1-1-1994

Fault-Tolerant Circuit-Switching Networks

Nicholas Pippenger
Harvey Mudd College

Geng Lin
Bell-Northern Research

Recommended Citation

Nicholas Pippenger and Geng Lin. "Fault-Tolerant Circuit-Switching Networks", Society for Industrial and Applied Mathematics Journal of Discrete Mathematics, 7, 108 (1994).

SIAM J. DISC. MATH.
Vol. 7, No. 1, pp. 108–118, February 1994

FAULT-TOLERANT CIRCUIT-SWITCHING NETWORKS*

NICHOLAS PIPPENGER† AND GENG LIN‡

Abstract. The authors consider fault-tolerant circuit-switching networks under a random switch failure model. Three circuit-switching networks of theoretical importance—nonblocking networks, rearrangeable networks, and superconcentrators—are studied. The authors prove lower bounds for the size (the number of switches) and depth (the largest number of switches on a communication path) of such fault-tolerant networks and explicitly construct such networks with optimal size $\Theta(n(\log n)^2)$ and depth $\Theta(\log n)$.

Key words. nonblocking networks, rearrangeable networks, superconcentrator

AMS subject classifications. 94C15, 68E10, 05C35

1. Introduction.

In this paper, we study some fault-tolerant circuit-switching networks under a random switch failure model. In this model, each electrical switch in the network is independently in one of the following three states: (1) *open failure* (the switch is permanently *off* and fails to be *on*) with probability $0 < \varepsilon_1 < \frac{1}{2}$, (2) *closed failure* (the switch is permanently *on* and fails to be *off*) with probability $0 < \varepsilon_2 < \frac{1}{2}$, and (3) *normal state* (the switch functions correctly) with probability $1 = \varepsilon_1 - \varepsilon_2$. For simplicity of notation, we assume that $\varepsilon_1 - \varepsilon_2 = \varepsilon$. The measure of fault tolerance is the probability of the network fulfilling the communication task in the presence of switch failures. This model is essentially equivalent to that of Moore and Shannon [MS], who introduced it in the context of relay circuits computing Boolean functions. The model retains its relevance, since open and closed failures represent the two dominant failures modes both for metallic-contact switches (still frequently used, especially for video switching) and MOSFETs (metal-oxide semiconductor field-effect transistors), a common switching element in VLSI circuits.

2. The networks.

The circuit-switching networks we study in this paper are *nonblocking networks*, *rearrangeable networks*, and *superconcentrators*. Nonblocking networks were introduced by Clos [Cl] in 1953 to epitomize the activity of telephone communication. Beneš [B] in 1964 described the rearrangeable network. Rearrangeable networks are useful architectures for parallel machines. Aho, Hopcroft, and Ullman [AHU] in 1974 posed the problem of *superconcentrators*. Although their purpose was to hope to use them to establish a nonlinear lower bound for the Boolean circuit complexity of multiplication, superconcentrators proved to be central in a number of communication networks. For example, superconcentrators provide support for the task queue scheme (see [Co]) in parallel computing. Tremendous efforts on these networks have been made, and significant results obtained.

In this paper, we describe a circuit-switching network in terms of an acyclic directed graph. *Terminals* of the network (wires that connect the network to the outside world) are represented by distinguished vertices called *inputs* and *outputs*. Electrical links are represented by vertices other than inputs and outputs, and switches (single-pole single throw, connecting two links) by edges between the two corresponding vertices. The three states of a switch in the random switch failure model are therefore interpreted as (1) the edge ceases to exist (open failure), (2) two vertices of the edge contract to one (closed

* Received by the editors April 20, 1992; accepted for publication (in revised form) October 15, 1992.

† Department of Computer Science, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada, (nicholas@cs.ubc.ca).

‡ Bell-Northern Research, Box 3511, Station C, Ottawa, Ontario K1Y 4H7, Canada, (geng@bnr.ca).

failure), and (3) the edge is unaffected (normal state). In this paper, we say "graph" and "network" without distinction, and the same is true for "edge" and "switch."

Given a directed graph with n inputs and n outputs, it is said to be a "*nonblocking n-network*" if, given any set of vertex-disjoint paths from inputs to outputs and given any input and output not involved in these established paths, a new path that is vertex-disjoint from the established paths can be found from the requesting input to the requesting output; it is said to be a "*rearrangeable n-network*" if, given any one-to-one correspondence between the inputs and the outputs, there exists a set of n vertex-disjoint paths joining each input to its corresponding output; it is said to be an "*n-superconcentrator*" if, for every $r \leq n$, every set of r inputs, and every set of r outputs, there exists a set of r vertex-disjoint paths from the given inputs to the given outputs. It is obvious that a nonblocking n-network is a rearrangeable n-network, and a rearrangeable n-network is an n-superconcentrator.

The networks considered in this paper are based on directed graphs and distinguish the roles of inputs and outputs as terminals. Variants of these definitions exist for networks based on undirected graphs, and for which there is but one class of terminals. Our definition of "nonblocking" is also referred to as "strictly nonblocking," to distinguish it from the somewhat weaker notion of "wide-sense nonblocking" that also appears in the literature. The networks we call "rearrangeable" are sometimes referred to as "permutation" networks, though the latter term is also used for some variants of this notion.

The measures of complexity applied to such networks are *size* (the number of edges) and *depth* (the largest number of edges on any path from an input to an output). Shannon [S] showed an $\Omega(n \log n)$ size lower bound of rearrangeable n-networks. Beneš [B] presented an $O(n \log n)$ size and $O(\log n)$ depth construction for rearrangeable n-networks. The existence of $O(n \log n)$ size and $O(\log n)$ depth nonblocking n-networks was proved by Bassalygo and Pinsker [BP]. For n-superconcentrators, an $\Omega(n)$ size lower bound is obvious, and Valiant [V] showed an $O(n)$ size upper bound.

