Available online at www.sciencedirect.com

SCIENCE DIRECT®

J. Parallel Distrib. Comput. 65 (2005) 1443 – 1452

Journal of Parallel and Distributed Computing

www.elsevier.com/locate/jpdc

Swapped interconnection networks: Topological, performance, and robustness attributes

Behrooz Parhami*

Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560, USA

Available online 23 June 2005

Abstract

Interconnection architectures range from complete networks, that have a diameter of $D = 1$ but are impractical except when the number n of nodes is small, to low-cost, minimally connected ring networks whose diameter $D = \lfloor n/2 \rfloor$ is unacceptable for large n. In this paper, our focus is on swapped interconnection networks that allow systematic construction of large, scalable, modular, and robust parallel architectures, while maintaining many desirable attributes of the underlying basis network comprising its clusters. A two-level swapped network with n^2 nodes is built of n copies of an n-node basis network using a simple rule for intercluster connectivity (node j in cluster i connected to node i in cluster j) that ensures its regularity, modularity, packageability, fault tolerance, and algorithmic efficiency. We show how key parameters of a swapped interconnection network are related to the corresponding parameters of its basis network and discuss implications of these results to synthesizing large networks with desirable topological, performance, and robustness attributes. In particular, we prove that a swapped network is Hamiltonian (respectively, Hamiltonian-connected) if its basis network is Hamiltonian (Hamiltonian-connected). These general results supersede a number of published results for specific basis networks and obviate the need for proving Hamiltonicity or Hamiltonian connectivity for many other basis networks of practical interest.

Keywords: Average internode distance; Bisection width; Fault diameter; Fault tolerance; Interconnection network; Modularity; OTIS network; Packageability; Robustness; Survivability

1. Introduction

Low-latency, high-communication bandwidth, modularity, packageability, energy efficiency, and robustness are some of the properties that are sought in networks for parallel and distributed computing [16]. Given that network performance parameters depend not only on its architecture but also on a number of factors pertaining to applications and their data-exchange characteristics, the challenge in interconnection network design is finding the appropriate match between communication needs of applications on one side and capabilities and limitations inherent in each architecture on the other. This, in turn, explains the proliferation of implemented and proposed connectivity schemes among multiple processors, sometimes characterized as the sea of interconnection networks [13,14,21].

Ideally, each network node is directly connected to every other node, thus allowing one-hop communication between any pair of nodes. This connectivity pattern is modeled by the n-node *complete graph* K_n. Physically, however, complete-graph connectivity is difficult to provide for large systems that are of practical interest. At the other extreme from K_n, the simplest possible physical connectivity pattern is that of n-node *ring* R_n. Here, each node has only two communication channels. Data exchange is direct only between each node and one or two neighbor(s) and indirect in all other cases. Intermediate architectures between K_n and R_n can be obtained in a variety of ways, providing tradeoffs in cost and performance. Network cost is affected, among other things, by the (maximum) node degree d, while indicators of network performance include *diameter* D and

* Fax: +1 805 893 3262.
E-mail address: parhami@ece.ucsb.edu.

0743-7315/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jpdc.2005.05.002

bisection width B. The degree-diameter product *dD* is sometimes used as a composite measure of cost-performance, or indicator of cost-effectiveness, in interconnection network comparisons.

Many of these intermediate architectures can be viewed as *chordal rings* [1], rings to which *bypass links* or *chords* have been added to reduce the network diameter, or richly connected graphs from which certain links are systematically removed via *pruning* [9,10,17] so as to reduce the node degree, wiring density, and network cost. Alternate mechanisms for deriving cost-effective interconnection networks from other networks include cross-product composition, recursive substitution, and hierarchical composition. These combining strategies lead to families of networks that are all based on the same component networks, thus sharing a number of topological, performance, and robustness attributes.

Our focus of discussion in this paper is swapped interconnection networks that allow systematic construction of large, scalable, and highly modular parallel architectures, while maintaining desirable topological properties of an underlying component or basis network. We show how key parameters of a swapped interconnection network are related to the corresponding attributes of its basis network and discuss the implications of these results to synthesizing large networks with desirable packageability, performance, and fault-tolerance properties. Studying general composite structures such as swapped networks is important in that it allows the derivation of results that pertain to a wide array of interconnection networks. For example, our proof of Hamiltonicity for a swapped network with a Hamiltonian basis network leads to Hamiltonicity results for a wide array of interconnection networks. This is clearly preferable to proving each network to be Hamiltonian in a separate study. The latter has been the norm in parallel processing research. For example, the recent result of Fu and Chen [5] follow from our Theorems 6 and 7.

After defining swapped networks and tracing their history and relevant related work in Section 2, we present results about their topological parameters in Section 3. We then discuss certain structural and fault-tolerance properties of swapped networks in Sections 4 and 5, respectively. Section 6 contains our conclusions and points to several directions for further research.

2. Definitions and background

Symmetric interconnection networks, in which node degree is uniformly equal to d and the network "looks the same from every node," are of particular interest due to their algorithmic simplicity and greater resilience resulting from the absence of weak spots. A symmetric interconnection network is characterized by its number n of nodes, along with one or more other parameters or rules that define its connectivity pattern. Focusing on the number n of nodes for the moment (e.g., by fixing other independent variables at default values), topological parameters can be expressed as conventional or asymptotic functions of n. This is shown in the case of network diameter in Fig. 1.

Practical interconnection networks used in parallel computers fall on the right-hand side of Fig. 1, where networks tend to be scalable, readily packageable, and low-cost. Interconnection network research over the past two decades, on the other hand, has focused on lower-diameter networks constituting the left half of Fig. 1. Swapped networks, defined below, and illustrated in Fig. 2, offer some of the advantages of each group in that they can have sublogarithmic diameters while remaining both packageable and relatively scalable.

Definition 1. *Swapped network*—the swapped network $Sw(G)$, derived from the n-node *nucleus* or *basis* graph G, is a graph with n copies of G (*clusters*) numbered 0 to $n-1$, so that node j in cluster i is connected to node i of cluster j for all $i \neq j$ and $0 \leqslant i, j \leqslant n-1$ [26].

