SUPERCONCENTRATORS, GENERALIZERS AND GENERALIZED CONNECTORS WITH LIMITED DEPTH

(Preliminary Version)

Danny Dolev
Hebrew University
Jerusalem, Israel

Cynthia Dwork
Cornell University
Ithaca, NY 14853

Nicholas Pippenger
IBM Research Laboratory
San Jose, CA 95193

Avi Wigderson
Princeton University
Princeton, NJ 08544

ABSTRACT: We show that the minimum possible size of an n-superconcentrator with depth $2k\geq 4$ is $\Theta(n\lambda(k, n))$, where $\lambda(k, .)$ is the inverse of a certain function at the k-th level of the primitive recursive hierarchy. It follows that the minimum possible depth of an n-superconcentrator with linear size is $\Theta(\beta(n))$, where β is the inverse of a function growing more rapidly than any primitive recursive function. Similar results hold for generalizers.

We give a simple explicit construction for a $(d_1 \ldots d_k)$-generalizer with depth k and size $(d_1+\ldots+d_k)d_1\ldots d_k$. This is applied to give a simple explicit construction for a generalized n-connector with depth 2k-3 and size $(2d_1+3d_2+\ldots+3d_{k-1}+2d_k)$ $d_1\ldots d_k$. These are the best explicit constructions currently available. We also show that, for each fixed $k\geq 2$, the minimum possible size of a generalized n-connector with depth k is $\Omega(n^{1+1/k})$ and $O((n \log n)^{1+1/k})$.

1. Introduction

The objects of our study in this paper are interconnection networks of various types. An <u>n-network</u> is an acyclic directed graph with n distinguished vertices called <u>inputs</u> and n other distinguished vertices called <u>outputs</u>. We shall be concerned with the minimum possible <u>size</u> (number of edges) and <u>depth</u> (number of edges in the longest path from an input to an output) that n-networks with various connectivity properties can possess. The properties that we shall be most interested in are the following. An <u>n-superconcentrator</u> is an n-network such that, for any subset of the inputs and any equinumerous subset of the outputs, there exist vertex-disjoint paths joining the chosen inputs to the chosen outputs. An <u>n-generalizer</u> is an n-network such that, for any assignment of non-negative integers summing to at most n to the inputs, there exist vertex-disjoint trees joining each input to the assigned number of distict outputs. An <u>n-connector</u> is an n-network such that, for any one-to-one correspondence between certain inputs and distinct outputs, there exist vertex-disjoint paths joining each chosen input to the corresponding output. Finally, a <u>generalized n-connector</u> is an n-network such that, for any one-to-many correspondence between certain inputs and disjoint sets of outputs, there exist vertex-disjoint trees joining each chosen input to the corresponding set of outputs.

An <u>n-crossbar</u> is an n-network with depth 1 and size n^2, with an edge joining each input to each output. For all of the problems considered here, a crossbar provides a solution with small depth and large size. Our goal is to find alternate solutions with larger but limited depth and smaller size.

These networks, which are the subject of an extensive literature, are relevant to theoretical computer science in several ways. Firstly, their use has been proposed as components of parallel computers (see Ofman [15] and Galil and Paul [7]);

upper bounds for networks thus give upper bounds for parallel computations. Secondly, they have been used in the construction of graphs that are hard to pebble (see Paul, Tarjan and Celoni [16], Lengauer and Tarjan [9] and Pippenger [23]); here, upper bounds for networks give lower bounds for pebbling. Finally, oblivious computations for many naturally occuring functions give rise to such networks (see Valiant [28] and Tompa [27]); here, lower bounds for networks give lower bounds for computations. Our interest in networks with limited depth is appropriate to the study of circuits with limited depth but unbounded fan-in (see, for example, Furst, Saxe and Sipser [5]). In particular, our lower bounds for weak superconcentrators (see Section 2 below) can be used to show the optimality of recent upper bounds by Chandra, Fortune and Lipton [2, 3] for the size and depth of circuits for binary addition.

2. Superconcentrators with Limited Depth

Let $d(n)$ denote the minimum possible size of an n-superconcentrator and let $d_k(n)$ denote the minimum possible size of an n-superconcentrator with depth at most k.

Valiant [28] showed that $d(n)=O(n)$. He showed, in fact that $d_k(n)=O(n)$ for $k=O(n^{\alpha})$, for some $0<\alpha<1$; Pippenger [18] improved this to $k=O(\log n)$. The sharpest upper bound to size currently available, due to Bassalygo [1], is

$$d_k(n) \le 36n + O(\log n)$$

with $k=O(\log n)$. The sharpest lower bound currently available, due to Lev and Valiant (see Lev [10]), is

$$d(n) \ge 5n + O(\log n).$$

For depth 1, we have $d_1(n)=n^2$, trivially. For depth 2, Pippenger [22] showed that

$$d_2(n) = O(n(\log n)^2)$$

and

$$d_2(n) = \Omega(n \log n).$$

To describe our results on superconcentrators of greater depth, we shall need to define some very rapidly and very slowly growing functions. Following Tarjan [25], let us define

$A(0, j) = 2j$ for $j\ge1$,

$A(i, 1) = 2$ for $i\ge1$, and

$A(i, j) = A(i-1, A(i, j-1))$ for $i\ge1$ and $j\ge2$.

For $i\ge0$, define

$$\lambda(i, x) = \min\{j: A(i, j)\ge x\}.$$

Then $\lambda(0, x)=\lceil x/2\rceil$ and

$$\lambda(i, x) = \min\{j: \lambda^{(j)}(i-1, x)\le1\}$$

for $i\ge1$, where $\lambda^{(1)}(i, x)=\lambda(i, x)$ and $\lambda^{(j)}(i, x)=\lambda(i, \lambda^{(j-1)}(i, x))$ for $j\ge2$. The function $\lambda(0, x)$ behaves roughly like $x/2$, $\lambda(1, x)$ like $\log_2 x$, $\lambda(2, x)$ like $\log^* x$ and so forth. Define

$$\beta(x) = \min\{i: A(i, i)\ge x\}.$$

Then

$$\beta(x) = \min\{i: \lambda(i, x)\le i\}.$$

The function β grows more slowly than the inverse of any primitive recursive function.