3. Fault tolerance. Given $0 < \varepsilon < \frac{1}{2}$, consider a network N subject to the random switch failure model. Let the event space Ω be the set of all graphs obtained from N. The probability measure on each graph is assigned in accordance to the number of failed edges. More precisely, if a graph $G \in \Omega$ has k failed edges, the probability that the random instance of N equals G is $(2\varepsilon)^k(1-2\varepsilon)^{n-k}$, where n is the number of edges in N. Given $0 < \delta < 1$, we say that N is an (ε, δ)-*nonblocking n-network* if the probability that the random instance of N contains a nonblocking n-network consisting of edges of normal state is greater than $1 - \delta$. Similarly, we define an (ε, δ)-*n-rearrangeable network* and an (ε, δ)-*n-superconcentrator*. We observe that an (ε, δ)-nonblocking n-network is an (ε, δ)-rearrangeable n-network, and an (ε, δ)-rearrangeable n-network is an (ε, δ)-n-superconcentrator. It is clear that, by choosing arbitrarily small δ, an (ε, δ)-nonblocking n-network or an (ε, δ)-rearrangeable n-network or an (ε, δ)-n-superconcentrator can fulfill its communication task with arbitrarily high probability.

The goal of this paper is to analyze the asymptotic behaviors of the size and depth of the (ε, δ)-nonblocking n-network, the (ε, δ)-rearrangeable n-network, and the (ε, δ)-*n-superconcentrator*. For this purpose, the exact values of $0 < \varepsilon < \frac{1}{2}$ and $0 < \delta < 1$ do not matter. To see this, we first need a result of Moore and Shannon [MS].

Define an $(\varepsilon, \varepsilon')$-1-*network* to be a directed graph with two distinguished vertices called *input* and *output*, in which each edge is randomly and independently subject to closed and open failures with probabilities of ε, respectively, and in which the probabilities that the input and the output contract into one vertex and that there is no path from the input to the output are both less than ε'.

PROPOSITION 1 (Moore and Shannon). *Given $0 < \varepsilon < \frac{1}{2}$ and $0 < \varepsilon' \leq \varepsilon$, there is an explicit construction of an $(\varepsilon, \varepsilon')$-1-network with $c_\varepsilon(\log_2(1/\varepsilon'))^2$ edges and $d_\varepsilon \log_2(1/\varepsilon')$ depth, where c_ε and d_ε are constants depending only on ε.*

To observe the fact that the exact value of ε does not affect the asymptotic behaviors of the size and depth, we suppose that $0 < \varepsilon_1 \leq \varepsilon_2 < \frac{1}{2}$ and that Φ is an (ε_1, δ)-n-superconcentrator with size L and depth D, for some $\delta < 1$. By Proposition 1, there is an $(\varepsilon_2, \varepsilon_1)$-1-network Ψ of size a and depth b (a and b are depending only on ε_2). The result of substituting this network Ψ for each edge in Φ is clearly an (ε_2, δ)-n-superconcentrator with size at most aL and depth at most bD. Similar arguments apply to (ε, δ)-rearrangeable n-networks and (ε, δ)-nonblocking n-networks as well.

To see the invariance with respect to the value of δ, we suppose that $0 < \delta_1 \leq \delta_2 < 1$ and that Φ is an (ε, δ_2)-n-superconcentrator, for some $\varepsilon < \frac{1}{2}$. The failure probability of Φ is a polynomial in ε and the constant term of this polynomial vanishes (since the network does not fail unless some switch fails). If we replace ε by $\varepsilon\delta_1/\delta_2$, every term in this polynomial decreases to at most δ_1/δ_2 times its previous value. Thus Φ is also an $(\varepsilon\delta_1/\delta_2, \delta_1)$-$n$-superconcentrator. Again, substitute each edge in Φ by an $(\varepsilon, \varepsilon\delta_1/\delta_2)$-1-network and the resulting network is an (ε, δ_1)-n-superconcentrator with the size and depth being affected by only a constant factor. Similar arguments apply to (ε, δ)-rearrangeable n-networks and (ε, δ)-nonblocking n-networks as well.

4. Main result and the overall strategy. In this paper, we show that the size and depth of (ε, δ)-n-superconcentrators, (ε, δ)-rearrangeable n-networks, and (ε, δ)-nonblocking n-networks are $\Theta(n(\log n)^2)$ and $\Theta(\log n)$.

The overall strategy is that we prove the $\Omega(n(\log n)^2)$ and $\Omega(\log n)$ lower bounds for size and depth of a $(\frac{1}{4}, \frac{1}{2})$-$n$-superconcentrator, and we construct $(10^{-6}, \delta)$-nonblocking n-networks with $O(n(\log n)^2)$ size and $O(\log n)$ depth for arbitrarily small δ. The success of this strategy is ascribed to an observation we made earlier, that, for any $0 < \varepsilon < \frac{1}{2}$ and $0 < \delta < 1$, an (ε, δ)-nonblocking n-network is an (ε, δ)-rearrangeable n-network, and an (ε, δ)-rearrangeable n-network is an (ε, δ)-n-superconcentrator. Thus a lower bound (for size or depth) of the (ε, δ)-n-superconcentrator is a lower bound of all three, and an upper bound of the (ε, δ)-nonblocking n-network is an upper bound of all three.

The lower bounds are proved in §5. In §6 we construct the (ε, δ)-nonblocking n-network. A few observations on our upper bound are in order. First, the upper bound is based on an explicit construction and is not merely an existence proof. Second, with high probability we can find a nonblocking network contained in the fault-tolerant network merely by discarding faulty components and their immediate neighbors, so no difficult computations are hidden here. Third, because the contained network is "strictly" nonblocking (see Feldman, Friedman, and Pippenger [FFP] for details), routing can be performed by a "greedy" application of a standard path-finding algorithm, so again no difficult computations are involved.

