Brief Contributions

An $O(\log_2 N)$ Depth Asymptotically Nonblocking Self-Routing Permutation Network

G.A. De Biase, C. Ferrone, and A. Massini

Abstract—A self-routing multi-log$N$ permutation network is presented and studied. This network has $3\log_2 N - 2$ depth and $N(\log_2 N)/2$ nodes, where $N$ is the number of network inputs and outputs very close to 1. A parallel routing algorithm runs in $3\log_2 N - 2$ time on this network. The overall system (network and algorithm) can work in pipeline and it is asymptotically nonblocking in the sense that its blocking probability vanishes when $N$ increases, hence the quasi-totality of the information synchronously arrives in $3\log_2 N$ steps at the network outputs. This network presents very good fault tolerance, a modular architecture, and it is suitable for information exchange in very large scale parallel processors and communication systems.

Index Terms—Permutation networks, self-routing algorithm, blocking probability, stack of banyan networks.

I. INTRODUCTION

To construct a PRAM-like computing machine built by a very large number of processor elements (PEs), the device devoted to the exchange of information between processors and memories is a very critical point. In fact, efficient nonblocking permutation networks (e.g., necessary for the construction of a PRAM EREW machine) become more and more expensive when the number of their inputs (outputs) $N$ increases (by using crossbar networks, which have $O(1)$ depth, topological complexity and cost increase with $O(N^2)$). For this reason various kinds of nonblocking permutation devices, built by means of blocking permutation networks with $N(\log_2 N)$ topological complexity, have been extensively studied.

Among blocking permutation networks, multistage banyan networks are attractive for their moderate depth which guarantees that information reaches its destination in $\log_2 N$ steps, for the possibility to route information in pipeline by simple self-routing algorithms, and for their topological complexity which is $N\log_2 N/2$ nodes. They are considered a cost-effective alternative to crossbar networks for large $N$, but the fact that multistage banyan networks are blocking, compromises the above mentioned advantages when connection requests are simultaneously presented at all network inputs.

In some recent papers Lea [6], [7], Shyy and Lea [8], [10], and Melen [9] point out that vertical stacks of $K$ banyan networks can be nonblocking or strictly nonblocking permutation networks under suitable conditions. These stacks (multi-log$N$ networks) maintain two important features of banyan networks:

1) the $\log_2 N$ or $O(\log_2 N)$ depth between inlet-outlet pairs, and
2) the self-routing capability.

Unfortunately these studies on multi-log$N$ networks do not provide routing algorithms with a time complexity comparable to the network depth.

A way to build routing algorithms on permutation networks is the probabilistic one [11], [12], [2], but, with this approach, some information cannot reach its destination, and therefore the probabilistic approach becomes valid if the amount of blocked information is negligible. Using this approach, in a recent work a quasi-nonblocking self-routing interconnection network with $\log_2 N$ depth was introduced [2]. In the present work a network, which is more efficient for very large $N$, is presented. It is a $3\log_2 N - 2$ depth multi-log$N$ network on which a probabilistic self-routing algorithm acts. The overall system (network and algorithm) is asymptotically nonblocking in the sense that its blocking probability $pb_N$ vanishes when $N$ increases.

II. ASYMPTOTICALLY NONBLOCKING NETWORKS

Let $B_N$ be a permutation network of size $N$ (with $N$ inputs and $N$ outputs) and let $I = \{i_n\}$ and $O = \{o_n\}$, $n = 1, ..., N$, be the sets of its inputs and outputs respectively. $B_N$ realizes $N$ simultaneous one-to-one connections between each $i_n$ and each $o_n$. In other words, $B_N$ performs a bijection $I \leftrightarrow O$. The one-to-one mappings of $I$ onto $O$ are characterized by the set of all permutations $P = \{p_k\}$, $j = 1, ..., N!$, of the elements of $I$ onto $O$. In nonblocking permutation networks all connection requests presented at the inputs $i_n$ reach their destinations. If $B_N$ is a blocking network, a certain number of requests cannot be honored. The ratio $pb_N = (r_{in} - r_{out})/N^2$, where $r_{in}$ is the number of simultaneous connection requests (input) and $r_{out}$ is the number of nonblocked requests (outputs), is the blocking probability of $B_N$ [11], [12], [2] and represents the probability that a request at the generic input $i_n$ cannot reach its destination $o_n$ when $N$ requests are simultaneously applied on the whole input set $I$. $pb_N$ can depend on $N$ (as an example, in banyan multistage networks $pb_N$ increases when $N$ increases [11], [12]). The quantity $\eta_N = 1 - pb_N$ is the probability that a request at an input $i_n$ reaches its destination $o_n$ when $N$ requests are simultaneously applied on the whole input set, and it will be called efficiency (efficiency is a measure of the nonblocking capability of a network, and nonblocking networks have $\eta_N = 1$ for any $N$). In banyan multistage networks $\eta_N$ decreases when $N$ increases [11], [12].

