Almost Optimal Explicit Selectors

Bogdan S. Chlebus [1,⋆] and Dariusz R. Kowalski [2, 3,⋆⋆]

[1] Department of Computer Science and Eng., UCDHSC, Denver, CO 80217, USA
[2] Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK
[3] Instytut Informatyki, Uniwersytet Warszawski, Banacha 2, Warszawa, Poland

Abstract. We understand *selection by intersection* as distinguishing a single element of a set by the uniqueness of its occurrence in some other set. More precisely, given two sets A and B, if $A \cap B = \{z\}$, then element $z \in A$ is *selected* by set B. Selectors are such families $\mathcal{S}$ of sets B of some domain that allow to select many elements from sufficiently small subsets A of the domain. Selectors are used in communication protocols for the multiple-access channel, in implementations of distributed-computing primitives in radio networks, and in algorithms for group testing. We give new explicit (n, k, r)-selectors of size $\mathcal{O}(\min\left[\, n, \frac{k^2}{k-r+1} \text{ polylog } n \,\right])$, for any parameters $r \leq k \leq n$. We establish a lower bound $\Omega(\min\left[\, n, \frac{k^2}{k-r+1} \cdot \frac{\log(n/k)}{\log(k/(k-r+1))} \,\right])$ on the length of (n, k, r)-selectors, which demonstrates that our construction is within a polylog n factor close to optimal. The new selectors are applied to develop explicit implementations of selection resolution on the multiple-access channel, gossiping in radio networks and an algorithm for group testing with inhibitors.

1 Introduction

Selection by intersection means distinguishing a single element of a set as the only element of some other set. More precisely, given a subset $A \subseteq X$ of a finite domain X, element $z \in A$ is *selected* by a set $B \subseteq X$ when $A \cap B = \{z\}$.

The power of such a selection is often considered in quantitative terms, which translate into efficiency in applications. A natural parameter to consider is the size of sets A from which we select. A family $\mathcal{S}$ of subsets of X is called *k-selective*, following Chlebus *et al.* [4], if we can select an element from *any* subset $A \subseteq X$ of size $|A| \leq k$ by a set in $\mathcal{S}$. Families $\mathcal{S}$ that are useful in application, because of their selection-related properties, are typically parametrized by the size n of the domain X, and we want the number k to be close to n, while keeping the size of $\mathcal{S}$ small. Additionally, we may want to have many elements $z \in A$ to be selected, for any $A \subseteq X$ of size k.

⋆ The work of this author is supported by the NSF Grant 0310503.
⋆⋆ The work of this author supported in part by the KBN Grant 4T11C04425.

M. Liśkiewicz and R. Reischuk (Eds.): FCT 2005, LNCS 3623, pp. 270–280, 2005.

This is captured by the following definition. Let n, k and r be positive integers so that $r \leq k \leq n$. Let $\mathcal{S}$ be a family of subsets of $[n] = [1..n]$. We say that $\mathcal{S}$ is an (n, k, r)-*selector* if, for each set $A \subseteq [n]$ of size $|A| = k$, there are at least r elements in A that can be selected from A by sets in $\mathcal{S}$.

The name "selectors" was coined by De Bonis, Gąsieniec and Vaccaro [13] in the context of their work on group testing. Their definition of selectors is in terms of binary matrices and corresponds to certain generalized superimposed codes. The notion of a selector generalizes many popular combinatorial structures. Among them there are (n, k)-*selective families*, introduced by Chlebus *et al.* [4], which are $(n, k, 1)$-selectors in selector terminology. Objects called simply (n, k)-*selectors* by Chrobak, Gąsieniec and Rytter [7] correspond to $(n, 2k, 3k/2)$-selectors. Finally, (n, k)-*strongly-selective families* introduced by Clementi, Monti and Silvestri [11] are nothing but (n, k, k)-selectors. Such (n, k, k)-selectors are closely related to $(k-1)$-*cover-free families*, in the hypergraph terminology [23], and to superimposed codes [17,24]. See Section 2 for an overview of the related combinatorics and matrix representations.

Selection by intersection is a notion that occurs in many disguises in combinatorial settings and in algorithmic and communication applications. Selectors can be applied in deterministic conflict resolution in multiple-access channels [3,25], in broadcasting and gossiping algorithms for ad-hoc radio networks [1,4,7,12,26] and in deterministic algorithms for group testing [13,14,15,16]. Related combinatorial structures called radio synchronizers are directly applicable in algorithms for waking up radio networks [5,6,20] and to implement distributed-computing primitives in radio networks [5,6], like leader election and synchronization of local clocks.