5. The lower bounds. The strategy of the lower bound proof is as follows. We associate with each input a neighborhood containing all vertices within a logarithmic distance of the input. We show that, for a large set of inputs, these neighborhoods are disjoint (otherwise, two inputs would be shorted by closed failures with high probability). This gives the lower bound for depth. We then partition the vertices in the neighborhoods of these inputs into zones according to their distance from the input. We show that, for a large number of inputs, each of these zones must have logarithmic size (otherwise, some input would be isolated by open failures with high probability). Summing over the zones of each neighborhood and the neighborhoods of the inputs gives the lower bound for size.

Given a graph G, we say the distance from vertex ξ_1 to vertex ξ_2, dist (ξ_1, ξ_2), is the number of edges in the shortest path (not necessarily directed) from ξ_1 to ξ_2; the distance from a vertex ξ to an edge $e = \langle \nu, \mu \rangle$, dist (ξ, e), is min $\{\text{dist }(\xi, \mu), \text{dist }(\xi, \nu)\} + 1$.

LEMMA 1. *A tree with l leaves, in which every internal node has degree at least* 3, *contains at least $l/42$ edge-disjoint paths, each joining* 2 *leaves, and each having length at most* 3.

Proof. We begin by observing that we may assume that each internal node has degree exactly 3. For, if not, we may replace each internal node with degree $d > 3$ by a "tree" comprising $d - 2$ new nodes with degree 3. If we find a set of edge-disjoint paths of length at most 3 in the resulting tree, these will correspond to edge-disjoint paths of no greater length in the original graph. Suppose then that T is a tree with l leaves in which each internal node has degree 3. Clearly, there must be $l - 2$ internal nodes. Let us say that a leaf L is "bad" if there is no other leaf with distance at most 3 from L. We show that there are at most $6l/7$ bad leaves. If L is bad, there are seven internal nodes with distance at most 3 from L (see Fig. 1). Let L "pay" one dollar to each of those nodes. We claim that each of the $l - 2$ internal nodes "collects" at most six dollars, from which it follows that there are at most $6(l - 2)/7 \leq 6l/7$ bad leaves. If some internal node V collects more than six dollars from bad leaves at distance at most 3, then more than one of these bad leaves must be adjacent to one of the six or fewer nodes at distance 2 from V. However, no more than one bad leaf can be adjacent to an internal node (see Fig. 2). Thus at least $l/7$ leaves are "good" (that is, not bad). Suppose that there are m good leaves. Let $\mathscr{L}$ be a maximal set of edge-disjoint paths, each joining two good leaves and each having length at most 3. Say that a good leaf is "lucky" if it is the endpoint of a path in $\mathscr{L}$, and that it is "unlucky" otherwise. If L is unlucky, there must be a path P in $\mathscr{L}$ within distance 2 of L. (There is a leaf within distance 3 of L, since L is good, and only a path in $\mathscr{L}$ could prevent L from being joined to such a leaf in the maximal set $\mathscr{L}$.) Let each unlucky leaf "pay" one dollar to some such path P. Each path P "collects" at most four dollars from unlucky leaves, since there are at most four leaves with distance at most 2 from P (see Fig. 3). It follows that there are at most four unlucky leaves for each path in $\mathscr{L}$. Since there are exactly two lucky leaves for each path in $\mathscr{L}$, and $m \geq l/7$ good leaves (lucky and unlucky), this implies that there are at most $m/6 \geq l/42$ paths in $\mathscr{L}$. □

Remark. The bound "$l/42$" in Lemma 1 can be improved to "$l/4$," but this requires a more elaborate analysis, which will be presented elsewhere (see Lin [L]).

COROLLARY 1. *A forest F of l leaves, in which every internal node has degree at least* 3, *contains at least $l/42$ edge-disjoint paths, each joining* 2 *leaves, and each having length at most* 3.

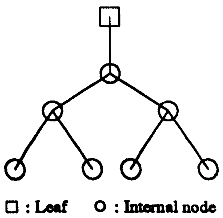

FIG. 1. *A bad leaf.*

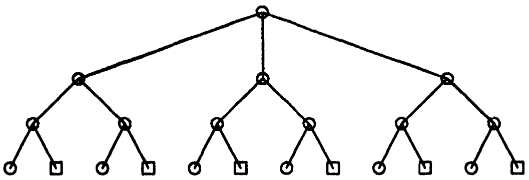

FIG. 2. *An internal node collects at most six dollars.*

LEMMA 2. *Let Φ be a $(\frac{1}{4}, \frac{1}{2})$-$n$-superconcentrator. For all sufficiently large n, at least $n/2$ inputs of Φ have distance (ignoring the direction of each edge) at leasi $(\frac{1}{8})\log_2 n$ from each other input.*

Proof. Suppose that each of $n/2$ inputs v has a path $\pi(v)$ of length at most j to some other input. We obtain a contradiction if $j = (\frac{1}{8})\log_2 n$ and n is sufficiently large. Define a forest by (1) starting with the empty forest, (2) considering each such input in turn, and (3) adding to the forest the longest initial segment of $\pi(v)$ that is edge-disjoint from the forest generated thus far. Thus the resulting forest F has at least $n/2$ leaves, and each "stretch" (sequence of consecutive vertices of degree 2) has length at most j. Let G be the forest obtained from F by replacing each stretch, together with the edges incident with its vertices, by a single edge. In G every internal node is of degree at least 3, so we may apply Corollary 1 to obtain at least $n/84$ edge-disjoint paths, each having length at most 3 and each joining one leaf to another. Replacing each edge of these paths by the corresponding stretch, we obtain in F at least $n/84$ edge-disjoint paths, each having length at most $3j$ and each joining one input of Φ to another. Note that, if each of the $3j$ edges on one of the $n/84$ paths is in the closed failure state, two inputs of Φ will contract to a single vertex, and the result will certainly not be an n-superconcentrator. Since this can happen with probability at most $\frac{1}{2}$, we have $1 - (1 - (\frac{1}{4})^{3j})^{n/84} < \frac{1}{2}$. If we set $j = (\frac{1}{6})\log_2(n/(84 \ln 2))$, we obtain a contradiction using the inequality $(1-x)^y < e^{-xy}$. Thus, if we set $j = (\frac{1}{8})\log_2 n$, we obtain a contradiction for all sufficiently large n. □