Theorem 1: For $k\ge2$,

$$d_{2k}(n) = O(n\lambda(k, n)).$$

Sketch of Proof: An ($\underline{a}$, $\underline{b}$)-partial n-superconcentrator is an n-network in which, for any $b+1\le m\le a$, any set of m inputs and any set of m outputs, there exist m-b vertex-disjoint paths joining chosen inputs to chosen outputs. Let $d_k(n, a, b)$ denote the minimum possible size of an (a, b)-partial n-superconcentrator with depth at most k. The inequalities

$d_k(n) \le d_k(n, n, 0)$,

$d_1(n, n, \lfloor n/2\rfloor) \le n$,

$d_k(a, a, c) \le d_k(n, a, b) + d_k(n, b, c)$ and

$d_2(n, a, 0) \le 2na$ are immediate.

The basic building blocks for our construction of partial superconcentrators will be concentrators. An ($\underline{n}$, $\underline{m}$, ℓ)-concentrator is a network with n inputs and m outputs such that, for every set of at most ℓ inputs, there exist vertex-disjoint paths joining the chosen inputs to outputs. Let $c_k(n, m, \ell)$ denote the minimum possible size of an (n, m, ℓ)-concentrator with depth at most k.

Lemma 1.1: If $n\ge2\ell$ and $m\ge\ell+1$, then

$$c_1(n, m, \ell) \le 9n(\log(n/\ell))/(\log(m/\ell)).$$

The proof is a standard probabilistic argument, as in Pinsker [17]. This is the only non-constructive argument in this section. All further networks will be built from concentrators by explicit constructions. This lemma is, to within a constant factor, the best possible, as has been shown by Nakamura and Masson [13].

To obtain superconcentrators of depth 2, we shall use the inequality

$$d_2(n, \ell, \lfloor m/2\rfloor) \le 2c_1(n, m, \ell).$$

Using this inequality twice, we obtain the first part of the following proposition.

Proposition 1.1: If $r \geq 2$, then

$$d_2(n, \lfloor n/r \rfloor, \lfloor n/2r \rfloor) \leq 135n\lambda(1, r),$$

$$d_2(n, n, \lfloor n/r \rfloor) \leq 135n(\lambda(1, r))^2$$

and

$$d_2(n) \leq 135n(\lambda(1, n))^2.$$

The remaining two parts are easy consequences of the first. This proposition gives an alternate proof of the upper bound in Pippenger [22].

To obtain superconcentrators of depth 4 and greater, we shall use a special case of Lemma 1.1.

Lemma 1.2: If $n \geq 2\ell$, then

$$c_1(n, \lfloor (n\ell)^{1/2} \rfloor, \ell) \leq 27n.$$

We shall also use the following key inequality, valid for $a \leq m \leq n$:

$$d_{2k}(n, a, b) \leq 2c_1(n, m, a) + d_{2k-2}(m, a, b).$$

An important consequence of this inequality together with Lemma 1.2 and Proposition 1.1 is the following.

Lemma 1.3: If $r \geq 2$, then

$$d_4(n, \lfloor n/r \rfloor, \lfloor n/4r^2 \rfloor) \leq 2400n.$$

Using this lemma twice together with the key inequality, Lemma 1.2 and Proposition 1.1, we obtain the first part of the following proposition.

Proposition 1.2: If $r \geq 2$, then

$$d_4(n, \lfloor n/r \rfloor, \lfloor n/2^r \rfloor) \leq 5000n,$$

$$d_4(n, n, \lfloor n/r \rfloor) \leq 5000n\lambda(2, r)$$

and

$$d_4(n) \leq 5000n\lambda(2, n).$$

Finally, using Lemma 1.3, the key inequality, Lemma 1.2 and Proposition 1.2, we obtain the first two parts of the following proposition by simultaneous induction on k.

Proposition 1.3: If $r \geq 2$, then

$$d_{2k}(n, \lfloor n/r \rfloor, \lfloor n/A(k-1, r) \rfloor) \leq 5000n,$$

$$d_{2k}(n, n, \lfloor n/r \rfloor) \leq 5000n\lambda(k, r)$$

and

$$d_{2k}(n) \leq 5000n\lambda(k, n).$$

This completes the proof of Theorem 1. □

Combining this result with a result of Bassalygo [1], we obtain linear-sized superconcentrators with the best available constant in the size bound, but much smaller depth.

Corollary 1.1: For some $k=2\beta(n)+O(\log \beta(n))$,

$$d_k(n) \leq 36n + O(n/\beta(n)).$$

Proof: Bassalygo [1] shows that

$$d_{2k}(n) \leq d_{2k-2}(\lceil 5n/9 \rceil) + 16n + 10.$$

Using this recurrence $\lceil 21\log_{9/5} \beta(n) \rceil$ times, then applying Theorem 1 with $k=2\beta(n)$, yields the corollary. □

Let us say that a network is synchronous if all paths from an input to an output have the same length. In a synchronous network with depth k, the vertices are partitioned into (k+1) ranks (the inputs form rank 0 and the outputs rank k), and the outputs are partitioned into k stages (the edges directed out of inputs form stage 1 and the edges directed into outputs stage k). Synchronous networks have obvious relevance in situations where "pipelining" is desired. Let $d^*(n)$ denote the minimum possible size of a synchronous n-superconcentrator.

Corollary 1.2:

$$d^*(n) = O(n\beta(n)).$$

This is an immediate consequence of Corollary 1.1 and the following easy lemma.

Lemma 1.4:

$$d^*(n) \leq kd_k(n).$$

We shall give lower bounds for networks having a weaker property than that of being a superconcentrator. Consider n-networks in which the inputs and ouputs are regarded as elements from the set $\{1, \ldots, n\}$. We shall say that a set $A=\{a_1, \ldots, a_m\}$of inputs and a set $B=\{b_1, \ldots, b_m\}$ of outputs are interleaved if $a_1<b_1<\ldots<a_m<b_m$. A weak n-superconcentrator is an n-network such that, for any subset of the inputs and any equinumerous and interleaved set of outputs, there exist vertex-disjoint paths joining the chosen inputs to the chosen outputs.