Algorithm $R_{Sw(G)}$. Routing in a swapped network $Sw(G)$ from v_{ij} to v_{kl} via at most one intercluster hop, using the routing algorithm R_G for G [26]: If $i = k$, then use R_G to route from v_{ij} to v_{il} within cluster i in $\delta_G(j,l)$ hops. Otherwise, first route from v_{ij} to v_{ik} within the source cluster i in $\delta_G(j,k)$ hops, then from v_{ik} to v_{ki} in a single intercluster hop, and finally from v_{ki} to v_{kl} within the destination cluster k in $\delta_G(i,l)$ hops, as depicted in Fig. 3.

Note that $R_{Sw(G)}$ is not a shortest-path routing algorithm. Referring to Fig. 3, we note that the $R_{Sw(G)}$ path $v_{ij} \to v_{ik} \to^{Sw} v_{ki} \to v_{kl}$, which is of length $\delta_G(j,k)+\delta_G(i,l)+1$, may be longer than the alternate path $v_{ij} \to v_{im} \to^{Sw} v_{mi} \to v_{mk} \to^{Sw} v_{km} \to v_{kl}$, via the intermediate cluster m, which is of length $\delta_G(j,m)+\delta_G(i,k)+\delta_G(m,l)+2$.

Example 1. *Routing in a torus-based swap network*: The alternate path of Fig. 3 is shorter, for example, if the node pairs j and m, i and k, and m and l are neighbors in G, leading to the alternate path of Fig. 3 being of length 5, whereas $\delta_G(j,k)+\delta_G(i,l) > 4$. This is the case, for instance, when G is a 4×4 torus network, with nodes i, j, k, l, and m situated as in Fig. 4.

Despite the foregoing observation, the fact that $R_{Sw(G)}$ uses at most one intercluster link makes it preferable in a modular system with hierarchical packaging where intracluster communication is significantly faster than intercluster data exchange.

Definition 1 can be extended and modified in a variety of ways. For example, swapped networks can be built up recursively beginning with a fairly small basis network. The network size increases from n, to n^2, to n^4, and so on. It is fairly easy to prove that when so constructed from a very small seed, the node degree of a recursive swapped network

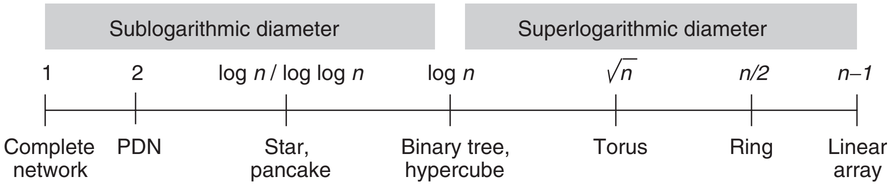

Fig. 1. The spectrum of interconnection networks in terms of diameter for size n.

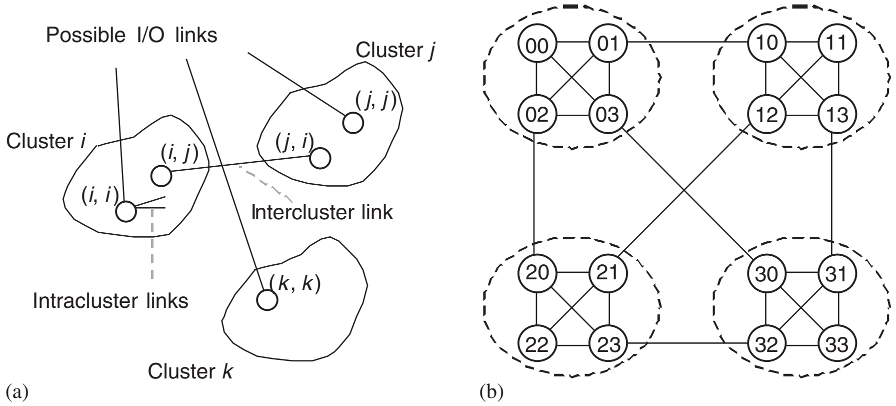

Fig. 2. The general structure of a swapped network and an example network with the 4-node complete graph as its basis.

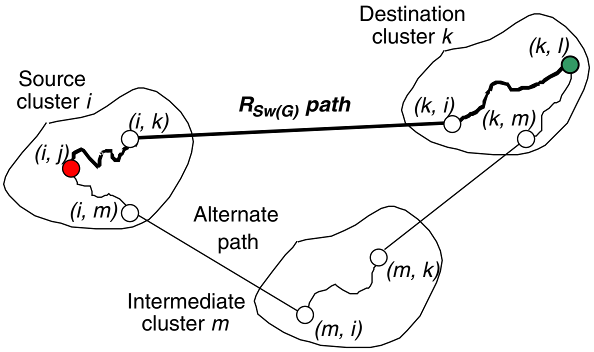

Fig. 3. The path prescribed by $R_{Sw(G)}$ and an alternate path from v_{ij} to v_{kl}.

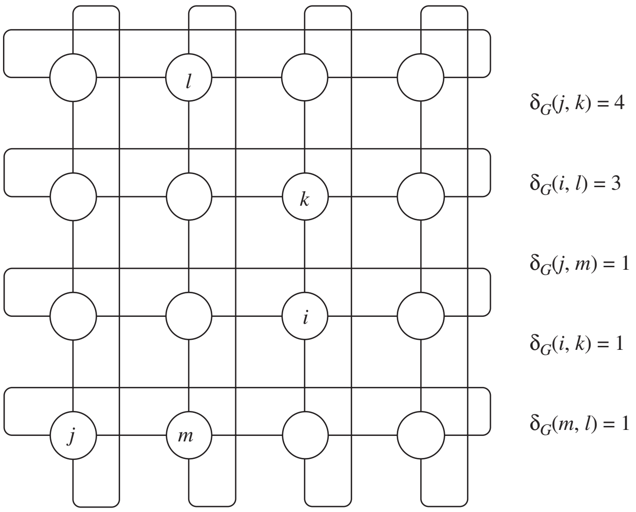

Fig. 4. Torus cluster G of a 256-node swapped network, with five of its nodes labeled.

is a double-logarithmic function of its size. In the remainder of this paper, we will focus exclusively on two-level networks, with the understanding that many of the results can be extended to higher levels in a straightforward manner. For example, if Hamiltonicity of the n-node network G implies the Hamiltonicity of the n^2-node network $Sw(G)$, the result extends to higher-level networks with $n^4, n^8, n^{16}, \ldots$ nodes by induction. Obviously, given the fast growth in the network size, the number of recursive levels is quite limited in practice.