THEOREM 1. *Let Φ be a $(\frac{1}{4}, \frac{1}{2})$-$n$-superconcentrator. For all sufficiently large n, Φ has size at least $(\frac{1}{256})n(\log_2 n)^2$ and depth at least $(\frac{1}{16})\log_2 n$.*

Proof. Say an input is "good" if it has distance at least $(\frac{1}{8})\log_2 n$ from each other input. By Lemma 2, there are at least $n/2$ good inputs. (Note that the existence of two good inputs implies, by the triangle inequality of the distance, that the depth is at least $(\frac{1}{16})\log_2 n$.) For each good input v, let $B(v)$ denote the set of all edges at distance at

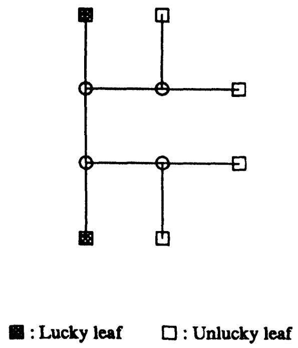

FIG. 3. *Each path collects at most four dollars from unlucky leaves.*

most $(\frac{1}{16})\log_2 n$ from v. For any pair v and w of good inputs, the sets $B(v)$ and $B(w)$ must be disjoint, since otherwise the distance between v and w would be less than $(\frac{1}{8})\log_2 n$. Thus it will suffice to show that, for each good input v, the set $B(v)$ contains at least $(\frac{1}{128})(\log_2 n)^2$ edges for all sufficiently large n. If an input v_0 has all n outputs adjacent to some edges in $B(v_0)$, then it is certainly true that $|B(v_0)| \geq (\frac{1}{128})(\log_2 n)^2$, since the number of edges in $B(v_0)$ cannot be less than the number of outputs adjacent to these edges, and $n \geq (\frac{1}{128})(\log_2 n)^2$ for all sufficiently large n. Thus we may assume that, for each good input v, there is an output $w(v)$ that is not adjacent to an edge in $B(v)$. Consider an arbitrary good input v. Set $i = \lfloor(\frac{1}{16})\log_2 n\rfloor \geq (\frac{1}{32})\log_2 n$. Partition $B(v)$ into subsets $B_1(v), \ldots, B_i(v)$, where $B_h(v)$ comprise those edges at distance h from v. Let $B^*(v)$ denote the set $B_h(v)$ with the minimum number of edges. It will suffice to show that each set $B^*(v)$ contains at least $(\frac{1}{4})\log_2 n$ edges. Let b be the cardinality of the set $B^*(v)$ with the minimum number of edges. It will suffice to show that $b \geq (\frac{1}{4})\log_2 n$ for all sufficiently large n. Consider an arbitrary good input v. Any path from v to $w(v)$ must contain an edge in $B^*(v)$, since the distance from v can increase at most 1 at each successive edge of a path. If edges of $B^*(v)$ are all in open state, all paths from v to $w(v)$ are broken, and the resulting network is certainly not an n-superconcentrator. This can happen with probability at most $\frac{1}{2}$. Thus we have $1 - (1 - (\frac{1}{4})^b)^{n/2} < \frac{1}{2}$. As before, this implies that $b \geq (\frac{1}{2})\log_2(n/2 \ln 2) \geq (\frac{1}{4})\log_2 n$ for all sufficiently large n. □

6. The upper bounds. In this section, we explicitly construct $(10^{-6}, \delta)$-nonblocking n-networks with $O(n(\log n)^2)$ edges and $O(\log n)$ depth for arbitrarily small δ.

The strategy of the upper bound proof is as follows. We use a standard recursive construction for nonblocking networks, but scale the construction up by a logarithmic factor and terminate the recursion with subnetworks of logarithmic size (rather than constant size). We then use networks (called "directed grids" in this paper) of logarithmic by logarithmic size based on the "hammock" of Moore and Shannon [MS] to interface the inputs and outputs to the terminal subnetworks.

The basic building blocks of the construction are *(c, c', t)-expanding graphs* and *(l, w)-directed grids*. A *(c, c', t)-expanding graph* is a bipartite directed graph with two distinguished sets of t vertices called *inlets* and *outlets*, respectively, where every set of c inlets is joined by edges to at least c' outlets (that is, for every set C of c inlets, there exist a set C' of c' outlets, such that, for every outlet ζ' and C', there is an inlet ζ in C and an edge (ζ, ζ')). The constructions of *(an, bn, n)-expanding graphs* (where $0 < a < b < 1$ are constants) with linear sizes (with respect to n) are quite standard. See Bassalygo and Pinsker [BP] for the probabilistic version, and see Gabber and Galil [GG] for the explicit construction. (We need to mention that the first explicit construction was presented by Margulis [M] and currently the best-known explicit construction is due to Lubotzky, Phillips, and Sarnak [LPS].) An *(l, w)-directed grid* is a directed graph with w stages and l vertices in each stage. A vertex in the jth stage and the ith row is denoted by a binary tuple (i, j), $1 \leq i \leq l$ and $1 \leq j \leq w$. An edge from vertex (i, j) to vertex (i', j') exists if and only if $i' = i$ and $j' = j + 1$ or $i' = i + 1$ and $j' = j + 1$. (See Fig. 4.)