DEFINITION 1. A blocking $B_N$ with efficiency $\eta_N$ is called asymptotically nonblocking if:

$$\lim_{N \to \infty} \eta_N = 1 \quad (1)$$

It is evident that the above defined interconnection structures have great importance in very massive systems (multiprocessors or communication systems).

Let $S_N = \{B_{N_1}\}$, $k = 1, ..., K$, be a set of $K$ identical and independent permutation networks $B_{N_1}$ of size $N_1$, the inputs and the outputs of $S_N$ belong to the set $I = \{i_n\}$, and $O = \{o_n\}$, $n = 1, ..., N_1; k = 1, ..., K$, respectively. $K$ permutations can act simultaneously on the set $S_N$ and $K$ one-to-one mappings $I \leftrightarrow O$ can be simultaneously performed (each mapping on each $B_{N_1}$), $P = \{p_k\}$, $k = 1, ..., K (P'' \subset P)$, is the set of $K$ uncorrelated permutations $p_k$, belonging to the set $P$, which are simultaneously applied on $B_{N_1}$ networks. If $B_{N_1}$ are blocking networks (with blocking probability $pb_{N_1}$), the overall blocking probability $pb_{N_1}$ of the whole set $S_N$ can be defined as:

DEFINITION 2. The overall blocking probability $pb_{N_1}$ of a set $S_N$ is the probability that, if $K$ requests are simultaneously presented each at an input $i_n$ of one $B_{N_1}$ network, no connection request reaches its destination (when $N \times K$ requests are simultaneously

0018-9340/95$04.00 © 1995 IEEE
applied at the whole input set \( I \) of \( S_N \).

If connection requests at the inputs of each \( B_{N_k} \) are completely independent (uncorrelated permutations \( p_k \) simultaneously act on each \( B_{N_k} \)), \( p_{b_k}^N \) is given by (see e.g., [4]):

\[
p_{b_k}^N = p_b^N
\]  

If \( 0 < p_b^N < 1 \), one can write:

\[
\lim_{k \to -\infty} p_{b_k}^N = 0
\]  

From (2) and (3) there follows that:

\[
\lim_{k \to -\infty} (1 - p_{b_k}^N) = \lim_{k \to -\infty} \eta_k^N = 1
\]  

where \( \eta_k^N \) is the overall efficiency of the set \( S_N \).

**DEFINITION 3.** The overall efficiency \( \eta_k^N \) of the set \( S_N \) is the probability that, if \( K \) requests are simultaneously presented each one at an input \( i \) of one \( B_{N_k} \) network, at least one connection request reaches its destination (when \( N \times K \) requests are simultaneously applied at the whole input set \( I \) of \( S_N \)).

Equation (4) shows that for any size \( N \), when \( K \) (the number of \( B_{N_k} \) networks) goes to \( \infty \), the overall efficiency \( \eta_k^N \) of the set \( S_N \) goes to 1 under the sole condition that \( 0 < p_b^N < 1 \). According to Definition 3, the condition expressed by (1), which states that a permutation network is asymptotically nonblocking when the size \( N \) increases, can be generalized for a set \( S_N \):

\[
\lim_{N \to -\infty} \eta_k^N = 1
\]

**THEOREM 1.** Let \( S_N \) be a set of identical and independent blocking networks \( B_{N_k} \) with blocking probability \( p_b^N \) and let \( K_N \) depending on \( N \), be the number of \( B_{N_k} \) networks of the set \( S_N \). The set \( S_N \) is asymptotically nonblocking if all permutations \( p_k \) presented at \( B_{N_k} \) networks are uncorrelated, and if:

\[
K_N = \left\lceil \frac{C(N)}{\ln p_b^N} \right\rceil
\]

where \( 0 < p_b^N < 1 \) for any \( N \), and \( C(N) \) is any function for which:

\[
\lim_{N \to -\infty} C(N) = +\infty
\]

**PROOF.** The overall efficiency of a set \( S_N \) is \( \eta_k^N = 1 - p_{b_k}^N \), and \( \lim_{N \to -\infty} \eta_k^N = 1 \) when \( \lim_{N \to -\infty} p_{b_k}^N = 0 \). There follows from (2) that:

\[
\lim_{N \to -\infty} p_{b_k}^N = \lim_{N \to -\infty} p_{b_k}^{K_N} = 0
\]

which is true if:

\[
\lim_{N \to -\infty} K_N \ln p_b^N = +\infty
\]

It is easy to see that, with the substitution \( K_N = -C(N)/\ln p_b^N \), (7) is always verified if \( \lim_{N \to -\infty} C(N) = +\infty \). The ceiling in (5) is necessary to guarantee integer \( K_N \) values. \( \square \)

Theorem 1 states that a set of blocking networks \( S_N \) is asymptotically nonblocking if \( K_N \) increases according to (5).

