Parallel Computing 3 (1986) 231-249
North-Holland

Fault-tolerance and performance analysis of beta-networks

John P. SHEN

Department of Electrical and Computer Engineering, Carnegie-Mellon University, Pittsburgh, PA 15213, U.S.A.

John P. HAYES

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, U.S.A.

Luigi CIMINIERA and Angelo SERRA

Dipartimento Automatica e Informatica, Politecnico di Torino, 10129 Torino, Italy

Received January 1985
Revised June 1985

Abstract. The relationship between fault tolerance and performance is explored for β-networks used as interconnection networks in multicomputer systems. The networks of interest are composed of 2×2 switches (β-elements) and are represented by a graph model called a β-graph. Two parameters derived from β-graphs are used to characterize β-networks. The fault tolerance (FT) parameter is the maximum number of β-element faults that can be tolerated. The communication delay (CD) parameter, representing the worst-case delay between any pair of computers, is used as a measure of the performance of the β-networks. Tight bounds for both FT and CD parameters are derived. Two important classes of β-networks are introduced, namely, DPR-networks and MISE-networks. It is shown that DPR-networks possess the maximal fault tolerance, and the class of DPR-networks is unique in achieving the maximum possible fault tolerance. The class of MISE-networks is minimally fault tolerant, but has the minimum communication delay. A class of β-networks, called RDTT-networks, that achieve an optimal balance of the FT and CD parameters is also presented.

Keywords. Interconnection networks, multicomputer systems, fault-tolerance, performance analysis, beta-networks.

1. Introduction

A number of recently proposed multicomputer systems use a class of interconnection networks called β-networks as intercomputer communication networks [7,8,10]. A *multicomputer* system is considered here to be a distributed system of computing units supported by an interconnection network which provides the communication paths among the computing units. An $N\times N$ *β-network* is an interconnection network with N input and N output terminals which is composed of 2×2 switching elements called β-elements. Each β-element can be set to one of two states, namely, the through (T) state or the cross (X) state, to provide interconnecting paths from the N input terminals to the N output terminals. As illustrated in Fig. 1, the set of N computing units in a multicomputer system can be equated with the set of input terminals and the set of output terminals of the β-network. Because of the existence of the N feedback paths through the N computing units, the N input links and the N output links of the β-network are considered to be identical and are called the N *terminal links* of the β-network.

0167-8191/86/$3.50 © 1986, Elsevier Science Publishers B.V. (North-Holland)

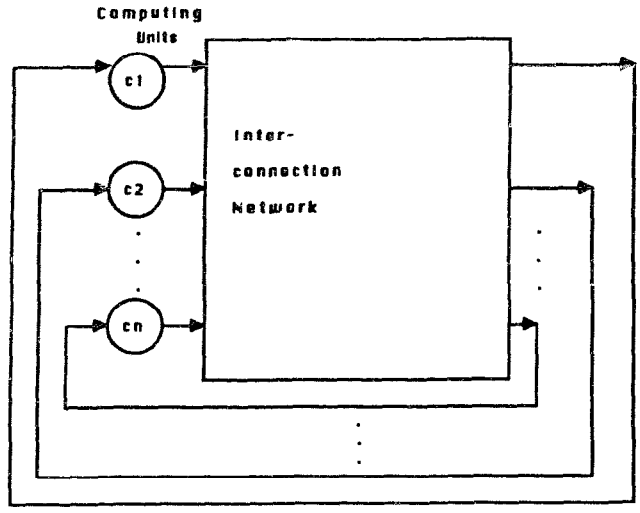

Fig. 1. A multicomputer system model.

In [12] a theoretical framework for fault-tolerance analysis of β-networks was introduced. Pertinent results from that work are now summarized here. A fault model is used which allow β-elements to be stuck in either of their two normal states, i.e., stuck-at-through (s-a-T) or stuck-at-cross (s-a-X). A connectivity property called *dynamic full access* (DFA) serves as the criterion for fault tolerance in β-networks. A β-network has the DFA property if each of its inputs can be connected to any one of its outputs via a finite number of passes through the β-network. A *fault* in a β-network is a collection of β-element stuck-at faults. A fault is said to be *critical* if it destroys the DFA property of the β-network. A *minimal critical fault* is a critical fault none of whose proper subsets constitutes a critical fault. The fault tolerance of a β-network is defined as its ability to maintain DFA in spite of the presence of stuck-at-T/X faults in its β-elements. A β-network with DFA is *k-fault tolerant* or *k*-FT if the failure, either s-a-T or s-a-X, of any k or fewer β-elements does not destroy DFA. The largest k for which a β-network is *k*-FT is called the *fault-tolerance (FT) parameter* of the β-network.