Focusing on the two-level swapped network, it is natural to try to come up with some use for the empty ports due to lack of intercluster links at nodes $v_{ii}, 0 \leqslant i \leqslant n-1$, for the sake of uniformity in node degree. It would have been nice to connect these n nodes into a ring, but that would require two extra links per node. Two ways of using these links suggest themselves.

Definition 2. *Folded swapped network*: If we connect each node v_{ii} of $Sw(G), 0 \leqslant i \leqslant n-1$, to node $v_{n-1-i,n-1-i}$, where n is even, we obtain a *folded swapped network*. This modification is inspired by folded hypercube in which nodes i and $2^q - 1 - i$ are connected via a diagonal link.

Definition 3. *Expanded swapped network*: If we add a new cluster n to $Sw(G)$ whose original clusters are numbered 0 to $n-1$, connecting node v_{ii} in the original cluster i to node i in the new cluster n for $0 \leqslant i \leqslant n-1$, we obtain an *expanded swapped network*. When built from an n-node basis network G, an expanded swapped network has $n(n+1)$ nodes.

In both modified definitions, the additional links provide alternate paths between nodes, thus potentially improving performance and robustness. Because folded and expanded swapped networks are more richly connected than $Sw(G)$, many of the results obtained for the latter apply to the former networks as well. In this paper, we do not study these two variants separately.

Before proceeding further, it is instructive to trace the history of swapped networks and related works in the literature. It was recently pointed out to the author by an anonymous reviewer that swapped networks are the same as optical transpose interconnection system (OTIS) architectures which have been extensively studied by other researchers. Tracing the history of OTIS, the author discovered that its roots go back to 1993, when Marsden et al. published a three-page note in *Optics Letters* [11] suggesting a topology in which nodes (i, j) and (j, i) are linked via an optical channel. It appears that transfer of the OTIS idea to the computer architecture and parallel processing community occurred, in part, due the 1998 Ph.D. dissertation of C.-F. Wang at the University of Florida, under Professor Sartaj Sahni, and publication of its results beginning in 1997 [12,20,22–24]. Architectural and some topological considerations for OTIS networks have been studied by Zane et al. [30] and Day and Al-Ayyoub [2], among others.

Concurrent with the developments just cited, and before any reference to OTIS appeared in the computer architecture or parallel processing literature, Chi-Hsiang Yeh (a former doctoral student of the author) proposed swapped networks [26–28] as tools for unifying and extending a number of known hierarchical networks. Prominent among these prior architectures were the two-level special case of WK-recursive networks [3,4] (beyond two levels, WK-recursive and swapped networks diverge in structure and do not have much in common), hierarchical cubic networks [6], and recursively fully connected networks [25]. The unification was due to the replacement of complete-graph or hypercube component networks of the prior architectures with an arbitrary graph.

It thus appears that OTIS and swapped architectures have evolved independently and in parallel. Incidentally, “swapped network” is a much more descriptive, as well as more appropriate, name than OTIS because many of the algorithms and topological properties discussed in the literature, and here, are independent of implementation technology. Reference to optical implementation may in fact serve to discourage researchers not involved in optical computing or communications from examining the published results, which may be useful in other contexts.

3. Topological parameters

In this section, we relate some of the topological parameters of a swapped network $Sw(G)$ to the respective parameters of the basis network G.

Theorem 1. *Degree and diameter*: *If G has node degree d and diameter D, the degree and diameter of $Sw(G)$ are $d+1$ and $2D+1$, respectively* [26].

An implication of Theorem 1 is that if the basis network G has logarithmic or sublogarithmic diameter, then so does the resulting swapped network $Sw(G)$. This is because $D = O(\log n)$ implies $2D + 1 = O(\log n^2)$ and $D = o(\log n)$ implies $2D + 1 = o(\log n^2)$.

Theorem 2. *Average internode distance*: *If the n-node graph G is node-symmetric and has an average internode distance Δ, then $Sw(G)$ has an average internode distance $\Delta_{Sw(G)}$ satisfying $\Delta_{Sw(G)} < 2\Delta + 1 - (\Delta + 1)/n$, where the right-hand side expression is the average internode distance with respect to the routing algorithm $R_{Sw(G)}$.*

Proof. Let v_{ij} denote node j in the ith cluster G and $\delta(v, v')$ denote the distance between nodes v and v' when the routing algorithm $R_{Sw(G)}$ is used. Then, the average internode distance $\Delta_{R[Sw(G)]}$ of $Sw(G)$ with respect to $R_{Sw(G)}$ is derived as follows:

$$
\begin{aligned}
n^4\Delta_{R[Sw(G)]} &= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\delta(v_{ij}, v_{kl}) \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{l=0}^{n-1}\delta(v_{ij}, v_{il}) \\
&\quad + \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{k=0(k\neq i)}^{n-1}\sum_{l=0}^{n-1}[\delta(v_{ij}, v_{ik}) \\
&\quad + 1 + \delta(v_{ki}, v_{kl})] \\
&= n^3\Delta + n\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{k=0(k\neq i)}^{n-1}\delta(v_{ij}, v_{ik}) \\
&\quad + n^3(n-1) + n\sum_{i=0}^{n-1}\sum_{k=0(k\neq i)}^{n-1}\sum_{l=0}^{n-1}\delta(v_{ki}, v_{kl}) \\
&= n^4 + n^3(\Delta - 1) + n^2\sum_{j=0}^{n-1}\sum_{k=1}^{n-1}\delta(v_{0j}, v_{0k}) \\
&\quad + n(n-1)\sum_{i=0}^{n-1}\sum_{l=1}^{n-1}\delta(v_{0i}, v_{0l}) \\
&= n^4 + n^3(\Delta - 1) + n^2\sum_{j=0}^{n-1}\sum_{k=1}^{n-1} \\
&\quad \delta(v_{0j}, v_{0k}) + n^3(n-1)\Delta \\
&= n^4(\Delta + 1) - n^3 + n^2\left[\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\delta(v_{0j}, v_{0k})\right. \\
&\quad \left. - \sum_{j=0}^{n-1}\delta(v_{0j}, v_{00})\right]
\end{aligned}
$$

$$= n^4(\Delta + 1) - n^3 + n^2[n^2\Delta - n\Delta]$$
$$= n^4(2\Delta + 1) - n^3(\Delta + 1).$$

Note that we have made use of the node-symmetry of G by replacing certain sums ranging over n values of an index by n times the summand for a specific index value. Dividing both sides by n^4, and noting that the true average internode distance is less than that obtained when paths are based on $R_{Sw(G)}$, yields the desired result. □

To address the bisection width of swapped networks, we need the following definition.