Suppose that we wish to construct an (ε, δ)-nonblocking n-network with $n = 4^{\nu}$. Set $\gamma = \lceil\log_4(34\nu)\rceil$, so that $136\nu \geq 4^{\gamma} \geq 34\nu$. We first construct a nonblocking $4^{\nu+\gamma}$-network through the recursive construction illustrated in Pippenger [P82, §9]. (This network is a directed graph with $2(\nu + \gamma) + 1$ stages, with $4^{\nu+\gamma}$ inputs on stage 0 and $4^{\nu+\gamma}$ outputs on stage $2(\nu + \gamma)$. Each other stage contains $64 \cdot 4^{\nu+\gamma}$ vertices. Edges only exist between some vertices in adjacent stages. The subgraph induced by inputs and vertices in stage 1 consists of $4^{\nu+\gamma-1}$ disjoint bipartite graphs, each having four inputs

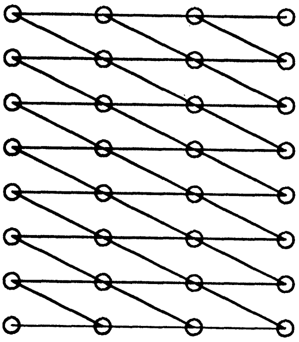

FIG. 4. *A* (4, 8)*-directed grid.*

on one side and 256 vertices on the other side. Similar property holds for the subgraph induced by outputs and vertices in the adjacent stage. The subgraph induced by vertices in stage i and stage $i + 1$ (for all $1 \le i \le \nu + \gamma - 1$) consists of $4^{\nu+\gamma-i}$ disjoint $(32 \cdot 4^i, 32(1 + (2 - \sqrt{3})/8) \cdot 4^i, 64 \cdot 4^i)$-expanding graphs, with each vertex on stage i having ten out-edges and vertex on stage $i + 1$ ten in-edges. The subnetwork from stage $\nu + \gamma$ to stage $2(\nu + \gamma)$ is a *mirror image* of that from stage 0 to stage $\nu + \gamma$. Network N' is a mirror image of network N if N' is obtained from N by (1) exchanging the inputs with the outputs and (2) reversing the direction of every edge.) We then remove vertices in the first and last γ stages and edges incident with them and let $\mathcal{M}$ be the remaining graph. The first stage of $\mathcal{M}$ consists of 4^ν disjoint sets vertices, each being the inlets of a $(32 \cdot 4^\gamma, 33.07 \cdot 4^\gamma, 64 \cdot 4^\gamma)$-expanding graph (note that $32(1 + (2 - \sqrt{3})/8) > 33.07$). Construct 4^ν $(\nu, 64 \cdot 4^\gamma)$-directed grids $\Phi_1, \dots, \Phi_{4^\nu}$. Joined to each vertex in the first stage of $\Phi_i (\, i = 1, \dots, 4^\nu)$ is an edge from an *input* vertex. Combine $\mathcal{M}$ with $\Phi_1, \dots, \Phi_{4^\nu}$ and the associated inputs by (1) letting the 4^ν $(32 \cdot 4^\gamma, 33.07 \cdot 4^\gamma, 64 \cdot 4^\gamma)$-expanding graphs in the first stage of $\mathcal{M}$ correspond to $\Phi_1, \dots, \Phi_{4^\nu}$ in any one-to-one fashion, and (2) in each such corresponding pair, identifying the inlets of the expanding graph with the vertices in the last stage of the directed grids in any one-to-one fashion. Similarly, construct 4^ν $(\nu, 64 \cdot 4^\gamma)$-directed grids $\Psi_1, \dots, \Psi_{4^\nu}$; join an *output* by edges to every vertex in the last stage of each $\Psi_j \, (j = 1, \dots, 4^\nu)$; combine $\Psi_1, \dots, \Psi_{4^\nu}$ and the associated outputs with $\mathcal{M}$ (and the above combined $\Phi_1, \dots, \Phi_{4^\nu}$ and the associated inputs) by identifying vertices of the first stage of $\Psi_1, \dots, \Psi_{2^\nu}$ with vertices in the last stage of $\mathcal{M}$. Call the resulting network $\mathcal{N}$. (See Fig. 5.)

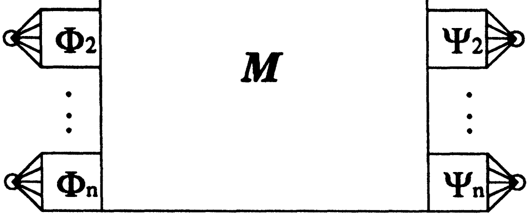

FIG. 5. *Network* $\mathcal{N}$.

Network $\mathcal{N}$ has $2(\nu + \nu) + 1 = 4\nu + 1$ stages. The 4^{ν} inputs and 4^{ν} outputs are on stage 0 and stage 4ν, respectively. Each other stage contains $64 \cdot 4^{\nu+\gamma}$ vertices. The subnetwork from stage 1 to stage ν and that from stage 3ν to stage $4\nu - 1$ consist of $\Phi_1, \ldots, \Phi_{4^{\nu}}$ and $\Psi_1, \ldots, \Psi_{4^{\nu}}$, respectively. The subnetwork from stage ν to stage 3ν is $\mathcal{M}$. Each input has out-degree $64 \cdot 4^{\gamma}$; a vertex in Φ_i has in-degree 2 and out-degree 2, except vertices on their first stages (in-degree 1) and last stages (out-degree 10); a vertex in the left-hand half of $\mathcal{M}$ (stage ν to stage 2ν of $\mathcal{N}$) has in-degree 10 and out-degree 10, except vertices on stage ν (in-degree 2). The subnetwork from stage 2ν to stage 4ν (called $\mathcal{N}_{\mathcal{R}}$) is a mirror image of that from stage 0 to stage 2ν (called $\mathcal{N}_{\mathcal{L}}$). In particular, the right-hand half of $\mathcal{M}$, called $\mathcal{M}_{\mathcal{R}}$, is a mirror image of $\mathcal{M}_{\mathcal{L}}$, the left-hand half of $\mathcal{M}$. Network $\mathcal{N}$ has $1408\nu 4^{\nu+\gamma}$ edges because there are $1280\nu 4^{\nu+\gamma}$ edges in $\mathcal{M}$, $128(\nu - 1)4^{\nu+\gamma}$ edges in Φ_i and Ψ_i, for all $1 \leq i \leq 4^{\nu}$, and $128 \cdot 4^{\nu+\gamma}$ edges adjacent to inputs and outputs.