For analysis purposes, β-networks are represented by graphs [5] called β-graphs. The *labeled β-graph* of a β-network is a labeled directed graph with vertices representing the β-elements, and edges representing the links of the β-network. An edge is labeled and called a *terminal edge* if it corresponds to a terminal link of the β-network, otherwise it is not labeled and is called an *intermediate edge*. An *unlabeled β-graph*, or simply a *β-graph*, is a labeled β-graph with all its edge labels deleted. Figure 2 illustrates the labeled β-graph of a β-network called the indirect binary 2-cube network [10] which connects four computing units $\{1, 2, 3, 4\}$. Each computing unit is implicitly represented by a terminal edge in the β-graph. Usually the terminal edges are labeled with the indices of the associated computing units as depicted in Fig. 2. Each β-element in a β-network is modeled by a vertex with two incoming and two outgoing edges in the corresponding β-graph. When the β-element is stuck at one of its two states, an incoming edge can only be connected to one of the outgoing edges. Hence, a β-element stuck-at fault can be modeled by the splitting of the corresponding β-graph vertex into two subvertices, each with one incoming and one outgoing edge. It is easily seen that a β-network has the DFA property if and only if the corresponding β-graph is strongly connected. Fault tolerance in terms of the

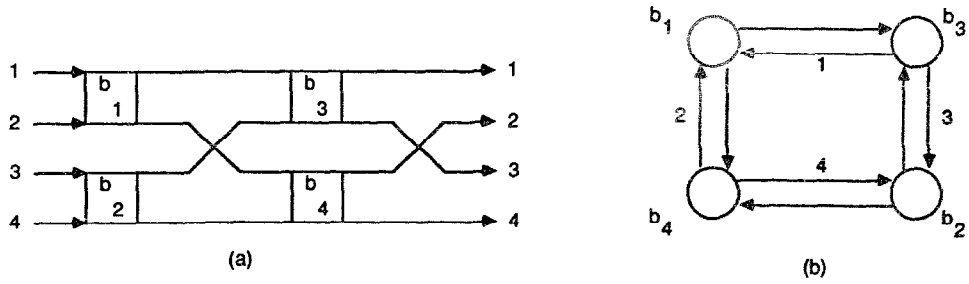

Fig. 2. (a) The indirect binary 2-cube network; (b) Its labeled β-graph.

β-graph can thus be defined as the ability of the β-graph to stay strongly connected in spite of the splitting of its vertices.

In this paper a second network parameter called the communication delay is introduced as a measure of the performance of a β-network. Tight bounds on both the fault-tolerance and the communication-delay parameters are derived in Section 2. A fundamental relationship between the two parameters is also established. Two important specific classes of β-networks are introduced, namely DPR-networks and MISE-networks. These networks represent two extreme design choices for fault-tolerant β-networks. It is shown in Section 3 that DPR-networks possess the maximal fault tolerance and maximal communication delay. It is further shown that the class of DPR-networks is unique in achieving the maximal fault tolerance. Section 4 presents MISE-networks which are demonstrated to be minimally fault tolerant, but have the minimum communication delay. Section 5 introduces a class of optimal β-networks called RDTT-networks that exhibits the best combination of fault tolerance and performance.

2. Fault tolerance vs. communication delay

Traditionally, high speed or performance has been the primary objective in the design of interconnection networks. Due to the proliferation of fault-critical applications of computers, fault tolerance is also becoming an important design requirement. It is probable that the β-networks of future multicomputer systems will attempt to strike a balance between performance and fault tolerance. This implies the need for a uniform approach to the measurement of these parameters. The fault-tolerance parameter k defined earlier can serve as a measure of the fault tolerance of a β-network. The performance of a β-network can be measured by a basic connectivity property such as dynamic full-access (DFA). However, since all useful β-networks have the DFA property, a finer, and preferably numerical, measure of performance is needed. In the following section a suitable performance parameter is introduced, which is based on the intercomputer communication delays imposed by a β-network.

The communication dealy from computing unit to computing unit is measured here by the minimum number of β-elements that need to be traversed by data being sent from unit i to unit j. A communication delay parameter d for a β-network is obtained by considering the communication delays between all pairs of computing units and choosing the maximum or worst-case value of these delays. We formalize the above definition by making use of the β-graphs model of β-networks. The *edge-distance*, or simply *distance*, from edge i to edge j in a β-graph is the number of intermediate vertices in the shortest directed path having edges i and j as its first and last edges, respectively. The *edge-diameter*, or simply *diameter*, of a β-graph is the longest distance between any two edges of the β-graph. The *communication delay (CD)*