Definition 4. *β-cut and bisection width*: The *β-cut*, $\beta < 1$, of an n-node network is the minimum number of links that must be removed to partition the network into two disconnected sections of sizes $n\beta$ and $n(1-\beta)$. In particular, the *bisection width* of a network is its 1/2-cut.

Theorem 3. *Bisection width*: *The bisection width $B_{Sw(G)}$ of the swapped network whose basis network G has bisection width B and* $(1/\sqrt{2})$-cut *B' is upper bounded by* $n \times \min(n/4, B, B'/\sqrt{2})$.

Proof. An upper bound u for the bisection width of a network can be established by showing a particular bisection cut of size u. Let n be even. A bisection cut placing clusters 0 through $n/2 - 1$ on one side, and clusters $n/2$ through $n-1$ on the other, consists entirely of swap links and is thus of size $n^2/4$. Bisecting each cluster, so that nodes 0 through $n/2 - 1$ are on one side and nodes $n/2$ through $n - 1$ on the other (renumbering nodes and clusters, if necessary, to accomplish this with a minimum-size cut), leads to the second bound nB. Note that in the latter case, all swap links are confined to the same side of the bisected network. Finally, the upper bound $nB'/\sqrt{2}$, inspired by that derived for OTIS mesh in [12], is obtained as follows. Consider $(1/\sqrt{2})$-cuts in the first $n/\sqrt{2}$ clusters in such a way that nodes with smaller indices are in the larger segment (of size $n/\sqrt{2}$) in each of these partitioned clusters. As in the previous case, nodes and clusters can be renumbered, if necessary, to accomplish this with a minimum-size $(1/\sqrt{2})$-cut. The proof is complete upon noting that nodes 0 to $n/\sqrt{2} - 1$ in the first $n/\sqrt{2}$ clusters, constituting half of the network's n^2 nodes, are not connected to the nodes on the other side by means of intercluster or swap links. Obviously, for any network (such as mesh or torus) whose $(1/\sqrt{2})$-cut is no greater than its bisection width, the latter bound is tighter than nB, but this is not always the case.

For n odd, the first bound becomes $(n-1)(n+1)/4$, which is strictly less than $n^2/4$. To derive the second bound, we bisect clusters 1 through $n-1$, putting $(n-1)/2$ nodes with smaller indices on side 1 and the rest on side 2. Based on this partitioning, there will be $(n-1)^2/2 = n^2/2 - n + 1/2$ nodes on side 1 and $(n-1)(n+1)/2 = n^2/2 - 1/2$ nodes on side 2. Putting the entire cluster 0 on side 1 will balance the two sides. Because all swap links of cluster 0 are confined to side 1, the width of the resulting cut is $(n-1)B$, which is strictly less than nB. The third bound is independent of the parity of n. If, with either parity, $n/\sqrt{2}$ is not an integer, we use $\lfloor n/\sqrt{2} \rfloor$, along with a method similar to that of the case just discussed to balance the two sides. □

We conjecture that the bound in Theorem 3 is tight for a wide class of symmetric basis networks, but have been unable to prove this, or to derive a formula for the exact bisection width. The upper bound, meanwhile, provides assurance that the bisection width, and thus wiring and associated area costs, can be kept in check by using a seed network of small bisection B.

4. Structural properties

In this section, we cover some structural properties of swapped networks. These results shed light on the relationships between different classes of swapped networks and facilitate not only systematic studies of such networks, but also comparison with competing networks.

Theorem 4. *Hierarchical structure*: *If H is a subgraph of G, then $Sw(H)$ is a subgraph of $Sw(G)$.*

Proof. Number the nodes of G from 0 through $n - 1$. Let the nodes of H be $i_1, i_2, \ldots, i_k$. Then, the subgraphs H in clusters $i_1, i_2, \ldots, i_k$ of $Sw(G)$, along with their swap links, form $Sw(H)$. □

Theorem 4 can be easily extended to the case of r disjoint subgraphs $H_1, H_2, \ldots, H_r$, leading to the associated r networks $Sw(H_i)$ forming disjoint subgraphs of $Sw(G)$. This is a generalized version of Theorem 1 in [2] which pertains to the case of all the H_i being identical subgraphs.

Hamiltonicity and Hamiltonian connectivity are useful properties of interconnection networks. These are defined as follows.

Definition 5. *Hamiltonian path and Hamiltonian cycle*: A *Hamiltonian path* between nodes u and v in graph G is a path that leads from u to v and visits each node of G exactly once. A *Hamiltonian cycle* in G is a Hamiltonian path from one node to itself.

Definition 6. *Hamiltonicity and Hamiltonian connectivity*: A graph G is *Hamiltonian* or *cyclic* if it contains a Hamiltonian cycle. It is *Hamiltonian-connected* if a Hamiltonian path exists between every pair of nodes in G.

Theorem 5. *Hamiltonian connectivity*: *If G is Hamiltonian-connected, then so is $Sw(G)$. Also, Hamiltonian connectivity of G ensures the Hamiltonicity of $Sw(G)$.*

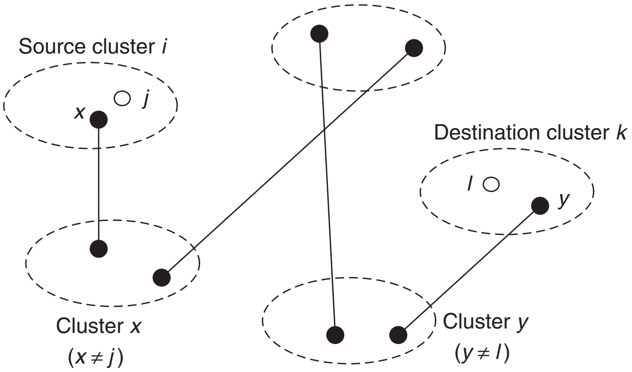

Fig. 5. A Hamiltonian path through the clusters.