Let η be a vertex of $\mathcal{N}$ that is not an input or an output. Say a vertex η of $\mathcal{N}$ is *faulty*, if an edge $\langle \tau, \eta \rangle$ or $\langle \eta, \xi \rangle$ is in open failure or closed failure state. Given a set of vertex-disjoint direct paths from inputs to outputs in $\mathcal{N}$, for an input, an output, or a vertex that is not faulty, it is said to be *idle* if it is not involved in these paths, *busy* otherwise. Say an (idle) vertex ξ_1 has *access* to another (idle) vertex ξ_2 if there is a path of idle vertices from ξ_1 to ξ_2. It is clear that, if ξ_1 has access to ξ_2 and ξ_2 has access to ξ_3, then ξ_1 has access to ξ_3. A network N is a *majority-access* network if, given any set of directed paths from inputs to outputs, every idle input has access to a majority (strictly more than half) of the outputs.

LEMMA 3. *Let ξ be an idle input of network $\mathcal{N}$. The probability that ξ has access to at least $32 \cdot 4^{\gamma} + 1$ vertices in the last stage of Φ_{ξ} (i.e., strictly more than half) is at least $1 - c_1\nu(144\varepsilon)^{\nu}$, where $c_1 = 1/(1 - 72\varepsilon)$.*

Proof. Let us begin by estimating the probability that ξ does not have access to any vertex at the last stage of Φ_{ξ}. There is no busy vertex in Φ_{ξ}, since ξ is idle and $\mathcal{N}$ is a directed and staged graph. By Menger's theorem (see, e.g., Chapter 5 of [CL]), there is a "(vertex) cut set" of Φ_{ξ} (i.e., the removal of which and their adjacent edges will separate ξ and the vertices at the last stage) consisting of faulty vertices only. Consider a cut set C of l vertices. Then it must be $l \geq 64 \cdot 4^{\gamma}$, since Φ_{ξ} has this many rows. If every vertex in C is faulty, the probability is at most $(24\varepsilon)^l$, since each vertex in Φ_{ξ} (other than ξ, which is not in any cut set we consider) is adjacent to at most twelve edges. For any given l, the number of such cut set C is at most $\nu 3^l$, since they are ν vertices at the first row of Φ_{ξ} to start C, and at most three ways (each along an edge) to continue at each step. Thus the probability that ξ does not have access to any vertex at the last stage of Φ_{ξ} is at most

$$\sum_{l \geq 64 \cdot 4^{\gamma}} \gamma 3^l (24\varepsilon)^l = c_1\nu(72\varepsilon)^{64 \cdot 4^{\gamma}}.$$

Consider an arbitrary set S of $32 \cdot 4^{\gamma}$ vertices in the last stage of Φ_{ξ}. The probability that ξ does not have access to any vertex in S is at most $c_1\nu(72\varepsilon)^{64 \cdot 4^{\gamma}}$. There are at most

$$\binom{64 \cdot 4^{\gamma}}{32 \cdot 4^{\gamma}} < 2^{64 \cdot 4^{\gamma}}$$

such S. This implies that the probability of ξ having access to at least $32 \cdot 4^{\gamma} + 1$ vertices in the last stage of Φ_{ξ} is at least $1 - c_1\nu(144\varepsilon)^{\nu}$, since $64 \cdot 4^{\gamma} > \nu$. □

LEMMA 4. *In a $(32 \cdot 4^{\mu}, 33.07 \cdot 4^{\mu}, 64 \cdot 4^{\mu})$-expanding graph in $\mathcal{M}_{\mathcal{L}}$, for any $\gamma \leq \mu \leq \nu + \gamma - 1$ (the expanding graph is in the subgraph from stage $\mu + \nu - \gamma$ to stage $\mu + \nu - \gamma + 1$ of $\mathcal{N}$), the probability that it has more than $0.07 \cdot 4^{\mu}$ outlets faulty is at most $e^{-0.06 \cdot 4^{\mu}}$.*

Proof. There are $1280\cdot 4^{\mu}$ edges incident with outlets of the $(32\cdot 4^{\mu}, 33.07\cdot 4^{\mu}, 64\cdot 4^{\mu})$-expanding graph (each vertex has ten in-edges and ten out-edges). For each such edge, let x_j be the random variable such that $x_j = 0$ if the edge is in normal state, $x_j = 0$ otherwise, for all $1 \le j \le 1280\cdot 4^{\mu}$. It is clear that $\Pr\,[x_j = 1] < 2\varepsilon$ and $\Pr\,[x_j = 0] > 1 - 2\varepsilon$. Let $T = \sum_{j=1}^{1280\cdot 4^{\mu}} x_j$,

$$\Pr\,[T > 0.07\cdot 4^{\mu}] = \Pr\,[e^{T} > e^{0.07\cdot 4^{\mu}}] < \mathrm{E}\,[e^{T}]/e^{0.07\cdot 4^{\mu}}$$

by Markov's inequality. As x_j's are independent,

$$\mathrm{E}\,[e^{T}] = \prod_{j=1}^{1280\cdot 4^{\mu}} \mathrm{E}\,[e^{x_j}] < (1 + 2\varepsilon e)^{1280\cdot 4^{\mu}} < e^{2560e\varepsilon\cdot 4^{\mu}},$$

since $(1 + x)^{y} < e^{xy}$, and $2560e\varepsilon < 0.01$ when $\varepsilon = 10^{-6}$. Thus the probability that there are more than $0.07\cdot 2^{\mu}$ outlets faulty is at most $e^{0.01\cdot 4^{\mu} - 0.07\cdot 4^{\mu}} = e^{-0.06\cdot 4^{\mu}}$. □

LEMMA 5. *The probability that there exists a* $(32\cdot 4^{\mu}, 33.07\cdot 4^{\mu}, 64\cdot 4^{\mu})$-*expanding graph in* $\mathcal{M}_{\mathcal{L}}$ *with more than* $0.07\cdot 4^{\mu}$ *faulty outlets, for some* $\gamma \le \mu \le \nu + \gamma - 1$, *is less than* $\nu(2/e)^{2\nu}$.