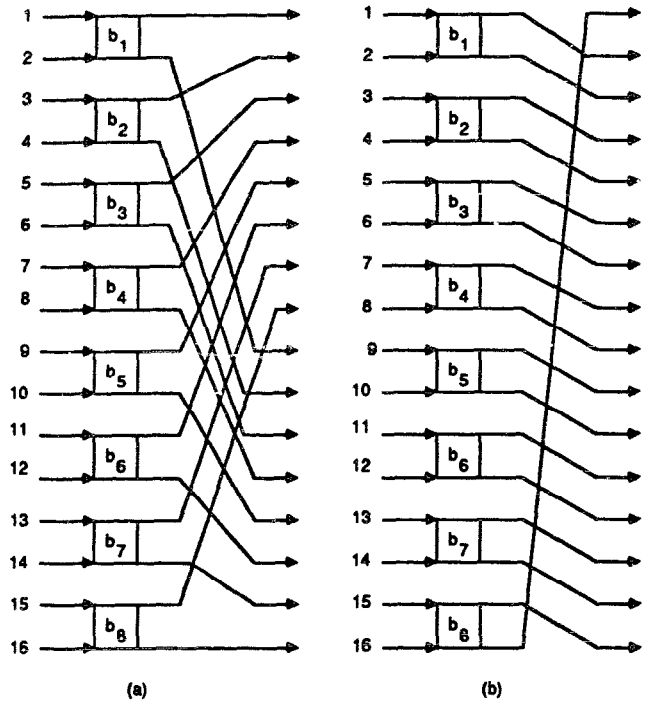

Fig. 3. (a) The 16×16 ISE-network; (b) The 16×16 SCS-network.

parameter d of a β-network is the diameter of its β-graph. The delays caused by the computing units themselves are not explicitly considered here. However, if two β-networks have the same number of stages of β-elements, their CD parameters will indicate their relative communication performance.

As an illustration, we now consider two single-stage β-networks having very different values of the CD parameter d. The β-network of Fig. 3(a) is the 16×16 inverse shuffle exchange network, or ISE-network [15]. We call the β-network of Fig. 3(b) the 16×16 *single cycle shift network*, or SCS-network. Figures 4(a) and (b) depict the β-graphs of Figs. 3(a) and (b) respectively. The diameters of these two β-graphs, and therefore the CD parameters of the corresponding β-networks, can be easily computed. The CD parameter for the 16×16 ISE-network is four, and that of the 16×16 SCS-network is eight. In the SCS-network, for example, the communication delay is eight when sending a message from computer 1 to computer 16. ISE-networks and SCS-networks of other sizes can be similarly constructed. In general, an $N \times N$ ISE-network, if $N = 2^m$ for some integer m, has CD parameter $d_1 = \log_2 N$, and an $N \times N$ SCS-network has CD parameter $d_2 = N/2$ [13].

Both the FT parameter k and the CD parameter d depend on the structure of a β-network, and are therefore properties of its β-graph. We now derive tight lower and upper bounds for k and d in terms of n, the number of β-elements in the β-network. We also establish a fundamental relationship between k and d. The smallest possible value for k is clearly zero, while the largest possible value of k at first glance seems to be n. However, if a β-network can tolerate faults affecting all its n β-elements, then the entire β-network is unnecessary; hence k cannot exceed $n - 1$. It can be shown that both the ISE-network and the SCS-network are 0-FT.

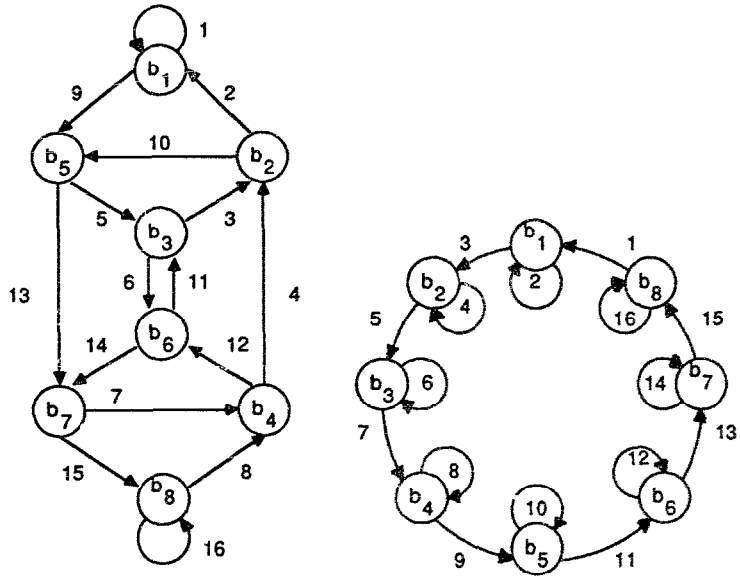

Fig. 4. (a) β-graph of the 16×16 ISE-network; (b) β-graph of the 16×16 SCS-network.

In subsequent section we demonstrate the existence of a class of $(n-1)$-FT β-networks. Hence 0 and $n-1$ are tight lower and upper bounds, respectively, for k. The corresponding bounds for d are given in the following lemma.

Lemma 1. *For any β-network with $n > 2$ β-elements and CD parameter d,*

$$\lfloor \log_2 n \rfloor + 1 \leqslant d \leqslant n.$$

These bounds are tight.