Proof. Consider nodes v_{ij} and v_{kl} in clusters i and $k (i \neq k)$ of $Sw(G)$ and a Hamiltonian path between their respective clusters in the n-node complete graph formed by the clusters (supernodes). This path begins in source cluster i, proceeds to cluster $x (x \neq j)$, and continues to cluster $y (y \neq l)$ before entering the destination cluster k (Fig. 5). Adding an intracluster Hamiltonian path between pairs of nodes shown in each of the clusters of Fig. 5 produces a Hamiltonian path between v_{ij} and v_{kl}. Even though our proof postulates the use of two clusters other than the source and destination clusters, the result is easily established for $n = 3$ as well. When $i = k$, we first draw an intracluster Hamiltonian path $v_{ij} \to \cdots \to v_{is} \to v_{it} \to \cdots \to v_{il}$ between v_{ij} and v_{il}, where $s \neq i$ and $t \neq i$ are not necessarily distinct from l. We next build a path from v_{is} to v_{it} that goes through all nodes in the other $n - 1$ clusters but not through any other node in cluster i, using the method of the preceding case. Replacing the transition $v_{is} \to v_{it}$ in the drawn path from v_{ij} to v_{il} with the path just built yields the desired Hamiltonian path between v_{ij} and v_{kl}. Noting that v_{ij} and v_{il} in the latter case can be neighboring nodes in cluster i proves the Hamiltonicity of $Sw(G)$. □

Hamiltonian connectivity of G is too strong a condition for Hamiltonicity of $Sw(G)$. We next prove that Hamiltonicity of G is sufficient for $Sw(G)$ to be Hamiltonian. We prove the result in two separate theorems, for odd and even values of n, because the methods used are different and the n odd case allows a much simpler construction.

Theorem 6. *Hamiltonicity for odd n: If G contains an odd number of nodes and is Hamiltonian, then so is $Sw(G)$.*

Proof. Let $n = 2h + 1$ and assume that a Hamiltonian cycle goes through nodes $0, 1, 2, \ldots, 2h, 0$ of G, in order, renumbering the nodes and clusters, if necessary, to accomplish this. Then, the following is a Hamiltonian cycle that begins and ends in node h of cluster 0. Each line contains a cluster number, the first node visited in that cluster (entry), and the last node visited (exit). The direction is always backward (toward smaller indices), so that the entry point h and the exit point $h+1$ represent the path $h, h-1, h-2, \ldots, h+2, h+1$ within the cluster.

Cluster	Entry	Exit
0	h	$h+1$
$h+1$	0	1
1	$h+1$	$h+2$
$h+2$	1	2
2	$h+2$	$h+3$
$h+3$	2	3
$\vdots$		
$h-1$	$2h-1$	$2h$
$2h$	$h-1$	h
h	$2h$	0 (back to cluster 0)

Note that the first (last) node visited in each cluster has the same index as the (preceding) following cluster, ensuring the availability of intercluster links. Fig. 6 provides an example. □

Theorem 7. *Hamiltonicity for even n: If G contains an even number $n = 2h$ of nodes and is Hamiltonian, then so is $Sw(G)$.*

Proof. We prove this by induction on n, assuming that an n-node cluster's Hamiltonian cycle is $0, 1, 2, \ldots, n-1, 0$. The result holds for $n = 4$, as easily seen from Fig. 2b after removing the two vertical links in each cluster to leave only its Hamiltonian cycle. Our induction basis is $n = 6$, for which a Hamiltonian cycle is shown as the heavy solid line in Fig. 7a (ignore clusters 6 and 7 and nodes 6 and 7 in other clusters for now). As part of the basis of our induction, we note that the Hamiltonian cycle for $n = 6$ is such that clusters 2 and 3 are the only clusters in which the link between nodes 0 and $n - 1 = 5$ is unused; this property will be maintained throughout. Now assuming that the theorem holds for n-node clusters, we construct a Hamiltonian path for the case of $(n+2)$-node clusters, while maintaining the property that the link between nodes 0 and $n+1$ is not part of the Hamiltonian path only in clusters 2 and 3. Start by drawing a Hamiltonian path through all the nodes 0 to $n - 1$ of clusters 0 to $n - 1$, assuming existence of a link between nodes 0 and $n - 1$, as would be the case in an n-node Hamiltonian cluster (Fig. 7a). Then draw a cycle encompassing all nodes in the remaining two clusters n and $n + 1$ and nodes n and $n + 1$ in clusters 0 and 1 (the heavy dotted line in Fig. 7a). Next, merge the two cycles of Fig. 7a into one via simple modifications in clusters 0 and 1, using nodes n and $n + 1$ in clusters 2 and 3, and modifying the paths within clusters 4 through $n - 1$ to cover nodes n and $n + 1$ in those clusters (Fig. 7b). Note that in the final drawing, the link between nodes 0 and $n + 1$ remains unused only in clusters 2 and 3, as postulated in our inductive argument. □

Modular construction of networks is of great significance. Given limitations of the currently available and forthcoming

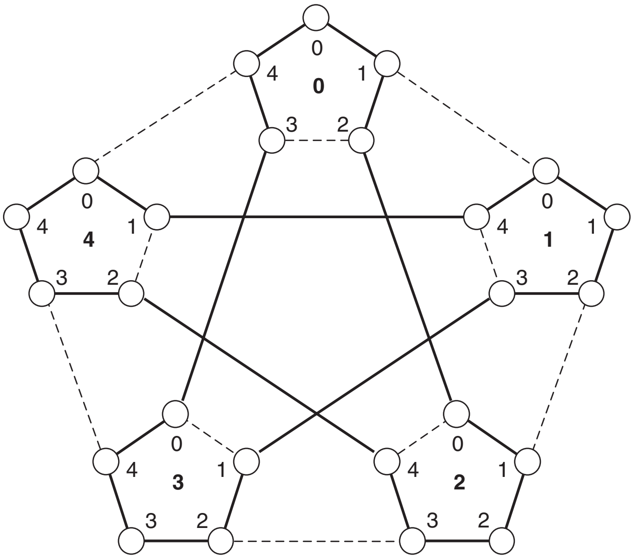

Fig. 6. A Hamiltonian cycle in a swapped network whose 5-node basis network is Hamiltonian.

packaging technologies for digital systems, a hierarchy of packaging levels is imposed so that crossing the packaging boundaries is undesirable, not only in terms of implementation cost, but also with regard to communication performance [16]. Swapped networks are naturally modular.