Proof. It is simply a problem of counting the number of expanding graphs with respect to the number of outlets. There are $4^{\nu+\gamma-\mu}$ $(32\cdot 4^{\mu}, 33.07\cdot 4^{\mu}, 64\cdot 4^{\mu})$-expanding graphs between stage $\nu + \mu - \gamma$ and stage $\nu + \mu - \gamma + 1$ of $\mathcal{M}_{\mathcal{L}}$, for all $\gamma \le \mu \le \nu + \gamma - 1$. By Lemma 4, the probability that there is an expanding graph with no more than $0.07\cdot 4^{\mu}$ outlets is at most

$$\sum_{\mu=\gamma}^{\nu+\gamma-1} 4^{\nu+\gamma-\mu} e^{-0.06\cdot 4^{\mu}} < \sum_{\mu=\gamma}^{\nu+\gamma-1} 4^{\nu} e^{-0.06\cdot 4^{\gamma}} < \nu 4^{\nu} e^{-2\nu},$$

since $4^{\gamma} \ge 34\nu$. □

LEMMA 6. *With probability at least* $1 - c_1\nu(144\varepsilon)^{\nu} - \nu(2/e)^{2\nu}$, $\mathcal{N}_{\mathcal{L}}$ *is a majority-access network.*

Proof. We may assume in this proof that each $(32\cdot 4^{\mu}, 33.07\cdot 4^{\mu}, 64\cdot 4^{\mu})$-expanding graph in $\mathcal{M}_{\mathcal{L}}$ has at least $0.07\cdot 4^{\mu}$ outlets faulty, and each idle input ξ has access to at least $32\cdot 4^{\gamma} + 1$ vertices in the last stage of Φ_{ξ}. It is clear by Lemmas 3 and 5 that the probability of the assumption failing is at most $1 - c_1\gamma(144\varepsilon)^{64\cdot 4^{\gamma}} - \nu(2/e)^{2\nu}$. For each pair of inputs ξ_1 and ξ_2, we say their relativity, relat (ξ_1, ξ_2) = relat (ξ_2, ξ_1), is d $(1 \le d \le \nu)$ if and only if the directed paths starting from the two inputs may share a vertex at or after the $(d + \nu)$th stage but cannot share any vertex before the $(d + \nu)$th stage. It is observed that, for each input ξ, there are $4^{d} - 4^{d-1} = 3\cdot 4^{d-1}$ other inputs ξ' with relat $(\xi, \xi') = d$, for any d with $1 \le d \le \nu$. Now suppose that ξ is an arbitrary idle input, let the subnetwork N_0 of $\mathcal{N}$ be Φ_{ξ}, and let N_k be the subnetwork induced by vertices that can only be reached by ξ and $4^{k} - 1$ other inputs ξ' with relat $(\xi, \xi') \le k$. It is clear that N_k has 4^{k} inputs and $64\cdot 4^{\gamma+k}$ outputs. We prove by induction on k that ξ has access to at least $32\cdot 4^{\gamma+k} + 1$ outputs of N_k, thus, in particular, has access to strictly more than half of the outputs of $\mathcal{N}_{\mathcal{L}} = N_{\nu}$. The base case N_0 is obviously true because of our assumption. Consider N_{k+1}. The outputs of N_k are linked to the output of N_{k+1} via four $(32\cdot 4^{\gamma+k}, 33.07\cdot 4^{\gamma+k}, 64\cdot 4^{\gamma+k})$-expanding graphs (with the inlets being the outputs of N_k and the outlets being four disjoint subsets of the outputs of N_{k+1}). By the induction hypothesis, ξ has access to at least $32\cdot 4^{\gamma+k} + 1$ outputs of N_k. These $32\cdot 4^{\gamma+k} + 1$ vertices are joined by edges to at least $4\cdot 33.07\cdot 4^{\gamma+k}$ outputs of N_{k+1} (via the four $(32\cdot 4^{\gamma+k}, 33.07\cdot 4^{\gamma+k}, 64\cdot 4^{\gamma+k})$-expanding graphs). By our assumption, there are at most $0.07\cdot 4^{\gamma+k+1}$ outputs of N_{k+1} are faulty. There are at most $4^{k+1} - 1$ outputs of N_{k+1} that are busy, because each busy output is one-to-one corresponding (via a directed path) to a busy input of N_{k+1}, and there are at most $4^{k+1} - 1$ such inputs. Thus, ξ has access to at least $4 \cdot 33.07 \cdot 4^{\gamma+k} - 0.07 \cdot 4^{\gamma+k+1} - 4^{k+1} + 1 > 32 \cdot 4^{\gamma+k+1} + 1$ outputs of N_{k+1}. This completes our induction. □

COROLLARY 2. *With probability at least* $1 - c_1\nu(144\varepsilon)^{\nu} - \nu(2/e)^{2\nu}$, *the mirror image of* $\mathcal{N}_{\mathcal{R}}$ *is a majority-access network.*

We observe that, if $\mathcal{N}_{\mathcal{L}}$ and the mirror image of $\mathcal{N}_{\mathcal{R}}$ are both majority-access networks and the inputs and outputs of $\mathcal{N}$ are distinct (no two input(s) and output(s) contracting to a single vertex), then $\mathcal{N}$ contains a nonblocking 4^{ν}-network of no-failure edges.