**III. AN ASYMPTOTICALLY NONBLOCKING SELF-Routing PERMUTATION DEVICE**

Based on the results given in Section II, a new self-routing asymptotically nonblocking permutation device, based on stacks of \( K_N \) blocking networks, can be carried out if the following assumptions hold:

- the permutation \( p_i \) (input of the whole device) is transformed into a set \( P' \) of \( K_N \) uncorrelated permutations \( p_k \),
- each permutation \( p_j \) of the set \( P' \) acts independently on one \( B_{N_k} \) network of the set \( S_N \).

Finally, every output of the device is obtained by the logical union of the corresponding outputs \( o_k \) of all \( B_{N_k} \) networks (see Definition 3).
uncorrelated permutations $p_k$, while the second one (router) addresses connection requests towards their destinations. The output ports $o_i$ of the whole device are obtained by the logical OR of the corresponding output ports of all $B_{N_i}$ (see Fig. 1b). To easily obtain self-routing capability, this permutation device is built by stacks, the planes of which are butterfly networks, as shown in Fig. 2. As one can see, in all planes the output nodes of a network are in common with the input nodes of the subsequent network, for this reason the device depth is $3 \log_2 N - 2$ stages.

The behavior of the device has been examined by numerical simulation under the assumption of permutation requests [11]. Simulations give the values of the device efficiency $\eta_N$, when a suitable function $f(N)$ is chosen. The numerical program simulates the behavior of all stacks. For each $N$ it utilizes, as input of the whole device, a randomly chosen permutation $p_i$. The randomization of requests is obtained only by setting all the nodes of each plane of the randomizer on randomly chosen states. The routing of requests is obtained on the routing stack by the simple distributed algorithm presented in Section II.B. Because the rapid increase with $N$ of the number of permutations $p_i$ ($N!$) generates very large computation times, to obtain the values of $\eta_N$ a number of attempts, sufficient to reach at least the 99% confidence level, has been executed in the interval $2^3 \leq N \leq 2^{13}$.

Simulated values of the device efficiency $\eta_N^2$ compared with $\eta_N$ computed values, when $K_N = \log_2 N$ and $K_B = \log_2 N - 1$, are presented in Fig. 3b. The blocking probability of the component banyan networks are computed by (8), and the efficiency of the whole device by (2). The model of Szymansky and Hamacher generates slightly overestimated $\eta_B$ values (see Fig. 3 in [11] and Fig. 4 in [12]), and consequently the simulated efficiency values $\eta_N'$ of the permutation device are slightly greater than the computed ones, $\eta_N$, (about 0.25% in the considered interval of $N$).

C. Number of Planes

Equation 5 gives the number of planes $K_B$ of the routing stack and, consequently, of the randomizer. The values of efficiency $\eta_N$, for $N = 2^1, 2^2, 2^3, \ldots$ of a banyan multistage network, under permutation requests, are recursively given with good accuracy by the model presented in [12] by Szymansky and Hamacher:

$$\eta_{N+1} = \sum_{i=1}^{2} \binom{2}{i} \eta_i \left(1 - \eta_i\right)^{2-i} \cdot \sum_{i=1}^{2} \binom{2}{i} \left(\frac{2^{H+i}}{2^{H+i-1}}\right)^{i-t} \cdot \left(\frac{2^{H+i}+1}{2^{H+i}}\right)^{i-t}$$

starting from: $\eta_1 = \frac{2^{H+1}}{2^{2^{H+1}-1}}$. In (8) $h = 1, \ldots, H - 1$, and $H = \log_2 N$ is the number of network stages. This model can be used to compute, by (2), the blocking probability of the router planes, in fact $p_B$ values can be obtained by (8) by the substitution $p_B = 1 - \eta_N$.

When in an interval $(N_m, N_s)$, $K_B$ values are given by a function $f(N)$, $K_B$ values are computed by $f(N) \geq -C(N)/\ln p_B$, Theorem 1 guarantees that the efficiency $\eta_N$ of the set $S_N$ increases with $N$ for all values of $N$ belonging to the same interval. Among these functions, $K_B = \log_2^2 N$ is a good compromise between network complexity and efficiency increase. In this case the topological complexity of the presented network is $N! \log_2^2 N (3 \log_2 V - 2)/2$ nodes and, also using $\gamma$ values very close to 1 ($\gamma = 1.05, 1.10, \ldots$), the network efficiency quickly increases in a wide interval of $N$, as shown in Fig. 3a.