Combinatorial structures used in implementations of algorithms as part of their code are said to be *explicit* when there are algorithms that produce them in time that is polynomial in the size of the output.

Our Results. The contributions are summarized as follows.

I. We construct explicit (n, k, r)-selectors of size $\mathcal{O}(\min[\, n, \frac{k^2}{k-r+1} \text{ polylog } n \,])$, for any configuration of parameters $r \leq k \leq n$. The design involves explicit dispersers. This result extends the ranges of two previously known explicit constructions. One is that of explicit superimposed codes of n codewords of length $\mathcal{O}(\min[\, n, k^2 \log^2 n \,])$ that are k-disjunct. This is the classical design by Kautz and Singleton [24]. The codes can be interpreted as (n, k, k)-selectors of size given by the length of codewords. The other is that of explicit $(n, k, 3k/4)$-selectors of size $\mathcal{O}(k \text{ polylog } n)$ given by Indyk [22].

II. We show that the length of an (n, k, r)-selector has to be $\Omega(\min[\, n, \frac{k^2}{k-r+1} \cdot \frac{\log(n/k)}{\log(k/(k-r+1))} \,])$. This demonstrates that the above mentioned explicit construction is within a polylog n factor close to optimal.

III. The new selectors are applied to obtain the following specific applications: (i) an explicit oblivious solution to a variant of a static selection problem for the multiple-access channel, (ii) an explicit implementation of gossiping in radio networks, and (iii) an algorithm for group testing with inhibitors.

Previous Work. Selectors generalize many kinds of families of finite sets, and the work on special cases of selectors has been motivated by either purely combinatorial interests, as in the case of cover-free families, or by applications of combinatorics, as in group testing and in communication in the multiple-access channel and ad-hoc radio networks. We summarize briefly the known facts about upper and lower bounds on the size of selectors, and on explicitness of known selectors.
Existence of small selectors: Komlós and Greenberg [25] showed that there are $(n,k,1)$ selectors of size $\mathcal{O}(k\log(n/k))$. Dyachkov and Rykov [18] showed that there exist (n,k,k)-selectors of size $\mathcal{O}(k^2\log n)$; see [19,23] for a simple proof and also [16] for a detailed account of existential upper bounds for superimposed codes. De Bonis, Gąsieniec and Vaccaro [13] showed that there exist (n,k,r)-selectors of size $\mathcal{O}\big(\frac{k^2}{k-r+1}\log(n/k)\big)$.
Lower bounds on size of selectors: Clementi, Monti and Silvestri [11] showed that $(n,k,1)$-selectors have to be of size $\Omega(k\log(n/k))$. Lower bounds on (n,k,r)-selectors with r close to k are stronger. In particular, (n,k,k)-selectors obey a lower bound $\Omega(\min[\,n,k^2\log n/\log k\,])$ on their size. The first component n in this bound follows from the observation that a family of all n singletons is an (n,k,k)-selector for any $k\leq n$. This lower bound was first showed by Dyachkov and Rykov [17] in a slightly weaker form $\Omega(c_k\cdot n)$, where $c_k=\Theta(k^2/\log k)$. It was rediscovered by Chaudhuri and Radhakrishnan [2] in a stronger form $\frac{k^2\ln n}{100\ln k}$ for $k\leq n^{1/3}$, which was later improved by Clementi, Monti and Silvestri [11] who showed that the constant 100 can be replaced by 16, for $k\leq\sqrt{2n}$. See also [16] for a detailed account of lower bounds for superimposed codes. De Bonis, Gąsieniec and Vaccaro [13] gave a general lower bound $\Omega\big(\min\big[n,\frac{(r-1)^2}{k-r+1}\cdot\frac{\log(n/(k-r+1))}{\log((r-1)/(k-r+1))}\big]\big)$ on the size of (n,k,r)-selectors.
Explicit constructions: Explicit (n,k,k)-selectors of size $\mathcal{O}(\min[\,n,k^2\log^2 n\,])$ were given by Kautz and Singleton [24]. Indyk [22] was the first to observe a relation between selectors and dispersers. He gave explicit $(n,k,3k/4)$-selectors of size $\mathcal{O}(\min[\,n,k\ \mathrm{polylog}\ n\,])$. Clementi *et. al* [10] explicitly constructed $(n,k,1)$-selectors of size $\mathcal{O}(\min[\,n,k\log k\log(n/k)\,])$.