Let $w_k^*(n)$ denote the minimum possible size of a synchronous weak n-superconcentrator with depth exactly k.

Theorem 2: For $k \geq 1$,

$$w_{2k}^*(n) = \Omega(n\lambda(k, n)).$$

Sketch of Proof: We shall begin with a technical lemma.

Lemma 2.1: For every $k \geq 1$ and every $x \geq 1$, there exist a set H and functions $c_1, \ldots, c_k$ such that the following conditions are satisfied.

(1) For all $h \in H$, $c_1(h) \geq 1, \ldots, c_k(h) \geq 1$.

(2) For all $h \in H$, $c_1(h) \ldots c_k(h) \leq x$.

(3) For all $y \geq 1$,

$$\sum_{h \in H, c_k(h) < y} c_k \leq 2y.$$

(4) For all $y \geq 1$,

$$\sum_{h \in H, c_k(h) \geq y} 1/c_k(h) \leq 2/y.$$

(5) For all $1 \leq j \leq k-1$ and $y \geq 1$,

$\Sigma_{h\varepsilon H, c_j(h)<y\leq c_j(h)\ldots c_k(h)}\ 1/c_j(h)\ldots c_{k-1}(h) \leq 1.$

(6) We have

$$\Sigma_{h\varepsilon H}\ 1/c_1(h)\ldots c_{k-1}(h) \geq \lambda(k, x)/6-1.$$

Sketch of Proof: For each $\ell\geq 1$, we shall construct a labeled ordered rooted tree $T_k(\ell)$. The tree $T_k(\ell)$ will have height k (every path from the root to a leaf will have length k) and its edges will be labeled with integral powers of 2. The out-degree of the root will be ℓ. The out-degree of every other vertex that is not a leaf will be equal to the label of the unique edge directed into that vertex. At each level of the tree (edges directed out of the root are at level k and edges directed into leaves are at level 1), the label of the first edge to be created will be 1. To determine the label of the j-th edfe to be created at a given level, consider the (j-1)-st edge created at that level and consider all paths starting with that edge and continuing to a leaf. The label of the j-th edge will be 2 times the maximum over all such paths of the product of the labels of the edges on that path. This completes the specification of $T_k(\ell)$.

It can be shown that the maximum over all paths from the root to a leaf of the product of the labels of the edges on that path is as most $A(k, 6\ell)$. Choose $\ell=\lfloor(\lambda(k, x)-1)/6\rfloor\geq\lambda(k, x)/6-1$, let H be the set of leaves in $T_k(\ell)$ and let $c_1(h), \ldots, c_k(h)$ be the successive labels of the edges on the path from the root to h. The conditions of the lemma follow easily from the construction of $T_k(\ell)$. □

Let G be a synchronous weak n-superconcentrator with depth 2k and size $w_{2k}^*(n)$. If v is a vertex in G, let f_v and g_v denote the in-degree and out-degree, respectively, of v. For $0\leq j\leq 2k$, let V_j denote the set of vertices in rank j. Then

$$w_{2k}^*(n) = \Sigma_{1\leq j\leq k}\ \Sigma_{v\varepsilon V_j}\ f_v + \Sigma_{k\leq j\leq 2k-1}\ \Sigma_{v\varepsilon V_j}\ g_v.$$

It remains to estimate these sums from below.

Set x=n/2 and let H and $c_1, \ldots, c_k$ be as in Lemma 2.1. For each $h\varepsilon H$, set $p(h)=1/c_1(h)\ldots c_k(h)$. Let A be a random subset of the inputs obtained by taking each of the inputs {1, ..., n} independently with probability p(h). Let B be a random subset of the outputs obtained by taking each of the outputs {1, ..., n} independently with probability p(h). From A and B, we can obtain interleaved subsets A' and B' by a "greedy algorithm" and show that the expected size of A' and B' is at least $(np(h)-1)/2$. It follows that there is a set Π of vertex-disjoint paths joining A and B with $E(|\Pi|)\geq(np(h)-1)/2$.

With each path $\pi=(v_0, \ldots, v_{2k})$ in Π we shall associate a vertex $w(\pi)$ in the following way. If $f_{v_1}>c_1(h)$, then $w(\pi)=v_1$. If $f_{v_1}\leq c_1(h)$ but $g_{v_{2k-1}}>c_1(h)$, then $w(\pi)=v_{2k-1}$. For $2\leq j\leq k-1$, if $f_{v_1}\leq c_1(h), \ldots, f_{v_{j-1}}\leq c_{j-1}(h)$ and $g_{v_{2k-1}}\leq c_1(h), \ldots, g_{v_{2k-j+1}}\leq c_{j-1}(h)$ but $f_{v_j}>c_j(h)$, then $w(\pi)=v_j$. For $2\leq j\leq k-1$, if $f_{v_1}\leq c_1(h), \ldots, f_{v_j}\leq c_j(h)$ and $g_{v_{2k-1}}\leq c_1(h), \ldots, g_{v_{2k-j+1}}\leq c_{j-1}(h)$ but $g_{v_{2k-j}}>c_j(h)$, then $w(\pi)=v_{2k-j}$. Finally, if $f_{v_1}\leq c_1(h), \ldots, f_{v_{k-1}}\leq c_{k-1}(h)$ and $g_{v_{2k-1}}\leq c_1(h), \ldots, g_{v_{k+1}}\leq c_{k-1}(h)$, then $w(\pi)=v_k$.

Let W be the set of vertices $w(\pi)$ associated with paths π in Π. Since π passes through $w(\pi)$, and since the paths in Π are vertex-disjoint, the associated vertices in W are distinct. Thus $|W|=|\Pi|$ and $E(|W|)\geq(np(h)-1)/2$.