Theorem 8. *Modularity*: *For any integer m that divides n, an n^2-node swapped network can be partitioned into m modules, each having $(n^2/m)(1 - 1/m)$ external links.*

Proof. Each of the m modules holds n/m clusters, or n^2/m nodes. Within a cluster, n/m of the nodes do not have intermodule links; these are nodes v_{ii} and nodes connected to the other $n/m - 1$ clusters in the same module. So, the number of intermodule links is $(n/m)(n - n/m)$. □

The modularity suggested by Theorem 8, requiring one external link per node in a module of n^2/m nodes, is better than those of the hypercube whose partitioning requires $(n^2/m)\log_2 m$ links per module. It also compares favorably with other networks of similar performance.

5. Fault tolerance

Swapped networks are quite robust. This is intuitively justified by noting that the complete connectivity among clusters is not much affected even if all nodes in a cluster, or parts of several clusters, fail. As exemplified by the pair of paths depicted in Fig. 3, there are typically many node- and edge-disjoint paths between pairs of nodes in various clusters. The existence of intercluster links also has a positive effect on fault tolerance within the same cluster in the sense that nodes becoming inaccessible in the cluster due to faults may be reachable through intermediaries in other clusters.

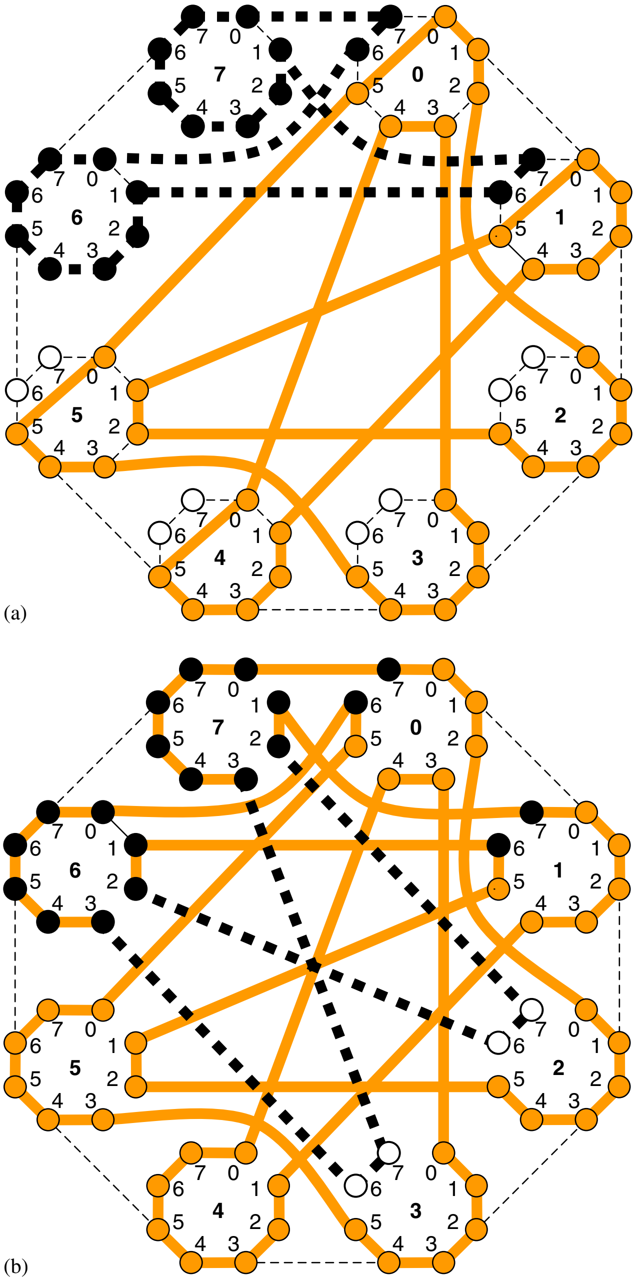

Fig. 7. Building a Hamiltonian cycle for $Sw(G)$, with $|G| = n+2$ (even): (a) two partial cycles and (b) merging the two cycles of (a).

We first present two key definitions relating to the notion of fault tolerance in networks.

Definition 7. *Graph connectivity*: A graph G is h-connected if it contains h different node- and edge-disjoint paths between any pair of nodes.

Definition 8. *Fault diameter of a network*: When a worst-case set of at most $h - 1$ faulty nodes are removed from an h-connected network, the diameter of the surviving part of the network, which is guaranteed to be connected, is the *fault diameter* of the original network.

A network that is considered robust generally has a high connectivity and a fault diameter that is the same as, or only slightly greater than, that of the fault-free network. The following three theorems collectively establish the strong fault-tolerance features of swapped networks.

Theorem 9. *Connectivity of swapped networks*: *If G is h-connected, Sw(G) is also h-connected* ([2, Theorem 6]).

Theorem 10. *Fault diameter of swapped networks*: *If the fault diameter of the basis network G is $D + \varepsilon$, the fault diameter of Sw(G) is no greater than $2D + 3 + \varepsilon$.*

Proof. Let the basis network G be h-connected. Then, the fault diameter of G being $D + \varepsilon$ means that for up to $h - 1$ faulty nodes in G, the diameter of the remaining (connected) network is $D + \varepsilon$ or less. Because the connectivity of $Sw(G)$ is also h, we need to consider the distance from a source node v_{ij} to a destination node v_{kl} in the surviving network when $Sw(G)$ has $h - 1$ or fewer faults. Note that h-connectivity implies that the minimum node degree is h or greater. Consider the source node j along with h of its neighbors in the source cluster i (Fig. 8). Let these neighbors be $x_1, x_2, \ldots, x_h$. These $h + 1$ nodes (node j and its h neighbors) collectively have at least h swap links to different clusters $m_1, m_2, \ldots, m_h$. Assume the worst case when none of these h clusters is the destination cluster k. Consider also nodes $m_1, m_2, \ldots, m_h$ in the destination cluster k. There should be some m_g, $1 \leqslant g \leqslant h$, for which both cluster m_g and node m_g in cluster k are fault-free. This is because at least h faults are needed to have a fault in cluster m_r or node m_r in cluster k for every r. Then, it is easy to see that the path from v_{ij}, perhaps via one of its neighbors in cluster i, to node i in cluster m_g (1 or 2 hops thus far), then to node k of the fault-free cluster m_g (at most D hops), to node m_g of cluster k (1 hop) and finally to node v_{kl} in cluster k (at most $D+\varepsilon$ hops) is of length no greater than $2D+3+\varepsilon$. □

We conjecture that the tighter bound $2D+2+\varepsilon$ applies to the fault diameter of a swapped network of any kind. Also, placing certain mild restrictions on the structure of G may allow the establishment of the optimal value $2D+1+\varepsilon$ as the fault diameter. These improvements are being investigated.