Explicit graphs with good expansion properties, on which we rely in our constructions, were given by Ta-Shma, Umans and Zuckerman [28].

Structure of This Document. Section 2 discusses interrelations between selection, in the sense of obtaining singleton sets as intersections, and superimposed coding. Section 3 describes the construction of explicit selectors with a matching lower bound. Section 4 discusses applications in the areas of multiple-access channel, radio networks and group testing. We conclude with a discussion in Section 5.

2 Selection and Superimposed Coding

Given a finite domain of size ℓ, or simply $[1..\ell]$, a subset A can be uniquely represented by its binary characteristic vector of length ℓ: an occurrence of 1 in position i means that number i belongs to A. This allows to represent families of

subsets of a finite domain as binary two-dimensional arrays. This may be defined in two ways, depending on the role of rows and columns. A representation is called *primal* when rows represent elements of the domain and columns represent subsets. A representation is called *dual* when columns represent elements of the domain and rows represent subsets. In the literature on selection in families of sets and on superimposed codes, the primal representation is typically used.

2.1 Selection by Intersection

Selection of elements of a finite set can be defined in terms of binary matrices as follows: for a subset A of the domain, represented as a set of rows, row $z \in A$ is *selected* by a column if there is exactly one occurrence of 1 in this column among the rows in A and this occurrence is at row z.

Let $\mathcal{B}$ be an $n \times m$ binary array. It represents, in a primal way, m subsets of set $[1..n]$. Array $\mathcal{B}$ is an $(n, k, 1)$-selector of size m if for any set $A \subseteq [1..n]$ of rows of $\mathcal{B}$, where $|A| = k$, there is a column with exactly one occurrence of 1 among the rows in A. In general, array $\mathcal{B}$ is an (n, k, r)-selector of size m if for any set $A \subseteq [1..n]$ of rows of $\mathcal{B}$, where $|A| = k$, there are at least r columns with exactly one occurrence of 1 among the rows in A.

A dynamic adversarial component, in binary arrays representing families of subsets of a finite domain, is added by a possibility to have rows shifted. By this we mean that the distance of a shift is at most the original number of columns and the obtained array has new entries filled with zeroes. We say that an array $\mathcal{B}$ has good *synchronization properties* when, for any set A of rows of a sufficiently small size, some column selects a single row among the rows in A after these rows have been shifted by arbitrary distances. When we want to be able to select against such adversaries from all sets A of size $|A| = k$, then $\mathcal{B}$ could be called k-*synchronizing*, following [5,6,20].

This synchronization terminology is motivated by the application in the multiple-access channel with collision detection we describe next. It was first considered by Gąsieniec, Pelc and Peleg [20]; see [3] for a detailed exposition of this model of communication. The model has the following properties. A single transmission by an attached station is heard by all stations. More than one simultaneous transmissions interfere with one another, and none can be heard by the stations, but the stations receive a feedback notifying them of the interference. Suppose there are n stations, some k of which wake up spontaneously and immediately start attempts to broadcast a message to all. The first successful transmission wakes up the whole network and allows to synchronize local clocks. A schedule of transmissions, for a station, is specified as a binary sequence. An occurrence of 1 as the i-th bit represents a transmission in the i-th step according to the local clock.

We say that a binary $n \times m$ array $\mathcal{B}$ is a (n, k)-*synchronizer of length* m if for any nonempty set $A \subseteq [1..n]$ of rows of size at most k, and for any shifts of rows in A, there is a column that selects exactly one (shifted) row in A. Such synchronizers were defined by Chrobak, Gąsieniec and Kowalski [6] in the context of their work on the problems of wake-up, leader election and synchronization

of local clocks in multi-hop radio networks. This notion was also implicitly used by Gąsieniec, Pelc and Peleg [20] in their work on waking up a multiple-access channel. The fastest known algorithm to wake up a multi-hop radio networks, given by Chlebus and Kowalski [5], uses *universal synchronizers*, which are arrays with properties stronger than those of radio synchronizers.

A construction of a (n, n)-synchronizers of length $\mathcal{O}(n^{1+\varepsilon})$, for any constant $\varepsilon > 0$, was given by Indyk [22]; it can be performed in a quasi-polynomial time $\mathcal{O}(2^{\text{polylog } n})$. Chlebus and Kowalski [5] described explicit (n, k)-synchronizers of a length $\mathcal{O}(k^2 \text{ polylog } n)$.