Proof: The CD parameter d is the diameter of the β-graph. The diameter of a β-graph of n vertices cannot exceed n, the total number of vertices. The value $d = n$ is achievable, for example, in the case of SCS-network containing n β-elements. Hence n is a tight upper bound on d. Since each vertex in a β-graph has two outgoing edges, the outdegree of a vertex is two. The maximum number of edges which are exactly at distance d from any edge i is 2^d. The maximum number of distinct edges reachable from i within distance d is $2^d + 2^{d-1} + \ldots + 2 = 2^{d+1} - 2$. We also know that a β-graph with n vertices has exactly $2n$ edges, and each edge must be able to reach the other $2n - 1$ edges. Hence the following inequality must hold:

$$2^{d+1} - 2 \geqslant 2n - 1$$

from which it follows that

$$d \geqslant \lceil \log_2 (n + 1/2) \rceil$$

Now $\lceil \log_2 x \rceil = \lfloor \log_2 x \rfloor + 1$ unless x is an integral power of 2, which is impossible when $x = n + 1/2$. Hence

$$d \geqslant \lfloor \log_2 n \rfloor + 1.$$

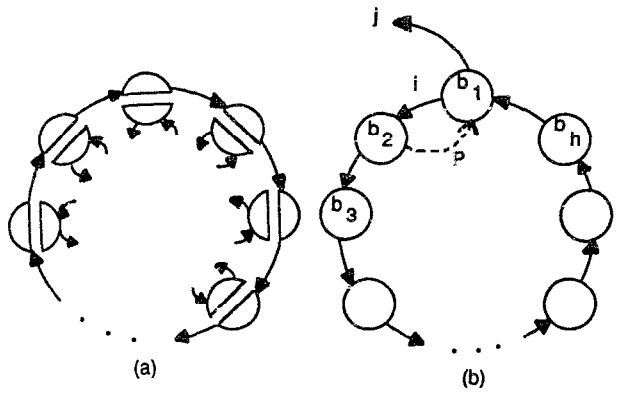

Fig. 5. (a) A critical fault corresponding to an elementary circuit; (b) An elementary circuit of length $h \geq k+1$.

This lower bound for d is achieved by the ISE-network of n β-elements, hence it also is tight. □

Lemma 2. *For any β-network with FT parameter k and CD parameter d, $k \leq d-1$.*

Proof. Every elementary circuit in a β-graph corresponds to a critical fault of the β-network. The β-elements represented by the vertices of the elementary circuit can fail in such a way that the edges of the elementary circuit are isolated from the rest of the β-graph as shown in Fig. 5(a). Hence the β-graph of a k-FT β-network must not contain any elementary circuit of length less than or equal to k.

Let C be an elementary circuit of minimal length $h \geq k+1$ in a k-FT β-graph. Assume that edges i and j have the same source vertex, and that C contains edge i, as illustrated in Fig. 5(b). Edge i can reach edge j via the edge of C, i.e., there exists a directed path from i to j consisting of all the edges of C. Since the length of C is h, the distance from i to j is at most h. In fact, it must be exactly h, otherwise there must exist a path P from the destination vertex of i to the source vertex of j containing at most $h-2$ edges. In that case the path P together with edge i constitutes an elementary circuit of length $h-1$ or less. This contradicts the minimal-length property of C. Hence the distance from i to j is exactly h. Since $h \geq k+1$, the distance from i to j must be $k+1$ or more. Hence the diameter of the β-graph cannot possibly be less than $k+1$. Therefore $k \leq d-1$. □

We can summarize the foregoing results in the following theorem.

Theorem 1. *Let N be a β-network with n β-elements, where $n > 2$. The FT parameter k and the CD parameter d of N are related as follows:*

$$0 \leq k \leq n-1, \tag{1}$$

$$\lfloor \log_2 n \rfloor + 1 \leq d \leq n, \tag{2}$$

$$k \leq d-1. \tag{3}$$

The bounds in (1) *and* (2) *are tight.*

Every β-network has a unique d. As faults occur, k tends to decrease, while d tends to increase. A β-network loses the DFA property when d becomes infinite. The synthesis of practical fault-tolerant β-network involves finding an appropriate balance between k and d.

3. Maximal fault tolerance

In Section 2 we showed that the largest possible value for the fault-tolerance parameter k is $n-1$, where n is the number of β-elements. In this section we prove the existence and uniqueness of a class of β-networks with a cyclic structure which meets this upper bound on k. A *double parallel ring* β-network, or DPR-network, of order n is a β-network with n β-elements, whose β-graph consists of two disjoint and parallel Hamiltonian circuits as shown in Fig. 6. (A *Hamiltonian circuit* of a directed graph is a circuit that passes through every vertex exactly once.)

Theorem 2. *The FT parameter k of the DPR-network of order n is* $n-1$.