If v_j is a vertex in V_j for some $1\leq j\leq k$, let X_{v_j} denote the set of inputs v_0 that are joined to v_j by a path $(v_0, \ldots, v_j)$ for which $f_{v_1}\leq c_1(h), \ldots, f_{v_{j-1}}\leq c_{j-1}(h)$. If $\xi=|X_{v_j}|$, then $\xi\leq f_{v_j} c_1(h)\ldots c_{j-1}(h)$. If v_j is a vertex in V_{2k-j} for some $1\leq j\leq k$, let Y_{v_j} denote the set of outputs v_{2k} that are joined to v_j by a path $(v_j, \ldots, v_{2k})$ for which $g_{v_{2k-1}}\leq c_1(h), \ldots, g_{v_{2k-j+1}}\leq c_{j-1}(h)$. If $\eta=|Y_{v_j}|$, then $\eta\leq g_{v_j} c_1(h)\ldots c_{j-1}(h)$. For $1\leq j\leq k-1$, if v_j appears in W, then X_{v_j} meets A. The probability of this event is $1-(1-p(h))^{\xi}\leq\xi p(h)$ (using Bernoulli's inequality). This probability is, of course, also at most 1. Thus the expected number of vertices in $V_1, \ldots, V_{k-1}$ that appear in W is at most

$$\Sigma_{1\leq j\leq k-1}\ \Sigma_{v\varepsilon V_j, f_v>c_j(h)}\ \min\{f_v c_1(h)\ldots c_{j-1}(h)p(h), 1\}.$$

Similarly, the expected number of vertices in $V_{k+1}, \ldots, V_{2k-1}$ that appear in W is at most

$$\Sigma_{1\leq j\leq k-1}\ \Sigma_{v\varepsilon V_{2k-j}, g_v>c_j(h)}\ \min\{g_v c_1(h)\ldots c_{j-1}(h)p(h), 1\}.$$

Finally, a vertex v_k in V_k can appear in W only if X_{v_k} meets A and Y_{v_k} meets B. Since A and B are independent random variables, the probability of this event is $(1-(1-p(h))^{\xi})(1-(1-p(h))^{\eta})\leq\xi\eta p(h)^2$

$\le((\xi+\eta)p(h)/2)^2$ (using the inequality between geometric and arithmetic means). This probability is again also at most 1. Thus the expected number of vertices in V_k that appear in W is at most

$$\Sigma_{v\varepsilon V_k} \min \{((f_v+g_v)c_1(h)\ldots c_{k-1}(h)p(h)/2)^2, 1\}.$$

Combining these upper bounds for the expected number of vertices in W with the lower bound derived previously, we obtain

$$\begin{aligned}&\Sigma_{1\le j\le k-1} \Sigma_{v\varepsilon V_j, f_v>c_j(h)} \\ &\quad \min \{f_v c_1(h)\ldots c_{j-1}(h)p(h), 1\} \\ &\quad + \Sigma_{1\le j\le k-1} \Sigma_{v\varepsilon V_{2k-j}, g_v>c_j(h)} \\ &\qquad \min \{g_v c_1(h)\ldots c_{j-1}(h)p(h), 1\} \\ &\quad + \Sigma_{v\varepsilon V_k} \\ &\qquad \min \{((f_v+g_v)c_1(h)\ldots c_{k-1}(h)p(h)/2)^2, 1\} \\ &\quad \ge (np(h)-1)/2.\end{aligned}$$

We shall now multiply the preceding inequality by $c_k(h)$ and sum over $h\varepsilon H$. After interchanging the order of summation and using $p(h)=1/c_1(h)\ldots c_k(h)$, we obtain the key inequality

$$\begin{aligned}&\Sigma_{1\le j\le k-1} \Sigma_{v\varepsilon V_j} \Sigma_{h\varepsilon H, c_j(h)<f_v} \\ &\quad \min \{f_v/c_j(h)\ldots c_{k-1}(h), c_k(h)\} \\ &\quad + \Sigma_{1\le j\le k-1} \Sigma_{v\varepsilon V_{2k-j}} \Sigma_{h\varepsilon H, c_j(h)<g_v} \\ &\qquad \min \{g_v/c_j(h)\ldots c_{k-1}(h), c_k(h)\} \\ &\quad + \Sigma_{v\varepsilon V_k} \Sigma_{h\varepsilon H} \\ &\qquad \min \{((f_v+g_v)/2)^2/c_k(h), c_k(h)\} \\ &\quad \ge \Sigma_{h\varepsilon H} c_k(h)(np(h)-1)/2.\end{aligned}$$

Using Lemma 2.1 to estimate the sums in this inequality, we can show that

$$w_{2k}^*(n) \ge n\lambda(k, (n/2))/144,$$

which completes the proof of Theorem 2. □

Let $w^*(n)$ denote the minimum possible size of a synchronous weak n-superconcentrator. Clearly, $w^*(n)\le d^*(n)$.

Corollary 2.1:

$$w^*(n) = \Omega(n\beta(n)).$$

The proof is immediate from Theorem 2.

Let $w_k(n)$ denote the minimum possible size of a weak n-superconcentrator with depth at most k. Clearly, $w_k(n)\le d_k(n)$.

Corollary 2.2: If k is a function of n such that

$$w_k(n) = O(n),$$

then $k=\Omega(\beta(n))$.

The proof is immediate from Corolary 2.1 and the analogue for weak superconcentrators of Lemma 1.4.

Let e(n) denote the minimum possible size of an n-generalizer and let $e_k(n)$ denote the minimum possible size of an n-generalizer with depth at most k.

Pippenger [19] showed that $e_k(n)=O(n)$ for some $k=O((\log)^2)$. The sharpest upper bound to size currently available, due to Bassalygo [1], is

$$e_k(n) \le 79.9n + O(n^\alpha)$$

with $k=O(n^\alpha)$ for some $0<\alpha<1$. For depth 1, we again have $e_1(n)=n^2$, trivially. For greater depths, we prove the following bounds.