Radio synchronizers have the properties of selective families. It follows that (n, k)-synchronizers have to be of lengths $\Omega(k \log(n/k))$. Using the probabilistic method, Gąsieniec, Pelc and Peleg [20] showed that there are (n, n)-synchronizers of a length $\mathcal{O}(n \log^2 n)$, and Chrobak, Gąsieniec and Kowalski [6] showed that there are (n, k)-synchronizers of a length $\mathcal{O}(k^2 \log n)$.

2.2 Superimposed Coding

Superimposed codes are typically represented as binary arrays, with columns used as binary codewords. Take an $a \times b$ binary array with the property that no boolean sum of columns in any set D of $d = |D|$ columns can cover a column not in D. This is a superimposed code of b binary codewords of length a each that is *d-disjunct*. When columns are representing sets, then d-disjunctness means that no union of up to d sets in any family of sets D could cover a set outside D.

A book by Du and Hwang [16] provides a contemporary exposition of superimposed coding and its relevance to nonadaptive group testing. There is a natural correspondence between such codes and strongly selective families, which we give for completeness sake. Using this correspondence, the explicit superimposed codes given by Kautz and Singleton [24] can be interpreted as (n, k, k)-selectors of size $\mathcal{O}(k^2 \log^2 n)$.

The correspondence is obtained by using the representations, primal and dual, of families of sets as boolean arrays. Take a (n, k)-strongly-selective family $\mathcal{S}$, that is, an (n, k, k)-selector, of some length m. This means there are m sets in $\mathcal{S}$, and the domain is of size n. A dual boolean representation of $\mathcal{S}$ is an $m \times n$ binary array $\mathcal{A}$. Let us interpret this array in the primal way. This representation yields a superimposed code: it consists of n codewords of length m each. Observe that this code is $(k-1)$-disjunct. To show this, suppose, to the contrary, that some $(k-1)$ columns of a set C of columns can cover column x of $\mathcal{A}$. Then the columns in $C \cup \{x\}$ represent a subset of $[1..n]$ of size k. By the property of $\mathcal{S}$ being a strongly selective family, there is a row in $\mathcal{A}$ with an occurrence of 1 in column x and only occurrences of 0 in columns in C. This means that column x is not covered by the columns in C, which is a contradiction. A reasoning in the opposite direction is similar.

Generalizations of superimposed codes can be proposed, which correspond to (n, k, r)-selectors being a generalization of (n, k)-strongly-selective families. This was already done by Dyachkov and Rykov [17]. De Bonis and Vaccaro [14,15] considered such generalizations in the context of their work on group testing.

Similar, but more restricted, generalized superimposed codes were considered by Chu, Colbourn and Syrotiuk [8,9] in their work on distributed communication in ad-hoc multi-hop radio networks.

3 Explicit Selectors

We show how to construct (n,k,r)-selectors of size $\mathcal{O}(\min\lceil n, \frac{k^2}{k-r+1}\ \text{polylog}\ n\rceil)$, for *any* configuration of parameters $r \le k \le n$, in time polynomial in n. The construction is by combining strongly selective families with dispersers. Strongly selective families are (n,k,k)-selectors. We show how to use dispersers to decrease the third parameter r in (n,k,r)-selectors while also gracefully decreasing the size of the family of sets.

If $r \le 3k/4$ then we can use the construction of an $(n,k,3k/4)$-selector given by Indyk [22]. Assume that $r > 3k/4$. Let $0 < \varepsilon < 1/2$ be a constant.

A bipartite graph $H = (V,W,E)$, with set V of inputs and set W of outputs and set E of edges, is a (ℓ,d,ε)-*disperser* if it has the following two properties:

Dispersion: for each $A \subseteq V$ such that $|A| \ge \ell$, the set of neighbors of A is of size at least $(1-\varepsilon)|W|$.
Regularity: H is d-left-regular.

Let graph $G = (V,W,E)$, where $|V| = n$, $|W| = \Theta((k-r+1)d/\delta)$, be a $(k-r+1,d,\varepsilon)$-disperser, for some numbers d and δ. (The amount $\log\delta$ is called the *entropy loss* of this disperser.) An explicit construction of such graphs, that is, in time polynomial in n, was given by Ta-Shma, Umans and Zuckerman [28], for any $n \ge k \ge r$, and some $\delta = \mathcal{O}(\log^3 n)$, where $d = \mathcal{O}(\text{polylog}\ n)$ is a bound on the left-degrees.