IV. SIMULATIONS

The behavior of the device has been examined by numerical simulation under the assumption of permutation requests [11]. Simulations give the values of the device efficiency $\eta_N$ versus $N$ when a suitable function $K_N = f(N)$ is chosen. The numerical program simulates the behavior of all stacks. For each $N$ it utilizes, as input of the whole device, a randomly chosen permutation $p_i$. The randomization of requests is obtained only by setting all the nodes of each plane of the randomizer on randomly chosen states. The routing of requests is obtained on the routing stack by the simple distributed algorithm presented in Section III.B. Because the rapid increase with $N$ of the number of permutations $p_i$ ($N!$) generates very large computation times, to obtain the values of $\eta_N$ a number of attempts, sufficient to reach at least the 99% confidence level, has been executed in the interval $2^3 \leq N \leq 2^{13}$.

Simulated values of the device efficiency $\eta_N^2$ compared with $\eta_N$ computed values, when $K_N = \log_2 N$ and $K_B = \log_2 N - 1$, are presented in Fig. 3b. The blocking probability of the component banyan networks are computed by (8), and the efficiency of the whole device by (2). The model of Szymansky and Hamacher generates slightly overestimated $\eta_B$ values (see Fig. 3 in [11] and Fig. 4 in [12]), and consequently the simulated efficiency values $\eta_N'$ of the permutation device are slightly greater than the computed ones, $\eta_N$, (about 0.25% in the considered interval of $N$).
V. CONCLUDING REMARKS

The architecture presented is inherently fault tolerant, in fact it consists of three vertical stacks each of $K_N$ banyan networks which implement many physical paths for each logical path. Faults on every network of each stack act on the overall performance of the structure. In absence of faults the overall efficiency of $M$ cascaded stacks $S_{ij}$ is given by: $\eta_{ij}^* = \prod_{j=1}^{M} \eta_N^{s_{ij}} = \prod_{j=1}^{M} \eta_N^{b_{ij}}$. In the presented device $\eta_{ij}^* = \eta_N^{s_{ij}}$, because $\eta_N^{s_{ij}} = \eta_N^{b_{ij}} = 1$ in the randomizer, while $\eta_N^{b_{ij}}$ is given by (2) and (8).

If in presence of faults the efficiency of each component network decreases by a quantity $\Delta \eta_{ij}$, the degradation of the efficiency of the whole device $\Delta \eta_N$ can be computed by:

$$\eta_N - \Delta \eta_N = \prod_{j=1}^{3} \left( \eta_N - \Delta \eta_N \right)^{K_N}$$

(9)

Now, by using (9) the degradation of the efficiency of the whole permutation device can be examined versus the efficiency degradation of component networks. In Fig. 4, $\Delta \eta_N$ values versus the size $N$, for several $\gamma$ values, are plotted. A small (10%) and a strong (40%) value for the efficiency degradation of all component networks is considered. The efficiency values of component banyan networks, in the absence of faults, are assumed $\eta_N = 1$ for the randomizer, while $\eta_N$ values are computed by (8) for the router. As one can see from the figures, also with $\gamma$ values very close to 1, fault tolerance quickly increases with $N$.

The multistage structure of the permutation device, the distributed self-routing algorithm, and the forward direction of the information flux, guarantee that the information wavefront synchronously passes
through the network stages in $3\log_2 N - 2$ steps. Then the system can work in pipeline, and information can be presented at the device inputs at each time interval $T$, where $T$ is the stage-to-stage propagation time.

The behavior of the efficiency of the device has been examined under the assumption of permutation request patterns. The efficiency values, closer and closer to 1 for large $N$, guarantee that, also under the assumption of random request patterns [12], the behavior of the permutation device is very close to that of nonblocking networks [2].

In the presented permutation device the two most important features of banyan networks are maintained: moderate depth ($3\log_2 N - 2$ stages) and simple request routing (obtainable by a self-routing distributed algorithm, which permits pipelined operations). This permutation network becomes asymptotically nonblocking when a suitable increase with $N$ of the number of planes, $K_p$, is chosen (see Theorem 1). When the planes are banyan networks, this behavior is already possible with $K_p = \log_2 N$. In this case the topological complexity (number of nodes) is $O(\log_2^2 N)$, which is the same as the Koppelman-Oruç and Batcher networks [5], [11], that are nonblocking but have a worse depth ($O(\log_2^2 N)$ instead of $O(\log_2 N)$). With a very little increase in topological complexity, the desired efficiency of this permutation device can be quickly reached for any $N$. The behavior of efficiency and fault tolerance clearly highlights that this network becomes more and more advantageous as device size increases.

REFERENCES