Theorem 3:

$$e_2(n) = O(n(\log n)^3).$$

Theorem 4: For $k\ge2$,

$$e_{2k}(n) = O(n(\lambda(k, n))^2).$$

Let $e_k^*(n)$ denote the minimum possible size of a synchronous n-generalizer with depth exactly k.

Theorem 5: For $k\ge1$,

$$e_{2k}^*(n) = \Omega(n\lambda(k, n)).$$

The proofs of these three theorems are similar to those of the analogous results for superconcentrators and will be omitted in this preliminary version.

3. A Simple Explicit Construction for Generalizers

Extending work by Ofman [15] and Garmash and Shor [8], Thompson [26] showed by a simple explict construction that for $d_1\ge2, \ldots, d_k\ge2$,

$$e_{2k-1}(d_1\ldots d_k) \le (d_1+2d_2+\ldots+2d_k)d_1\ldots d_k.$$

(We have not defined the notion of a "simple explicit construction" precisely. A somewhat arbitrary but technically useful definition is that a Turing machine running in space O(log n) can write out a description of an n-network.) Chung and Wong [4] showed by a quite different explicit construction that

$$e_k(2^{k-1}m) \le (2(k-1)+m)2^{k-1}m.$$

We shall present a result that improves the construction of Thompson [26] and extends the construction of Chung and Wong [4].

Theorem 6: By a simple explicit construction, for $d_1\ge2, \ldots, d_k\ge2$,

$$e_k(d_1\ldots d_k) \le (d_1+\ldots+d_k)d_1\ldots d_k.$$

The heart of the proof will be a proposition concerning the coloring of balls distributed among boxes. Consider dm boxes 0, ..., dm-1. Let them be cyclically assigned the d colors 0, ..., d-1, so that box i is assigned color i (modulo d). An (d, m)-distribution is an assignment of dm balls to these dm boxes (the balls are indistinguishable, so all that matters is the number of balls assigned to each box). A coloring of a (d, m)-distribution is an assignment one of the d colors 0, ..., d-1 to each of the dm balls (since the balls are indistinguishable, all that matters is the number of balls of each color in each box). A coloring is consistent if, among any d cyclically consecutive boxes i-d+1, ..., i (modulo dm) at most one contains a ball of color i (modulo d). A coloring is balanced if each of the d colors is assigned to m balls.

Proposition 6.1: For every (d, m)-distribution, there is a consistent and balanced coloring.

For the proof, we shall need the notion of a partial coloring.

For $0 \le k \le d(m-1)$, a k-partial coloring of a (d, m)-distribution is an assignment of the colors to some of the balls such that only boxes 0, ..., k-1 contain uncolored balls. (A 0-partial coloring is simply a coloring.) A partial coloring is consistent if, among any d cyclically consecutive boxes i-d+1, ..., i (modulo dm), if any contains an uncolored ball, then none contains a ball of color i (modulo d) and, in any case, at most one contains a ball of color i (modulo d). (A 0-partial coloring is consistent if and only if, regarded as a coloring, it is consistent. Thus our definitions of "consistent" are consistent.)

Let $a=(a_0, \ldots, a_{d-1})$ be a sequence of d integers satisfying $a_0 \ge 0, \ldots, a_{d-1} \ge 0$ and $a_0+\ldots+a_{d-1}=dm$. For each $0 \le i \le d-1$, consider the smallest integer $1 \le h \le d$ such that $a_{i-h+1}+\ldots+a_i \le hm$ (subscripts modulo d). Define $\Delta(a, i)$ to be $hm-(a_{i-h+1}+\ldots+a_i)$ (subscripts modulo d) and define R(a, i) to be the set {i-h+1, ..., i} (modulo d).

Let X be a partial coloring of a (d, m)-distribution. For $0 \le i \le d-1$, let $A_i(X)$ denote the number of uncolored balls in boxes of color i and let $B_i(X)$ denote the number of balls of color i. Let $a_i(X)=A_i(X)+B_i(X)$ and let $b_i(X)=A_i(X)+B_{i-1}(X)$ (subscripts modulo d). Then $a_0(X)+\ldots+a_{d-1}(X)=dm$ and $b_0(X)+\ldots+b_{d-1}(X)=dm$. We shall say that X is promising if, for all $0 \le i \le d-1$, R(a, i) is contained in R(b, i).

Lemma 6.1: For $0 \le k \le d(m-1)$, every consistent and promising k-partial coloring X of a (d, m)-distribution can be extended to a consistent and balanced coloring.

Sketch of Proof: By double induction on d and m. If d=1, the lemma is trivial. Suppose that $d \ge 2$.

If k=0, then X is a consistent coloring and it remains to show that it is balanced. For $0 \le i \le d-1$, $A_i(X)=0$ and we must show that for $0 \le i \le d-1$, $B_i(X)=m$. If this is not the case, let i be such that $B_i(X)>m$ but $B_{i-1}(X) \le m$. Then $|R(a(X), i)| \ge 2$ but $|R(b(X), i)|=1$, contradicting the assumption that X is promising. Suppose that $k \ge 1$.

Let i be the color of box k-1. If i belongs to R(a(X), i-1) or there are fewer than $\Delta(a(X), i-1)$ uncolored balls in box k-1, then color all uncolored balls in box k-1 with color i-1. It can be shown that the resulting (k-1)-partial coloring is consistent and promising, and thus by inductive hypothesis can be extended to a consistent and balanced coloring, completing the proof in this case.

If i does not belong to R(a(X), i-1) and there are at least $\Delta(a(X), i-1)$ uncolored balls in box k-1, then color $\Delta(a(X), i-1)$ of these balls with color i-1. For $0 \le j \le d-1$, move all balls with color j to the cyclically succeeding box with color j. Partition the boxes into two classes: those with colors in R(a(X), i-1) and those with other colors. It can be shown that the boxes in the each class form a consistent and promising k'-partial coloring of a (d', m)-distribution for some $1 \le d' \le d-1$ and $0 \le k' \le d'(m-1)$. By inductive hypothesis, these partial colorings can be extended to consistent and balanced colorings. Restoring all balls to their initial positions yields a consistent and balanced extension of X and completes the proof of the lemma. □

Proof of Proposition 6.1: By induction on d. If d=1, the proposition is trivial. Suppose that $d \ge 2$.