Let $\mathcal{M} = \{M_1,\ldots,M_m\}$ be an explicit $(n, c\delta\frac{k}{k-r+1})$-strongly-selective family, for a sufficiently large constant $c > 0$ that will be fixed later, of size $m = \mathcal{O}(\min\lceil n, \delta^2(\frac{k}{k-r+1})^2\log^2 n\rceil)$, as constructed by Kautz and Singleton [24].

We define an (n,k,r)-selector $\mathcal{S}(n,k,r)$ of size $\min[\,n, m|W|\,]$, which consists of sets $F(i)$, for $1 \le i \le \min[\,n, m|W|\,]$. There are two cases to consider, depending on the relation between n and $m|W|$. The case of $n \le m|W|$ is simple: take the singleton containing only the i-th element of V as $F(i)$. Consider a more interesting case when $n > m|W|$. For $i = am + b \le m|W|$, where a and b are non-negative integers satisfying $a + b > 0$, let $F(i)$ contain all the nodes $v \in V$ such that v is a neighbor of the a-th node in W and $v \in M_b$.

Theorem 1. *The family $\mathcal{S}(n,k,r)$ is an (n,k,r)-selector of size*

$$\mathcal{O}(\min\lceil n, \frac{k^2}{k-r+1}\ \text{polylog}\ n\rceil)\ .$$

Proof. First we show that $\mathcal{S}(n,k,r)$ is an (n,k,r)-selector. The case $n \le m|W|$ is clear, since each node in a set A of size k occurs as a singleton in some set $F(i)$. Consider the case $n > m|W|$. Let set $A \subseteq V$ be of size k. Suppose, to the contrary, that there is a set $C \subseteq A$ of size $k-r+1$ so that none among the

elements in C is selected by sets from $\mathcal{S}(n,k,r)$, that is, $F(i) \cap A \neq \{v\}$, for each $v \in C$ and $1 \leq i \leq m|W|$.

Claim: Every $w \in N_G(C)$ has more than $c\delta\frac{k}{k-r+1}$ neighbors in A.

The proof is by contradiction. Assume, for simplicity of notation, that $w \in W$ is the w-th element of set W. Suppose, to the contrary, that there is $w \in N_G(C)$ which has at most $c\delta\frac{k}{k-r+1}$ neighbors in A, that is, $|N_G(w) \cap A| \leq c\delta\frac{k}{k-r+1}$. By the fact that $\mathcal{M}$ is a $(n, c\delta\frac{k}{k-r+1})$-strongly-selective family we have that, for every $v \in N_G(w) \cap A$, the equalities

$$F(w \cdot m + b) \cap A = (M_b \cap N_G(w)) \cap A = M_b \cap (N_G(w) \cap A) = \{v\}$$

hold, for some $1 \leq b \leq m$. This holds in particular for every $v \in C \cap N_G(w) \cap A$. There is at least one such $v \in C \cap N_G(w) \cap A$ because set $C \cap N_G(w) \cap A$ is nonempty since $w \in N_G(C)$ and $C \subseteq A$. The existence of such v is in contradiction with the choice of C. Namely, C contains only elements which are not selected by sets from $\mathcal{S}(n,k,r)$ but $v \in C \cap N_G(w) \cap A$ is selected by some set $F(w \cdot m + b)$. This makes the proof of Claim complete.

Recall that $|C| = k - r + 1$. By dispersion, the set $N_G(C)$ is of size larger than $(1-\varepsilon)|W|$, hence, by the Claim above, the total number of edges between the nodes in A and $N_G(C)$ in graph G is larger than

$$(1-\varepsilon)|W| \cdot c\delta\frac{k}{k-r+1} = (1-\varepsilon)\Theta((k-r+1)d/\delta) \cdot c\delta\frac{k}{k-r+1} > kd\ ,$$

for a sufficiently large constant c. This is a contradiction, since the total number of edges incident to nodes in A is at most $|A|d = kd$. It follows that $\mathcal{S}(n,k,r)$ is an (n,k,r)-selector.