Swapped networks are also provably survivable in the sense that they do not contain any obvious vulnerability points [7]. Consider, for example, the behavior of swapped networks under complete cluster failures, as opposed to a small number of random node failures. The following result demonstrates robustness of $Sw(G)$ in the face of such failures.

Theorem 11. *Robustness under multiple cluster failures*: *The diameter of an incomplete swapped network with $h - 1$ or fewer of its n clusters completely removed*, *where h is the connectivity of the basis network G*, *does not exceed $2D + 3$* (*i.e.*, *it is at most 2 hops more than the diameter of the fault-free swapped network*).

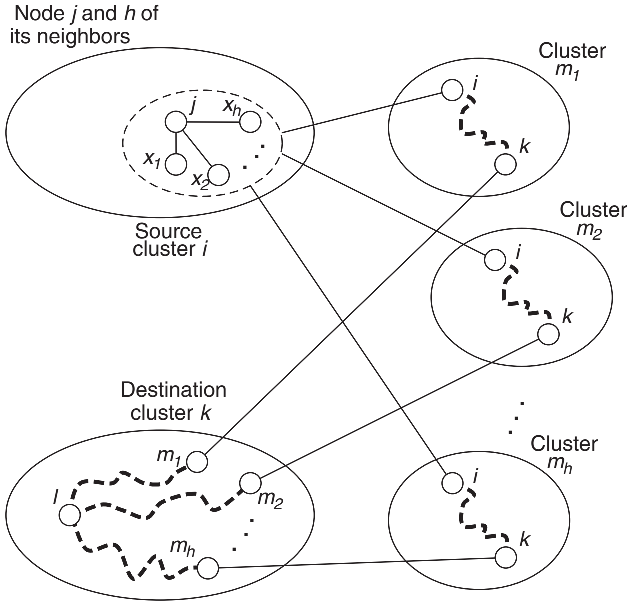

Fig. 8. The nodes and intermediate clusters of $Sw(G)$ used in the proof of Theorem 10.

Proof. The proof uses the same notation as in proving Theorem 10. Of clusters $m_1, m_2, \ldots, m_h$ in Fig. 8, at least one must be fault-free. Let it be cluster $m_s = q$. Then, given nodes v_{ij} and v_{kl} in fault-free clusters i and k, the path $v_{ij} \to v_{iq} \to v_{qi} \to \cdots \to v_{qk} \to v_{kq} \to \cdots \to v_{kl}$ is of length at most $2D + 3$, given that each of the two paths represented by the ellipses is of length D or less. □

6. Conclusion

We derived some general properties of a swapped network $Sw(G)$ based on parameters and structure of the basis network G. In particular, we showed that swapped networks are cost-effective (low diameter with fairly small node degree), modular, packageable, and quite robust. We also showed that the Hamiltonian connectivity and Hamiltonicity properties are preserved in hierarchical structures built as swapped networks. In support of the cost-effectiveness claim, note that, in going from G to $Sw(G)$, the degree-diameter product grows by the factor $(d + 1)(2D + 1)/(dD) = 2 + 2/d + 1/D + 1/(dD)$. This ratio is significantly less than 4, the corresponding growth factor for the hypercube, when the basis network G is neither too sparse nor too dense. For example, when G is a modest-size square 2D mesh, the ratio is close to 2.5. Table 1 indicates how the main topological parameters of certain classes of interconnection networks change as the network size is scaled from n to n^2. The various parameters are formulated in terms of the respective parameters for the network of size n. As evident from Table 1, swapped networks offer a mechanism for increasing the size of a network with

Table 1
Change in topological parameters for various ways of increasing network size from n to n^2

Network	Degree	Diameter	Avg. dist.	Bisection
Complete	$d^2 + 2d$	D	Δ	$4B^2$
Squared network	$2d$	$2D$	2Δ	nB
Hypercube	$2d$	$2D$	2Δ	$2B^2$
2D square torus	d	D^2	$4\Delta^2/3$	B^2
Swapped network	$d + 1$	$2D + 1$	$2\Delta + 1$	See Thm. 3

For example, an entry of D for diameter means that there is no change when the network size is squared and $2D$ means that the diameter is doubled. Squared network refers to the cross product of a network with itself.

relatively small cost increase, while limiting the deterioration of topological parameters and ensuring strong fault tolerance.

Work in progress includes expanding our results on the properties of swapped network to include other topological properties, as well as a deeper analysis of their performance parameters and robustness attributes. In particular, a fault-tolerant routing algorithm is required if the existence of multiple node- or edge-disjoint paths is to be practically and efficiently exploited. Other fault tolerance considerations, such as fault-survivability and fault-Hamiltonicity (study of Hamiltonicity of the surviving part of a network in the presence of faults), are also under study. Obtaining sharper bounds for the corresponding parameters of swapped-network variants, such as the ones in Definitions 2 and 3, are other possible directions for further investigation. Also of interest are variants of swapped networks that preserve symmetry and other desirable structural properties of a basis network G. For example, it would be desirable to have an alternate or modified form of swapped network that is a Cayley graph when the basis network G is a Cayley graph. This latter goal has thus far eluded us.

It is also possible, though less appealing, to study the properties of swapped networks with particular types of basis networks. Candidates include popular networks such as mesh, torus (both 2D and higher dimensional), hypercube, generalized hypercube, as well as lesser-known or newer networks such as hypernets [8], certain classes of chordal rings [15], and diameter-2 perfect difference networks [18,19]. Finally, comparison of the swapped-network mechanism with competing mechanisms, such as product networks [29], for systematically increasing the network size merits some attention.

References

[1] B.W. Arden, H. Lee, Analysis of chordal ring networks, IEEE Trans. Comput. 30 (1981) 291–295.

[2] K. Day, A. Al-Ayyoub, Topological properties of OTIS-networks, IEEE Trans. Parallel Distributed Systems 13 (2002) 59–366.