Proof. This theorem can be proven by examining the minimal critical faults (MCFs) of the DPR-network. We need to show that every MCF consists of more than $n-1$ β-element stuck-at faults. From Fig. 6 we see that there is no elementary circuit of length less than n in the β-graph of the DPR-network of order n. Each circuit partition [12] of the β-graph must consist of exactly two elementary circuit each of length n. Hence every CA-graph [12] or Fig. 6 must consist of two vertices connected by n edges. (The vertices of a CA-graph represent elementary circuits, while its edges represent vertices of the original β-graph that are common to two elementary circuits). Clearly each CA-graph has only one cut-set consisting of all its n edges. There is a one-to-one correspondence between the MCFs of a β-network and the cutsets of its CA-graphs (Theorem 1 of [12]) Hence every MCF of the DPR-network consists of n faulty β-elements. Therefore all n β-elements must be stuck at T/X to destroy the DFA property of the DPR-network. Consequently, the DPR-network of order n is $(n-1)$-FT. □

From Theorem 1 we know that $n-1$ is a tight upper bound for k in β-networks with n β-elements. The DPR-network clearly meets this upper bound. We can therefore say that the DPR-network of order n is maximally fault tolerant among all β-networks with n β-elements. Thus in a DPR-network only one fault-free β-element is needed to ensure DFA. The maximal fault tolerance of the DPR-network can also be viewed from another perspective which is developed below.

The fault tolerance of a β-network can be analyzed in terms of the number of Eulerian circuits in its β-graph. An *Eulerian circuit* of a directed graph is a circuit that includes every edge in the graph exactly once. Eulerian circuits represent maximal noncritical states of a β-network. Intuitively, we expect the number of Eulerian circuits to be proportional to the degree of fault tolerance. A classical result in graph theory [1] states that the number of Eulerian

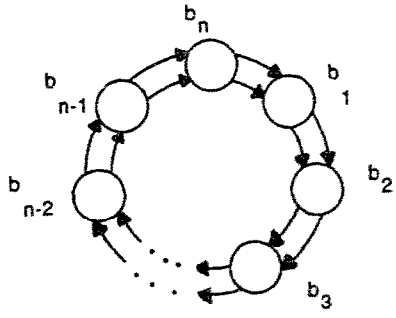

Fig. 6. β-graph of a DPR network of order n.

circuits in a β-graph is equal to the number of distinct spanning trees rooted at any one of the vertices. A *rooted spanning tree* of a β-graph of n vertices consists of $n-1$ edges each from a different vertex and all directed toward the root vertex. Since every vertex in the β-graph has two outgoing edges, either one can be chosen for inclusion in a rooted spanning tree. Hence in a β-graph of n vertices, there are at most 2^{n-1} distinct spanning trees rooted at a particular vertex. As a result, a β-graph of n vertices has at most 2^{n-1} distinct Eulerian circuits. From the structure of the β-graph of the DPR-network, we see that all 2^{n-1} combinations of $n-1$ edges indeed correspond to distinct rooted spanning trees. Hence we have the following result:

Lemma 3. *The β-graph of the DPR-network of order n has 2^{n-1} distinct Eulerian circuits, which is the maximum number possible in any β-graph of n β-elements..*

A β-network state is determined by the states of the n β-elements. If $s_i = s(b_i) \in \{T, X\}$ denotes the state of the β-element b_i, then a *state* of the β-network is represented by an n-tuple $s(b_1, b_2, \ldots, b_n) = (s_1, s_2, \ldots, s_n)$. An *Eulerian circuit* (EC) state is a β-network state that specifies an Eulerian circuit in the β-graph. We next show that being $(n-1)$-FT is equivalent to having 2^{n-1} Eulerian circuits. First, we need the following lemma.

Lemma 4. *If e_1 and e_2 are two distinct EC states of a β-graph, then e_1 and e_2 cannot differ in only one of their entries, i.e., at least two β-elements must be in different states.*

Proof. Assume that the EC states e_1 and e_2 are identical except in the ith entry, i.e., $e_1 = (s_1, s_2, \ldots, s_i, s_{i+1}, \ldots, s_n)$ and $e_2 = (s_1, s_2, \ldots, \bar{s}_i, s_{i+1}, \ldots, s_n)$, where if $s_i = T$ (X) then $\bar{s}_i = X$ (T). We illustrate the Eulerian circuit corresponding to e_1 in Fig. 7(a), highlighting the ith β-element. The Eulerian circuit of e_2 is identical to that shown in Fig. 7(a) except in the ith β-element as shown in Fig. 7(b). Clearly Fig. 7(b) consists of two disjoint circuits; it is therefore impossible for e_2 to represent an Eulerian circuit. Hence e_1 and e_2 must differ in more than one entry. □

Theorem 3. *A β-network of n β-elements is $(n-1)$-FT if and only if its β-graph has 2^{n-1} Eulerian circuits.*

Proof. The necessary condition is straightforward. If a β-network is $(n-1)$-FT, then any set of $n-1$ β-elements can be stuck in any of the 2^{n-1} possible faulty states without destroying DFA. Each of these 2^{n-1} states must be compatible [12] with an EC state. Hence, there must exist 2^{n-1} distinct Eulerian circuits in an $(n-1)$-FT β-graph.