The size of this selector is

$$\begin{aligned}
\min[\, n, m|W|\,] &= \mathcal{O}(\min\left[\, n, \delta^2(\frac{k}{k-r+1})^2 \log^2 n \cdot (k-r+1)d/\delta\,\right]) \\
&= \mathcal{O}(\min\left[\, n, d\delta\frac{k^2}{k-r+1}\log^2 n\,\right]) \\
&= \mathcal{O}(\min\left[\, n, \frac{k^2}{k-r+1}\ \text{polylog}\ n\,\right])\ ,
\end{aligned}$$

since $d = \mathcal{O}(\ \text{polylog}\ n)$ and $\delta = \mathcal{O}(\log^3 n)$.

Indyk [22] gave an explicit construction of $(n,k,3k/4+1)$-selectors of size $\mathcal{O}(k\ \text{polylog}\ n)$. His method does not appear to be directly adaptable to produce (n,k,r)-selectors in the case when $k-r$ is significantly smaller than k.

Theorem 2. *The length of an (n,k,r)-selector has to be*

$$\Omega(\min\left[\, n, \frac{k^2}{k-r+1} \cdot \frac{\log(n/k)}{\log(k/(k-r+1))}\,\right])\ .$$

Proof. We show that this fact follows from the lower bounds given in [11] and [13]. We may assume that $\frac{k^2}{k-r+1} \cdot \frac{\log(n/k)}{\log(k/(k-r+1))} = o(n)$, because otherwise it is sufficient to take a family of n singletons to obtain a selector of size n.

A bound on the size of $(n, k, 1)$-selectors given in [11] is $\Omega(k \log(n/k))$; we call it CMS.

A bound on the size of (n, k, r)-selectors given in [13] is $\Omega(\min\lceil n, \frac{(r-1)^2}{k-r+1} \cdot \frac{\log(n/(k-r+1))}{\log((r-1)/(k-r+1))}\rceil)$; we call it DGV.

Suppose the parameters k and r are functions of n. If $k = \mathcal{O}(1)$, then the size of an (n, k, r)-selector is $\Omega(\log n)$, which is consistent with the three bounds mentioned. Suppose $k = \omega(1)$. We consider two cases.

Case $1 \le r \le k/2$:

Apply the CMS bound. Observe that $\frac{k^2}{k-r+1} = \Theta(k)$ since $k/(k-r+1)$ is $\Theta(1)$.

Case $k/2 < r \le k$:

Apply the DGV bound. Observe that $(r-1) = \Theta(k)$ and $\log(n/(k-r+1)) = \Omega(\log(n/k))$.

This completes the proof.

4 Applications

Theorem 2 demonstrates that the construction of Theorem 1 is close to optimal within a polylog n factor. It follows that any algorithmic application of selectors can be made explicitly instantiated with only an additional poly-logarithmic overhead factor in performance. We describe three such applications.

4.1 Multiple Access Channel

There are n stations attached to a multiple-access channel. A transmission performed by exactly one station is heard by every station, while more simultaneous transmissions interfere with one another, which prevents hearing any of the transmitted messages. The channel is said to be *with collision detection* if each station receives a feedback notifying about an interference of many messages sent simultaneously. We consider the weaker channel *without* collision detection.

The problem of *k-selection* is defined as follows. Suppose each among some k of the stations stores its own input value, and the goal is to make at least one such a value heard on the channel. This problem can be solved deterministically in time $\mathcal{O}(\log n)$ applying the binary-search paradigm. It requires expected time $\Omega(\log n)$, as was shown by Kushilevitz and Mansour [27]. This selection problem can be generalized to (k, r)-selection as follows: we want to hear at least r values from among k held by the stations.

Corollary 1. *The (k, r)-selection problem for n stations, where $r \le k \le n$, can be solved deterministically by an explicitly instantiated oblivious algorithm in the multiple-access channel without collision detection in time*

$$\mathcal{O}(\min\lceil n, \frac{k^2}{k-r+1} \operatorname{polylog} n\rceil) .$$

Proof. An (n, k, r)-selector $\mathcal{S}$ can be used to provide an oblivious deterministic solution to the selection problem as follows. The sets in $\mathcal{S}$ are ordered, and station i performs a transmission if i is in the i-th set of $\mathcal{S}$. The performance bound follows from Theorem 1.

4.2 Gossiping in Radio Networks

The fastest known distributed algorithm for gossiping in directed ad-hoc multi-hop radio networks, given by Gąsieniec, Radzik and Xin [21], employs general (n, k, r)-selectors. The bound $\mathcal{O}(n^{4/3} \log^4 n)$ on time obtained in [21] relies on existence of (n, k, r)-selectors of size $\mathcal{O}\big(\frac{k^2}{k-r+1} \log(n/k)\big)$ shown in [13].