If there do not exist d cyclically consecutive empty boxes, let $0 \le i \le d-1$ be a color such that there are $\ell \le m$ balls in boxes of color i. Color these ℓ balls with color i. There must be at least $m-\ell$ empty boxes of color i. For each such empty box, at least one of the d-1 cyclically preceding boxes must contain at least one ball. Color $m-\ell$ such balls with color i. At this point we have colored m balls with color i. Furthermore, removing all boxes and balls with color i yields a (d-1, m)-distribution which, by inductive hypothesis, has a consistent and balanced coloring. Restoring the boxes and balls with color i yields a consistent and balanced coloring of the original (d, m)-distribution and completes the proof in this case.

If there exist d cyclically consecutive empty boxes, cyclically shift the boxes and their contents so that boxes d(m-1), ..., dm-1 are empty. This yields a consistent and promising (d(m-1))-partial coloring which, by Lemma 6.1, can be extended to a consistent and balanced coloring. Restoring the boxes and their contents to their initial positions yields a consistent and balanced coloring of the original (d, m)-distribution and completes the proof of Proposition 6.1. □

Proof of Theorem 6: By induction on k. If k=1, the theorem is established by a d_1-crossbar. For $k\geq 2$, let $d=d_1$ and $m=d_2 \ldots d_k$. Let F be the (dm)-network with depth 1 and size d^2m having inputs {0, ..., dm-1}, outputs {0, ..., dm-1} and, for $0\leq i\leq dm-1$, edges from input i to outputs i, ..., i+d-1 (modulo dm). By inductive hypothesis, let G be a $(d_2 \ldots d_k)$-generalizer with depth k-1 and size $(d_2+\ldots+d_k)d_2 \ldots d_k$. Let H be the (dm)-network obtained by identifying, for each $0\leq j\leq d-1$, the outputs of F that are congruent to j modulo d with the inputs of a copy of G. The network H has depth k and size $d^2m + d(d_2+\ldots+d_k)d_2 \ldots d_k = (d_1+\ldots+d_k)d_1 \ldots d_k$ and can be shown, using Proposition 6.1, to be a $(d_1 \ldots d_k)$-generalizer. □

7. Generalized Connectors of Limited Depth

Let f(n) denote the minimum possible size of an n-connector, and let $f_k(n)$ denote the minimum possible size of an n-connector with depth at most k.

Using a classical construction due to D. Slepian, A. M. Duguid and J. Le Corre, Pippenger [20] showed that

$$f_{2k-1}(n) \leq 2k(1/2)^{1/k}\, n^{1+1/k} + O(n)$$

for each fixed k, and that

$$f_k(n) \leq 6n \log_3 n + O(n(\log n)^{1/2})$$

for some $k=2\log_3 n + O(\log\log n)$. The second of these results is within a constant factor of the best possible; Pippenger [21] has shown that

$$f(n) \geq 6n \log_6 n + O(n).$$

For the first result, the sharpest lower bound available is

$$f_k(n) \geq kn^{1+1/k},$$

due to Pippenger and Yao [24]. Note that the exponent 1+1/k applies to networks of depth 2k-1 in the upper bound, but depth k in the lower bound. Pippenger and Yao [24] have shown that the lower bound lies closer to the truth, in that

$$f_k(n) = O(n^{1+1/k} (\log n)^{1/k})$$

for each fixed k.

Let g(n) denote the minimum possible size of a generalized n-connector, and let $f_k(n)$ denote the minimum possible size of a generalized n-connector with depth at most k.

Clearly,

$$g(n) \geq f(n) \geq 6n \log_6 n + O(n).$$

and

$$g_k \geq f_k(n) \geq kn^{1+1/k}.$$

No lower bounds better than these are yet available.

If the outputs of an n-generalizer are identified with the inputs of an n-connector, the resulting network is a generalized n-connector. Thus,

$$g(n) \leq e(n) + f(n)$$

and

$$g_{k+\ell}(n) \leq e_k + f_\ell(n).$$

Using these inequalities to combine our bounds for generalizers in the preceding section with the bounds for connectors cited above (together with an ad hoc trick for reducing depth and size), we obtain the following result.

Corollary 6.1: By a simple explicit construction,

$$g_{3k-2}(n) \leq 3k(4/9)^{1/k}\, n^{1+1/k} + O(n)$$

for each fixed k, and

$$g_{3k-2}(n) \leq 9n \log_3 n + O(n(\log n)^{1/2})$$

for some $k=\log_3 n + O(\log\log n)$.

For completeness, two other explicit constructions for generalized connectors, not involving generalizers, should be mentioned. Firstly, Masson and Jordan [12] have given a construction that can be used to show that

$$g_3(n) = O(n^{5/3}).$$

Secondly, Nassimi and Sahni [14] have given a quite different construction that can be used to show that

$$g_k(n) = O(kn^{1+1/\sigma(k)}),$$

where σ(k) is defined by σ(1)=1 and the recurrence

$$\sigma(k) = 1 + \max_{1\leq j\leq k-1} (1 - 1/j)\, \sigma(k-j).$$

Both constructions give generalized n-connectors with depth 3 and size $O(n^{5/3})$; no other explicit constructions for generalized connectors with depth 3 are known. For $k\geq 4$, $\sigma(k)<\lfloor (k+2)/3 \rfloor$, so the construction of Nassimi and Sahni is inferior to Corollary 6.1 above. In fact, $\sigma(k)=O(k^{1/2})$, so that even for large k, their construction gives only

$$g_k(n) = O(n(\log n)^2)$$

for some $k=O((\log n)^2)$.

Our final result extends the upper bound of Pippenger and Yao [24] from connectors to generalized connectors.