We now need to show that a β-graph with 2^{n-1} Eulerian circuits is $(n-1)$-FT. If not, there exists a critical fault f involving $n-1$ β-elements. Let β-elements $b_1, b_2, \ldots, b_{n-1}$ be faulty,

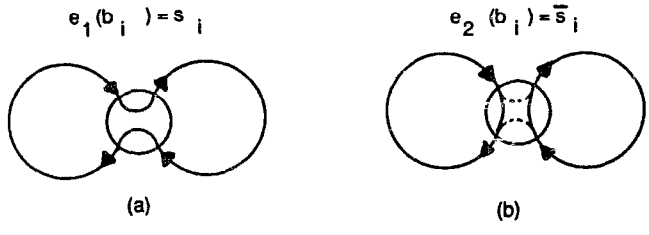

Fig. 7. Relationship between any two Eulerian circuits.

and let b_n be fault-free. The critical fault state of f must not be compatible with any of the 2^{n-1} EC-states [12]. This implies that none of the 2^{n-1} EC-states has the same first $n-1$ entries as f. There are 2^{n-1} possible distinct partial states involving only the first $n-1$ entries. We know that at least one of these 2^{n-1} partial states, namely, the one corresponding to f, does not appear in the set of 2^{n-1} EC-states. Hence there can be at most $2^{n-1}-1$ distinct partial states involving the first $n-1$ entries of the 2^{n-1} distinct EC-states. Therefore at least two of the EC-states, e_1 and e_2, must be identical in the first $n-1$ entries. This means e_1 and e_2 differ only in the nth entry, which is impossible by Lemma 4. Hence if a β-graph has 2^{n-1} Eulerian circuits, the corresponding β-network must be $(n-1)$-FT. □

So far we have shown that the DPR-network of order n has FT parameter $k = n-1$ and has 2^{n-1} Eulerian circuits. We further show that these two properties are equivalent, and characterize the maximal fault tolerance achievable. An interesting question is whether the DPR-network is the only β-network having the property. The answer is yes, as the following theorem asserts.

Theorem 4. *The DPR-network of order n is unique among all β-networks of n β-elements in achieving the maximum possible fault tolerance $k = n-1$.*

Proof. Let G be the β-graph of the DPR-network of order n. We need to show that if a β-network of order n is $(n-1)$-FT then its β-graph G' must be isomorphic to G. If G' is the β-graph of an $(n-1)$-FT β-network, it cannot contain any elementary circuits of length $n-1$ or less, because such an elementary circuit would correspond to a critical fault involving $n-1$

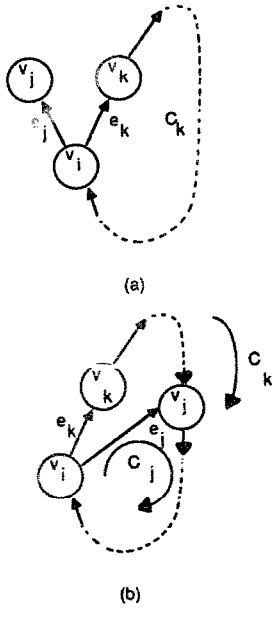

Fig. 8. (a) Elementary circuit C_k of G'; (b) Elementary circuit C_j of G'.

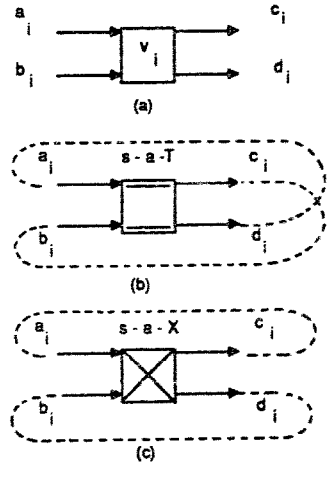

Fig. 9. (a) A fault-free β-element v_i; (b) Tolerance of s-a-T by v_i; (c) Tolerance of s-a-$\bar{X}$ by v_i.

or fewer β-elements. Hence G' must contain only elementary circuits of length n. Consequently, to prove that G' and G are isomorphic, all we need to show is that for every vertex in G', the two outgoing edges terminate at the same vertex.

Assume there exist a vertex v_i in G' whose two outgoing edges e_j and e_k terminate at vertices v_j and v_k, respectively, and $v_j \neq v_k$, as depicted in Fig. 8(a). Since G' contains only elementary circuits of length n, the edge e_k must belong to an elementary circuit C_k of length n. G' has only n vertices and the length of C_k is n, therefore v_j must be a vertex in C_k, as shown in Fig. 8(b). From Fig. 8(b) we see that e_j along with some of the edges in C_k constitute another elementary circuit C_j. Since $v_j \neq v_k$, the length of C_j must be less than n. This contradicts the fact that G' only has elementary circuits of length n. Both outgoing edges of every vertex in G' must terminate at the same vertex. Hence G' is isomorphic to G, and the DPR-network of order n is unique in achieving the maximal fault tolerance $k = n - 1$. □

4. Fault tolerance with minimal delay

In the previous section we introduced DPR-networks, and showed that they have the maximum fault tolerance $k = n - 1$. However, it is easily seen that while a DPR-network is $(n-1)$-FT, it has the worst possible communication delay $d = n$. In this section we present a complementary class of β-networks which has the minimum communication delay $d = \lceil \log_2 n \rceil + 1$, and the minimum fault tolerance $k = 1$.