[3] G. Della Vecchia, C. Sanges, Recursively scalable networks for message passing architectures, in: Proceedings of the Conference on Parallel Processing and Applications, 1987, pp. 33–40.

[4] G. Della Vecchia, C. Sanges, A recursively scalable network for VLSI implementation, Future Generation Comput. Systems 13 (1998) 235–243.

[5] J.S. Fu, G.H. Chen, Hamiltonicity of the hierarchical cubic network, Theory Comput. Systems 35 (2002) 59–79.

[6] K. Ghose, K.R. Desai, Hierarchical cubic networks, IEEE Trans. Parallel Distributed Systems 6 (1995) 427–435.

[7] A.M. Hobbs, Network survivability, in: J.G. Michaels, K.H. Rosen (Eds.), Applications of Discrete Mathematics, McGraw-Hill, New York, 1991, pp. 332–353.

[8] K. Hwang, J. Ghosh, Hypernet: a communication efficient architecture for constructing massively parallel computers, IEEE Trans. Comput. 36 (1987) 1450–1466.

[9] D.-M. Kwai, B. Parhami, Pruned three-dimensional toroidal networks, Inform. Process. Lett. 68 (1998) 179–183.

[10] D.-M. Kwai, B. Parhami, A unified formulation of honeycomb and diamond networks, IEEE Trans. Parallel Distributed Systems 12 (2001) 74–80.

[11] G. Marsden, P. Marchand, P. Harvey, S. Esener, Optical transpose interconnection system architectures, Opt. Lett. 18 (1993) 1083–1085.

[12] A. Osterloh, Sorting on the OTIS-mesh, in: Proceedings of the 14th International Parallel and Distributed Processing Symposium, 2000, pp. 269–274.

[13] B. Parhami, Introduction to Parallel Processing: Algorithms and Architectures, Plenum Press, New York, 1999.

[14] B. Parhami, Computer Architecture: From Microprocessors to Supercomputers, Oxford University Press, New York, 2005.

[15] B. Parhami, D.-M. Kwai, Periodically regular chordal rings, IEEE Trans. Parallel Distributed Systems 10 (1999) 658–672 See also corrections to printer errors in IEEE Trans. Parallel Distributed Systems 10 (1999) 767–768.

[16] B. Parhami, D.-M. Kwai, Challenges in interconnection network design in the era of multiprocessor and massively parallel microchips, in: Proceedings of the International Conference on Communications in Computing, 2000, pp. 241–246.

[17] B. Parhami, D.-M. Kwai, Incomplete k-ary n-cube and its derivatives, J. Parallel Distributed Comput. 64 (2004) 183–190.

[18] B. Parhami, M. Rakov, Perfect difference networks and related interconnection structures for parallel and distributed systems, IEEE Trans. Parallel Distributed Systems, 16 (2005), to appear.

[19] B. Parhami, M. Rakov, Performance, algorithmic, and robustness attributes of perfect difference networks, IEEE Trans. Parallel Distributed Systems, 16(2005), to appear.

[20] S. Rajasekaran, S. Sahni, Randomized routing, selection, and sorting on the OTIS-mesh, IEEE Trans. Parallel Distributed Systems 9 (1998) 833–840.

[21] I.D. Scherson, A.S. Youssef, Interconnection Networks for High-Performance Parallel Computers, IEEE Computer Society Press, Silver Spring, MD, 1994.

[22] C.-F. Wang, S. Sahni, Basic operations on the OTIS-mesh optoelectronic computer, IEEE Trans. Parallel Distributed Systems 9 (1998) 1226–1236.

[23] C.-F. Wang, S. Sahni, Image processing on the OTIS-mesh optoelectronic computer, IEEE Trans. Parallel Distributed Systems 11 (2000) 97–109.

[24] C.-F. Wang, S. Sahni, Matrix multiplication on the OTIS-mesh optoelectronic computer, IEEE Trans. Comput. 50 (2001) 635–646.

[25] C.-H. Yeh, B. Parhami, Recursively fully-connected networks: a class of high-performance low-degree interconnection networks, in: Proceedings of the 11th International Conference on Computers and their Applications, 1996, pp. 227–230.

[26] C.-H. Yeh, B. Parhami, Swapped networks: unifying the architectures and algorithms of a wide class of hierarchical parallel processors, in: Proceedings of the International Conference on Parallel and Distributed Systems, 1996, pp. 230–237.

[27] C.-H. Yeh, B. Parhami, Hierarchical swapped networks: efficient low-degree alternatives to hypercube and generalized hypercube, in: Proceedings of the International Symposium on Parallel Architectures, Algorithms, and Networks, 1996, pp. 90–96.

[28] C.-H. Yeh, B. Parhami, Recursive hierarchical swapped networks: versatile interconnection architectures for highly parallel systems, in: Proceedings of the Eighth IEEE Symposium Parallel and Distributed Processing, 1996, pp. 453–460.

[29] A. Youssef, Design and analysis of product networks, in: Proceedings of the Symposium on Frontiers of Massively Parallel Computation, 1995, pp. 521–528.

[30] F. Zane, P. Marchand, R. Paturi, S. Esener, Scalable network architectures using the optical transpose interconnection system, J. Parallel Distributed Computing 60 (2000) 521–538.

Behrooz Parhami received his Ph.D. in computer science from University of California, Los Angeles, in 1973. Presently, he is Professor in the Department of Electrical and Computer Engineering, University of California, Santa Barbara. His research deals with parallel architectures and algorithms, computer arithmetic, and reliable computing. In his previous position with Sharif University of Technology in Tehran, Iran (1974–88), he was also involved in the areas of educational planning, curriculum development, standardization efforts, technology transfer, and various editorial responsibilities, including a five-year term as Editor of Computer Report, a Farsi-language computing periodical.

Dr. Parhami's technical publications include over 220 papers in journals and international conferences, a Farsi-language textbook, and an English/Farsi glossary of computing terms. Among his latest publications are two graduate-level textbooks on parallel processing (Plenum, 1999) and computer arithmetic (Oxford, 2000) and an introductory textbook on computer architecture (Oxford, 2005).

Dr. Parhami is a Fellow the IEEE, a Chartered Fellow of the British Computer Society, a member of the Association for Computing Machinery, and a Distinguished Member of the Informatics Society of Iran for which he served as a founding member and President during 1979–84. He also served as Chairman of IEEE Iran Section (1977–86) and received the IEEE Centennial Medal in 1984.