Theorem 7: For each fixed k,

$$g_k(n) = O((n \log n)^{1+1/k}).$$

Sketch of Proof: An ($\underline{r}$, $\underline{s}$,)-restricted generalized n-connector is an n-network in which, for any one-to-many correspondence between certain inputs and disjoint sets of outputs in which each input i corresponds to $d(i)\le r$ outputs, there exist vertex-disjoint paths joining each input i with $d(i)\ge s+1$ to $d(i)-s$ corresponding outputs. Let $g_k(n, r, s)$ denote the minimum possible size of an (r, s)-restricted generalized n-connector with depth at most k. The inequalities $g_k(n) \le g_k(n, n, 0)$, $g_k(n, r, t) \le g_k(n, r, s) + g_k(n, s, t)$ and $g_2(n, r, 0) \le 2n\lfloor n/r \rfloor$ are immediate.

An ($\underline{a}$, $\underline{b}$)-partial ($\underline{r}$, $\underline{s}$)-restricted generalized n-connector is an n-network in which, for any $b+1\le m\le a$ and any one-to-many correspondence between certain inputs and disjoint sets of outputs in which each input i corresponds to $d(i)\le r$ outputs and m inputs i correspond to $d(i)\ge s+1$ outputs, there exist vertex-disjoint trees joining m-b inputs i to $d(i)-s$ corresponding outputs. Let $g_k(n, r, s, a, b)$ denote the minimum possible size of an (a, b)-partial (r, s)-restricted generalized n-connector with depth at most k. The inequalities $g_k(n, r, s) \le g_k(n, r, s, \lfloor n/r \rfloor, 0)$, $g_k(n, r, s, a, c) \le g_k(n, r, s, a, b) + g_k(n, r, s, b, c)$ and $g_2(n, r, s, a, 0) \le 2na$ are immediate.

The basic building blocks for our construction of partial restricted generalized connectors will be strong couplers, which we shall adapt from Pippenger and Yao [24].

A set $P=\{X_1, \ldots, X_r\}$ is an x-packing of a set A if $X_1, \ldots, X_r$ are mutually disjoint x-element subsets of A. An x-packing P of A is tight if $|P|\ge|A|/16$.

If G is a network and X a set of inputs, let G(X) denote the set of outputs reachable through paths in G from inputs in X.

An ℓ-network G is an (ℓ, x, y)-coupler if, for every tight x-packing $P=\{X_1, \ldots, X_r\}$ of the inputs of G, there exists a tight y-packing $Q=\{Y_1, \ldots, Y_s\}$ of the outputs of G such that, for every $1\le j\le s$, there exists $1\le i\le r$ such that Y_j is contained in $G(X_i)$.

A synchronous m-network G is a strong ($\underline{m}$, $\underline{x}$, $\underline{y}$)-coupler if, for every $m/2\le\ell\le m$, each ℓ-network obtained from G by deleting $m-\ell$ vertices from each rank (together with all edges incident with these vertices) is an (ℓ, x, y)-coupler. Let $h_k(m, x, y)$ denote the minimum possible size of a strong (m, x, y)-coupler with depth at most k.

Lemma 7.1: If $512x \ln m \le y \le m/16$, then

$$h_1(m, x, y) \le 32my/x.$$

The proof is almost identical to the proof of Proposition 3.1 in [24].

Using the inequality

$$h_{k+\ell}(m, x, z) \le h_k(m, x, y) + h_\ell(m, y, z),$$

we have the following lemmas.

Lemma 7.2: If $512 \ln m \le x$ and $x^{k-1} \le m/16$, then

$$h_{k-2}(m, x, x^{k-1}) \le 32(k-2)mx.$$

Lemma 7.3: If $512 \ln m \le x$ and $xy^{k-3} \le m/16$, then

$$h_{k-3}(m, x, xy^{k-3}) \le 32(k-3)my.$$

These lemmas allow us to prove the following propositions.

Proposition 7.1: For every $k\ge2$, there is a function $\phi(r, n)=O((r \log n)^{k-1})$ such that

$$g_k(n, r, \lfloor r/2 \rfloor, a, \lfloor a/2 \rfloor) = O(n(ar \log n)^{1/k})$$

for all n, r, and a such that $ar\le n$ and $a\ge\phi(r, n)$.

Proof: Choose $x=\lceil((7/2)ar \ln n)^{1/k}\rceil$. The conditions $x\ge(7/2)r \log_{e/2} n$, $x\ge(15/2) \log_{e/2} n$, $x\ge512 \ln ar$ and $x^{k-1}\le ar/16$ can be ensured by the hypothesis $a\ge\phi(r, n)$ for some function $\phi(r, n)=O((r \log n)^{k-1})$. By Lemma 7.2, there exists a strong (ar, x, x^{k-1})-coupler G with depth k-2 and size at most 32(k-2)arx. Let the n-network H be obtained by adjoining an edge from each input of H to each input of G independently with probability $p=8x/ar$ and an edge from each output of G to each output of H independently with probability p. The depth of H is k. It can be shown (along the lines of the proof of Proposition 4.1 in [24]) that with probability at least 1/2, the size of H is at most $32(k-2)arx + 4narp = O(n(ar \log n)^{1/k})$ and H is an (a, $\lfloor a/2 \rfloor$)-partial (r, $\lfloor r/2 \rfloor$)-restricted generalized n-connector. □

Proposition 7.2: For every $k\ge3$, there is a function $\psi(n)=O((\log n)^{k-2})$ such that

$$g_k(n, r, \lfloor r/2 \rfloor, a, \lfloor a/2 \rfloor) = O(n(a \log n)^{1/(k-1)})$$

for all n, r and a such that $ar\le n$ and $a\ge\psi(n)$.

Proof: Choose $x=\lceil r((15/2)a \ln n)^{1/(k-1)}\rceil$ and $y=\lceil((15/2)a \ln n)^{1/(k-1)}\rceil$. The conditions $x\ge(7/2)r$

$\log_{e/2} n$, $y \geq (15/2) \log_{e/2} n$, $y \geq 512 \ln ar$ and $xy^{k-3} \leq ar/16$ can be ensured by the hypothesis $a \geq \psi(n)$ for some function $\psi(n)=O((\log n)^{k-2})$.

