 9a), or $q = p + t - 1$, which is too close to p (Fig. 9b).

So p' lies in the left branch of T. Now, if p' is the leftmost element of T then $q = p + \tau(d + 1) \in S$, and we are done. Else $p + \tau - 1 \in S_t(p)$, $p + \tau + 1 \in T$, but $p + \tau$ is in neither t-sphere (Fig. 9c). So $p + \tau$ must be either the leftmost or the rightmost element of some other t-sphere $T' = S_t(q')$, $q' \in S$. It cannot be the leftmost element since in this case $p + d + \tau - 1$ is in both T' and $S_t(p)$. Thus, it is the rightmost element of T', in which case $q' = p - \tau(d - 1)$. □



(a) Labeled are: $p + \tau - 1$ in the lower row, p and p' in the upper row. Here $(d, t) = (9, 6)$.
(b) Labeled are: p, p' in the upper row, q in the lower row. Here $(d, t) = (10, 4)$.
(c) Labeled are: p, $p + \tau$, p' in the upper row, q in the lower row. Here $(d, t) = (9, 4)$.

Figure 9: For the proof of Lemma 4.5

Now, if S is s-optimal then by Lemma 4.5 $t^2/2$ divides either $(d + 1)\tau$ or $(d - 1)\tau$, which proves Thm. 4.1b. Note that by Thm. 4.1b and Thm. 3.3 whenever the even construction is s-optimal there exists an s-optimal two-offset construction. However, since the even construction is simpler, it is still interesting to investigate when exactly it is s-optimal.

Let D_t be the set of all values of d such that the even construction is s-optimal. We computed $\min(D_t)$ and the first 20-30 elements of D_t for each $t \leq 42$. This data motivated several conjectures:

- Let p be the smallest prime that does not divide $t/2$. Then $\min(D_t) = pt - 1$.
- Let $d \equiv 1 \pmod{t}$. Then $d \in D_t$ iff $d - 2 \in D_t$.
- Consider the sequence of intervals between consecutive elements of D_t. This sequence is periodic, starting from the very first element of the sequence. Let $p_0, p_1, \ldots, p_n$ be the distinct prime divisors of $t/2$. Then the length of the period is $2 \times \prod_{j=0}^{n}(p_j - 1)$, and the sum of the elements in a period is $t \times \prod_{j=0}^{n} p_j$.

References

[1] M. Blaum, J. Bruck and A. Vardy, "Interleaving schemes for multidimensional cluster errors," *IEEE Trans. Inform. Theory*, vol.44, pp. 730-743, 1998.

[2] M. Blaum, J. Bruck and P.G. Farell, "Two-dimensional interleaving schemes with repetitions," *IBM Research Report RJ 10047*, Oct. 1996.

[3] B. Elspas and J. Turner, "Graphs with circulant adjacency matrices," *J. Combin. Theory*, No.9, pp. 297-307, 1970.

[4] T. Etzion and A. Vardy, "Two-dimensional interleaving schemes with repetitions: constructions and bounds," *IEEE Trans. Inform. Theory*, vol.48, pp. 428-458, 2002.

[5] "MaxiCode Briefing Document," UPS Inc., *http://www.maxicode.com*.

[6] M.A.Neifeld and M.McDonald, "Error correction for increasing the usable capacity of photorefractive memories," *Opt. Lett.*, vol. 19, pp.1483, 1994.