In a typical fault-tolerant system, diagnostic software is run periodically, and when a faulty β-element is detected it is replaced. If the rate of diagnosis is much higher than the average rate of β-element failures, then it is reasonable to assume that only single β-element failures can occur in β-networks. In light of this single-fault assumption, single-fault tolerance can be treated as a fundamental design goal. As long as single β-element faults can be tolerated and removed before a second fault occurs, DFA can be continuously maintained with sufficiently high probability. We now present a relatively simple characterization of 1-FT β-networks. For each β-element v_i of a β-network, we denote its two inputs by a_i and b_i, and its two outputs by c_i and d_i as shown in Fig. 9(a). In order for v_i to tolerate the s-a-T fault, there must exist paths in the β-graph from edges c_i and d_i back to edges b_i and a_i, respectively, as shown in Fig. 9(b). Similarly, paths from c_i and d_i to a_i and b_i, respectively, are needed for v_i to tolerate the s-a-X fault, as shown in Fig. 9(c). The foregoing reasoning leads to the following result.

Lemma 5. *Let G be the β-graph of a β-network N. N is 1-FT if and only if for every vertex v_i, there exist four paths in G, one from each of the two outputs of v_i to each of the two inputs of v_i. These paths must not include the vertex v_i.*

Using this lemma, single-fault tolerance can be verified by checking for feedback paths from the outgoing edges of each vertex back to the incoming edges. If a β-network is not 1-FT, all its single critical faults can be identified in the same checking procedure. The identification and removal of single critical faults in a β-network is a basic step in fault-tolerant design. The characterization of Lemma 5 is straightforward, but its usefulness is limited because of the large amount of computation which may be needed to identify the feedback paths. The sufficient condition in the next lemma is more useful.

Lemma 6. *A β-network is 1-FT if its β-graph contains a Hamiltonian circuit and no self-loops.*

A β-element and the corresponding vertex in its β-graph is *critical* if one of its stuck-at states constitutes a single critical fault. Lemma 6 can be restated in the following way:

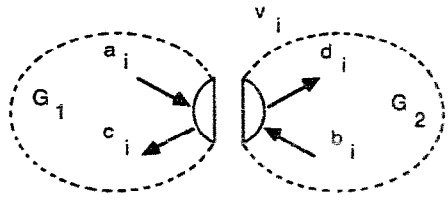

Fig. 10. A single critical fault v_i in the β-graph G.

Lemma 7. *If the β-graph G of a β-network N contains a Hamiltonian circuit, then a vertex v_i of G and the corresponding β-element are critical if and only if there exists a self-loop at v_i.*

Proof. If there is a self-loop at v_i, then it is obvious that v_i must be critical. Suppose that v_i is critical and has no self-loop. There must then exist a split of v_i which disconnects G into two components G_1 and G_2. Since G contains a Hamiltonian circuit C, one incoming and one outgoing edge of v_i must belong to C. Assume, without loss of generality, that the edges a_i and d_i defined in Fig. 10 belong to C. There must exist a path from G_2 back to G_1 not containing v_i, in order to complete the Hamiltonian circuit C. This is impossible because it implies that v_i is not critical. If instead of a_i and d_i, a_i and c_i belong to C, then d_i and b_i must constitute a self-loop. Hence if v_i is critical, then v_i must contain a self-loop. □

Lemma 7 implies that critical β-elements can be easily identified in β-networks that are known to contain a Hamiltonian circuit. Next we look for examples of 1-FT β-networks. A well-known single-stage β-network is the $N \times N$ *shuffle-exchange* network or SE-network [15]. The connecting-link pattern used resembles the perfect shuffling of a deck of cards and is called the *perfect shuffle*. If the terminals are numbered from 0 to $N-1$, then the perfect shuffle can be represented by a permutation σ which can be defined as follows:

$$\sigma(i) = (2i + \lfloor 2i/N \rfloor)_{\bmod N} \quad \text{for } i = 0, 1, \ldots, N-1.$$