Choose $m=\lfloor n/r \rfloor$. By Lemma 1.1, there exists an (n, 2m, m)-concentrator F with depth 1 and size at most $9n \log_2 n$. By Lemma 7.3, there exists a strong (ar, x, xy^{k-3})-coupler G with depth k-3 and size at most 32(k-3)ary. Let H be the n-network obtained by identifying the inputs of F with the inputs of H, adjoining an edge from each output of F to each input of G independently with probability $p=8x/ar$ and adjoining an edge from each output of G to each output of H independently with probability $q=8y/ar$. The depth of H is k. It can be shown that with probability at least 1/2, the size of H is at most $9n \log_2 n + 32(k-3)ary + 2marp + 2narq = O(n(a \log n)^{1/(k-1)})$ and H is an $(a, \lfloor a/2 \rfloor)$-partial $(r, \lfloor r/2 \rfloor)$-restricted generalized n-connector. □

We can now complete the proof of Theorem 7. If k=2, we use Proposition 7.1 for r and a equal to integral powers of 2 meeting its conditions. The total size of these networks is $O((n \log n)^{3/2})$. The bounds $g_2(n, r, 0) \leq 2n\lfloor n/2 \rfloor$ and $g_2(n, r, s, a, 0) \leq 2na$ dispose of the remaining cases.

If k≥3, we again use Proposition 7.1 for r and a equal to integral powers of 2 meeting its conditions. The total size of these networks is $O((n \log n)^{1+1/k})$. We then use Proposition 7.2 for those remaining cases meeting its conditions. The total size of these networks is $O(n^{1+1/k}(\log n)^{1/k+1/(k-1)})$. The bound $g_2(n, r, s, a, 0) \leq 2na$ disposes of the further remaining cases. This completes the proof of Theorem 7. □

5. References

[1] L. A. Bassalygo, "Asymptotically Optimal Switching Circuits", Prob. Info. Trans., 17 (1981) 206-211.

[2] A. K. Chandra, S. Fortune and R. Lipton, "Unbounded Fan-In Circuits and Associative Functions", ACM Symp. on Theory of Computing, 15 (1983).

[3] A. K. Chandra, S. Fortune and R. Lipton, "Lower Bounds for Constant Depth Monotone Circuits for Prefix Functions", Internat. Colloq. on Autom., Lang. and Prog., 10 (1983).

[4] K. M. Chung and C. K. Wong, "Construction of a Generalized Connector with 5.8n $\log_2$ n Edges", IEEE Trans. on Comp., 29 (1980) 1029-1032.

[5] M. Furst, J. B. Saxe and M. Sipser, "Parity, Circuits and the Polynomial-Time Hierarchy", IEEE Symp. on Found. of Comp. Sci., 22 (1981) 260-270.

[7] Z. Galil and W. J. Paul, "An Efficient General-Purpose Parallel Computer", ACM Symp. on Theory of Comp., 13 (1981) 247-262.

[8] V. A. Garmash and L. A. Shor, "Multistage One-Shot Batch Switching Arrangements", Prob. of Info. Trans., 13 (1978) 320-323.

[9] T. Lengauer and R. E. Tarjan, "Asymptotically Tight Bounds on Time-Space Trade-Offs in a Pebble Game", J. ACM, 29 (1982) 1087-1130.

[10] G. Lev, "Size Bounds and Parallel Algorithms for Networks", Thesis, Department of Computer Science, University of Edinburgh, September 1980.

[12] G. M. Masson and B. W. Jordan, Jr., "Generalized Multi-Stage Connection Networks", Networks, 2 (1972) 191-209.

[13] S. Nakamura and G. M. Masson, "Lower Bounds on Crosspoints in Concentrators", IEEE Trans. on Comp., (to appear).

[14] D. Nassimi and S. Sahni, "Parallel Permutation and Sorting Algorithms and a New Generalized Connection Network", J. ACM, 29 (1982) 642-667.

[15] Yu. P. Ofman, "A Universal Automaton", Trans. Moscow Math. Soc., 14 (1965) 200-215.

[16] W. J. Paul, R. E. Tarjan and J. R. Celoni, "Space Bounds for a Game on Graphs", Math. Sys. Theory, 10 (1977) 239-251.

[17] M. S. Pinsker, "On the Complexity of a Concentrator", Internat. Teletraffic Conf., 7 (1973) 318/1-4.

[18] N. Pippenger, "Superconcentrators", SIAM J. Comp., 6 (1977) 298-304.

[19] N. Pippenger, "Generalized Connectors", SIAM J. Comp., 7 (1978) 510-514.

[20] N. Pippenger, "On Rearrangeable and Non-Blocking Switching Networks", J. Comp. and Sys. Sci., 17 (1978) 145-162.

[21] N. Pippenger, "A New Lower Bound for the Number of Switches in Rearrangeable Networks", SIAM J. Alg. and Disc. Meth., 1 (1980) 164-167.

[22] N. Pippenger, "Superconcentrators of Depth 2", J. Comp. and Sys. Sci., 24 (1982) 82-90.

[23] N. Pippenger, "Advances in Pebbling", Internat. Colloq. on Autom., Lang. and Prog., 9 (1982) 407-417.

[24] N. Pippenger and A. C.-C. Yao, "Rearrangeable Networks with Limited Depth", SIAM J. Alg. and Disc. Meth., to appear.

[25] R. E. Tarjan, "Efficiency of a Good but Not Linear Disjoint Set Union Algorithm", J. ACM, 22 (1975) 215-225.

[26] C. D. Thompson, "Generalized Connection Networks for Parallel Processor Interconnection", IEEE Trans. on Comp., 27 (1978) 1119-1125.

[27] M. Tompa, "Time-Space Tradeoffs for Computing Functions, Using Connectivity Properties of Their Circuits", J. Comp. and Sys. Sci., 20 (1980) 118-132.

[28] L. G. Valiant, "Graph-Theoretic Properties in Computational Complexity", J. Comp. and Sys. Sci., 13 (1976) 278-285.