Corollary 2. *Gossiping in directed ad-hoc radio networks of n nodes can be performed in time $\mathcal{O}(n^{4/3}$ polylog $n)$ by an explicitly instantiated distributed algorithm.*

Proof. Use our explicit selectors in the algorithm of [21], instead of those known to exist only, to make the algorithm explicit. The performance bound follows from the estimates in [21] and Theorem 1. The additional overhead is of order polylog n.

4.3 Group Testing with Inhibitors

There is a set of n objects, some k of which are categorized as *positive*. The task of group testing it to determine all positive elements by asking queries of the following form: does the given subset of objects contain at least one positive element? The efficiency is measured by the number of queries.

The c-stage group testing consists of partitioning all objects c times into disjoint pools and testing the pools separately in parallel in each among c stages. Groups testing with inhibitors allows a category of some r objects, called *inhibitors*, so that a presence of such an element in a query hides the presence of a positive item. De Bonis, Gąsieniec and Vaccaro [13] showed how to implement 4-stage group testing with inhibitors relying on (n, k, r)-selectors.

Corollary 3. *There is an explicit implementation of a 4-stage group testing on a set of n objects with k positive items and r inhibitors, which consist of $\mathcal{O}(\min\big[\, n, \frac{k^2}{k-r+1}$ polylog $n\,\big])$ queries, if only $k < n - 2r$.*

Proof. Instantiate the scheme of tests developed in [13] with our explicit selectors. The bound on the number of tests follows from the estimates given in [13] and from Theorem 1. The additional overhead for explicitness is of order polylog n.

5 Conclusion

We showed how to construct (n, k, r)-selectors in time that is polynomial in n, for any configuration $r \leq k \leq n$ of parameters. The obtained selectors are close

to optimal, in terms of size, within a polylog n factor. Our construction is by way of combining explicit dispersers with explicit superimposed codes to obtain a family of a prescribed size with the desired degree of selectiveness.

This construction has a number of applications, as exemplified in Section 4. Such applications are fairly direct in the case of selection in multiple-access channel. A general scheme of application works by using any algorithm relying on selectors and making it explicit by plugging in the explicit selectors given in Theorem 1. We presented this for gossiping in radio networks and group testing with inhibitors. Since our construction is within a polylog-n factor from optimal, the additional overhead factor in efficiency is always of order polylog n.

Synchronizers are closely related to selectors. They are more robust, in that they exhibit selection-related properties even if rows of arrays representing them are shifted arbitrarily. The best know explicit (n, k)-synchronizers of a length $\mathcal{O}(k^2 \text{ polylog } n)$ were given in [5]. It is an open problem if explicit synchronizers of length $\mathcal{O}(k \text{ polylog } n)$ can be developed.

Known explicit constructions of dispersers, of a quality we need in construction of almost optimal selectors, are fairly complex. Simpler explicit dispersers applicable to obtain close to optimal selectors would be interesting to construct.

Exploring further a connection between selectors and graphs with expansion properties is an interesting topic of research.

References

1. R. Bar-Yehuda, O. Goldreich, and A. Itai, On the time complexity of broadcast in radio networks: an exponential gap between determinism and randomization, *Journal of Computer and System Sciences*, 45 (1992) 104 - 126.
2. S. Chaudhuri, and J. Radhakrishnan, Deterministic restrictions in circuit complexity, in *Proc., 28th ACM Symposium on Theory of Computing (STOC)*, 1996, pp. 30 - 36.
3. B.S. Chlebus, Randomized communication in radio networks, in *"Handbook of Randomized Computing,"* P.M. Pardalos, S. Rajasekaran, J. Reif, and J.D.P. Rolim (Eds.), Kluwer Academic Publishers, 2001, Vol. I, pp. 401 - 456.
4. B.S. Chlebus, L. Gąsieniec, A. Gibbons, A. Pelc, and W. Rytter, Deterministic broadcasting in unknown radio networks, *Distributed Computing*, 15 (2002) 27 - 38.
5. B.S. Chlebus, and D.R. Kowalski, A better wake-up in radio networks, in *Proc., 23rd ACM Symposium on Principles of Distributed Computing (PODC)*, 2004, pp. 266 - 274.
6. M. Chrobak, L. Gąsieniec, and D.R. Kowalski, The wake-up problem in multihop radio networks, in *Proc., 15th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2004, pp. 985 - 993.
7. M. Chrobak, L. Gąsieniec, and W. Rytter, Fast broadcasting and gossiping in radio networks, *Journal of Algorithms*, 43 (2002) 177 - 189.
8. W. Chu, C.J. Colbourn, and V.R. Syrotiuk, Slot synchronized topology-transparent scheduling for sensor networks, *Computer Communications*, to appear.
9. W. Chu, C.J. Colbourn, and V.R. Syrotiuk, Topology transparent scheduling, synchronization, and maximum delay, in *Proc., 18th International Parallel and Distributed Processing Symposium (IPDPS)*, 2004, pp. 223 - 228.