LEMMA 7. *With probability at most* $c_2\nu^2(160\varepsilon)^{2\nu}$, *where* $c_2 = 4^{15}/(1 - 40\varepsilon)$, *there exist two input(s) and output(s) that contract to a single vertex.*

Proof. The correctness of the lemma follows four observations. First, any simple path joining two input(s) and output(s) must contain at least 2ν edges. Second, for any $l \geq 2\nu$, there are at most $(64 \cdot 4^{\gamma})^2(40)^{l-2}$ such paths of length l, since the degree of inputs and outputs is $64 \cdot 4^{\gamma}$ and that of the other vertices is at most 40. Note that $(64 \cdot 4^{\gamma})^2(40)^{l-2} < 4^{14}\nu^2(40)^{l}$, since $4^{\gamma} \leq 136\nu$. Third, the probability that a path of length l gets "shorted" (all edges on the path are in closed failure state) is less than ε^{l}. Last, there are at most $(2 \cdot 4^{\nu})^2$ such input or output pairs. □

THEOREM 2. *Network* $\mathcal{N}$ *is a* $(10^{-6}, \delta)$*-nonblocking n-network with at most* $4^9 n(\log_4 n)^2$ *edges and* $5 \log_4 n$ *depth for arbitrarily small* δ, *when n is sufficiently large.*

Proof. We have seen that network $\mathcal{N}$ contains at most $1408\nu 4^{\nu+\gamma}$ edges and has $4\nu + 1$ depth, where $n = 4^{\nu}$ and $\gamma = \lceil \log_4 34\nu \rceil$. Work out the constant using $4^{\gamma} \leq 136\nu$. The probability that $\mathcal{N}$ fails to contain a nonblocking n-network of no-failure edges is less than $2(c_1\nu(144\varepsilon)^{\nu} - \nu(2/e)^{2\nu}) + c_2\nu^2(160\varepsilon)^{2\nu}$, by Lemma 6, Corollary 2, and Lemma 7. This value can be arbitrarily small when $n = 4^{\nu}$ is sufficiently large, given $\varepsilon = 10^{-6}$. □

Acknowledgment. The authors thank Bruce Maggs for helpful discussions.

REFERENCES

[AHU] A. V. AHO, J. E. HOPCROFT, AND J. D. ULLMAN, *The Design and Analysis of Computer Algorithms*, Addison-Wesley, Reading, MA, 1974.

[ALM] S. ARORA, T. LEIGHTON, AND B. MAGGS, *On-line algorithms for path selection in a nonblocking network*, ACM Sympos. on Theory of Computing, 22 (1990), pp. 149–158.

[B] V. E. BENEŠ, *Optimal rearrangeable multistage connecting networks*, Bell System Tech. J., 43 (1964), pp. 1641–1656.

[BP] L. A. BASSALYGO AND M. S. PINSKER, *Complexity of optimum non-blocking switching network without reconnections*, Problems Inform. Transmission, 9 (1974), pp. 64–66.

[Cl] C. CLOS, *A study of non-blocking networks*, Bell System Tech. J., 32 (1953), pp. 406–424.

[Co] M. I. COLE, *Algorithmic Skeletons: A Structured Approach to the Management of Parallel Computation*, Ph.D. thesis, Computer Science, Univ. of Edinburgh, Oct. 1988.

[CL] G. CHARTRAND AND L. LESNIAK, *Graphs and Digraphs*, 2nd ed., Wadsworth, Belmont, CA, 1986.

[FFP] P. FELDMAN, J. FRIEDMAN, AND N. PIPPENGER, *Wide-sense non-blocking networks*, SIAM J. Discrete Math., 1 (1988), pp. 158–173.

[GG] O. GABBER AND Z. GALIL, *Explicit constructions of linear-sized superconcentrators*, J. Comput. System Sci., 22 (1981), pp. 407–420.

[L] G. LIN, *Edge-disjoint paths in a tree*, in preparation.

[L92] ———, *Fault-tolerant planar communication networks*, ACM Sympos. on Theory of Computing, 24 (1992), pp. 133–139.

[LM] T. LEIGHTON AND B. MAGGS, *Expanders might be practical: Fast algorithms for routing around faults on multibutterflies*, IEEE Sympos. on Foundation of Computer Science, 30 (1989), pp. 384–389.

[LPS] A. LUBOTZKY, R. PHILLIPS, AND P. SARNAK, *Ramanujan graphs*, Combinatorica, 8 (1988), pp. 261–277.

[M] G. A. MARGULIS, *Explicit constructions of concentrators*, Problems Inform. Transmission, 9 (1975), pp. 325–332. (In English.)

[MS] E. F. MOORE AND C. E. SHANNON, *Reliable circuits using less reliable relays*, Part I and Part II, J. Franklin Inst., 262 (1956), pp. 191–208, 281–297.

[PY] N. PIPPENGER AND A. C. YAO, *Rearrangeable networks with limited depth*, SIAM J. Algebraic Discrete Math., 3 (1982), pp. 411–417.

[P82] N. PIPPENGER, *Telephone switching networks*, AMS Proc. Sympos. Appl. Math., 26 (1978), pp. 101–133.

[P90] ———, *Communication networks*, in Handbook of Theoretical Computer Science, Chapter 15, J. van Leeuwen, ed., Elsevier, Amsterdam, 1990.

[S] C. E. SHANNON, *Memory requirements in a telephone exchange*, Bell System Tech. J., 29 (1950), pp. 343–349.

[U] E. UPFAL, *An $O(\log N)$ deterministic packet routing scheme*, ACM Sympos. on Theory of Computing, 21 (1989), pp. 241–250.

[V] L. G. VALIANT, *On nonlinear lower bounds in computational complexity*, ACM Sympos. on Theory of Computing, 7 (1975), pp. 45–53.