10. A.E.F. Clementi, P. Crescenzi, A. Monti, P. Penna, and R. Silvestri, On computing ad-hoc selective families, in *Proc., 5th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM-APPROX)*, 2001, LNCS 2129, pp. 211 - 222.
11. A.E.F. Clementi, A. Monti, and R. Silvestri, Distributed broadcast in radio networks of unknown topology, *Theoretical Computer Science*, 302 (2003) 337 - 364.
12. A. Czumaj, and W. Rytter, Broadcasting algorithms in radio networks with unknown topology, in *Proc., 44th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2003, pp. 492 - 501.
13. A. De Bonis, L. Gąsieniec, and U. Vaccaro, Generalized framework for selectors with applications in optimal group testing, in *Proc., 30th International Colloquium on Automata, Languages and Programming (ICALP)*, 2003, LNCS 2719, pp. 81 - 96.
14. A. De Bonis, and U. Vaccaro, Constructions of generalized superimposed codes with applications to group testing and conflict resolution in multiple access channels, *Theoretical Computer Science*, 306 (2003) 223 - 243.
15. A. De Bonis, and U. Vaccaro, Improved algorithms for group testing with inhibitors, *Information Processing Letters*, 67 (1998) 57 - 64.
16. D.Z. Du, and F.K. Hwang, *"Combinatorial Group Testing and its Applications,"* World Scientific, 2000.
17. A.G. Dyachkov, and V.V. Rykov, A survey of superimposed code theory, *Problems of Control and Information Theory,* 12 (1983) 229 - 244.
18. A.G. Dyachkov, and V.V. Rykov, Bounds on the length of disjunctive codes, *Problemy Peredachi Informatsii,* 18 (1982) 7 - 13.
19. Z. Füredi, On r-cover free families, *Journal of Combinatorial Theory (A),* 73 (1996) 172 - 173.
20. L. Gąsieniec, A. Pelc, and D. Peleg, The wakeup problem in synchronous broadcast systems, *SIAM Journal on Discrete Mathematics*, 14 (2001) 207 - 222.
21. L. Gąsieniec, T. Radzik, and Q. Xin, Faster deterministic gossiping in directed ad-hoc radio networks, in *Proc., 9th Scandinavian Workshop on Algorithm Theory (SWAT)*, 2004, LNCS 3111, pp. 397 - 407.
22. P. Indyk, Explicit constructions of selectors and related combinatorial structures, with applications, in *Proc., 13th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2002, pp. 697 - 704.
23. S. Jukna, *"Extremal Combinatorics,"* Springer-Verlag, 2001.
24. W.H. Kautz, and R.R.C. Singleton, Nonrandom binary superimposed codes, *IEEE Transactions on Information Theory*, 10 (1964) 363 - 377.
25. J. Komlós, and A.G. Greenberg, An asymptotically nonadaptive algorithm for conflict resolution in multiple-access channels, *IEEE Transactions on Information Theory*, 31 (1985) 303 - 306.
26. D.R. Kowalski, and A. Pelc, Time of deterministic broadcasting in radio networks with local knowledge, *SIAM Journal on Computing*, 33 (2004) 870 - 891.
27. E. Kushilevitz and Y. Mansour, An $\Omega(D \log(N/D))$ lower bound for broadcast in radio networks, *SIAM Journal on Computing,* 27 (1998) 702 - 712.
28. A. Ta-Shma, C. Umans, and D. Zuckerman, Loss-less condensers, unbalanced expanders, and extractors, in *Proc., 33rd ACM Symposium on Theory of Computing (STOC)*, 2001, pp. 143 - 152.