The *inverse* of an $N \times N$ β-network is another $N \times N$ β-network that is the same as the original network except that the direction of all the links is reversed. The input terminals become the output terminals, and vice versa. The $N \times N$ *inverse shuffle exchange* network, or ISE-network, is the inverse of the $N \times N$ SE-network. Figure 11(a) depicts the 8×8 ISE-network, while Fig. 3(a) shows the 16×6 case. The *order* of an ISE-network is the number of β-elements in the network. Hence the order of an $2n \times 2n$ ISE-network is n. For convenience, we restrict our attention to ISE-networks of order $n = 2^m$, where m is an integer. Each β-element in an ISE-network can therefore be designated by an m-bit binary number $b_m b_{m-1} \ldots b_1$, where $b_i \in \{0, 1\}$. The top β-element is designated $00 \ldots 0$. Following the same convention, all the $2n$ links can be labeled from top to bottom by $(m+1)$-bit binary numbers $b_m b_{m-1} \ldots b_0$, starting from $00 \ldots 0$ and terminating with $11 \ldots 1$, as illustrated in Fig. 11(a). With the labeling scheme each vertex $b_m b_{m-1} \ldots b_1$ of an ISE-network has two incoming links labeled $b_m \ldots b_1 0$ and $b_m \ldots b_1 1$, and two outgoing links labeled $0 b_m \ldots b_1$ and $1 b_m \ldots b_1$. When β-element $b_m b_{m-1} \ldots b_1$ is in the T-state, connections are established from $b_m \ldots b_1 0$ to $0 b_m \ldots b_1$ and from $b_m \ldots b_1 1$ to $1 b_m \ldots b_1$. If it is in the X-state, these connections are reversed. The labels for β-elements can be translated directly into β-graphs to identify corresponding vertices. The binary $(m+1)$-tuple labels for β-network links can be used to lable edges in the β-graphs and thereby implicitly identifying the computing units. The labeled β-graph of the ISE-network of Fig. 11(a) is shown in Fig. 11(b).

It has been shown that the network structure of Pease's indirect binary m-cube [10] is isomorphic to that of the omega network [6], which is actually a cascade of m stages of the

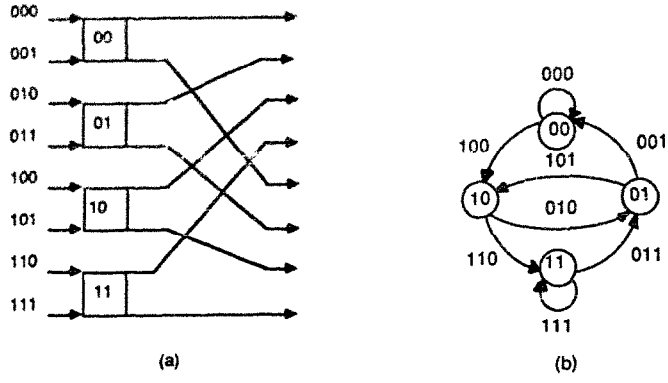

Fig. 11. (a) The ISE-network of order four; (b) Labeled β-graph of the ISE-network of order four.

ISE-network of order 2^{m-1} [9]; see Fig. 2. We know from Pease's work that the indirect binary m-cube has the *full access property*, that is, every input terminal of the network can reach any output terminal via one pass through the network. By a simple space-to-time transformation, the ith stage of the indirect binary m-cube can be mapped onto the ith pass through the ISE-network of order 2^{m-1}. Hence if an input terminal of an indirect binary m-cube can reach any one of the output terminals in m stages, then any input terminal of the ISE-network of order 2^{m-1} should be able to reach any other terminal within the distance m. The communication-delay parameter d of the ISE-network of order 2^{m-1} must therefore be m or less. In other words, for the ISE-network of order $n = 2^m$, $d \leqslant \log_2 n + 1$. Since we know that $\lfloor \log_2 n \rfloor + 1$ is the smallest possible value of d for any β-network with n β-elements (Theorem 1), the CD parameter of the ISE-network of order n must be $\log_2 n + 1$. Consequently the ISE-network of order $n = 2^m$ has the minimal communication delay among the β-networks of n β-elements. It is easy to see that the ISE-network of order n is 0-FT. Both the top and bottom β-elements contain self-loops; by Lemma 7 these self-loops correspond to single critical faults. The foregoing discussion leads to the following theorem.

Theorem 5. *The FT and CD parameters of the ISE-network of order $n = 2^m$ for some integer m, are $k = 0$ and $d = \log_2 n + 1$, respectively.*

The minimal communication delay of ISE-networks makes them very desirable for systems requiring very fast communication. In addition, they require very simple control algorithms [10]. Clearly, a serious drawback of ISE-networks is their lack of fault tolerance. Next we propose a modified ISE-network which is fault tolerant and still possesses the minimum communication delay. In order to make an ISE-network 1-FT, we must identify all its critical β-elements. We know that the β-elements labeled $00\ldots0$ and $11\ldots1$ in the β-graph have self-loops, and thus are critical; see Fig. 11(b). We first demonstrate that these are the only β-elements that are critical, and therefore need to be modified.

In the context of the mathematical treatment of shift register sequences, Good [4] and de Bruijn [2] introduced an important graph. A *Good–de Bruijn graph* of order m, G_m, is a directed graph with 2^m vertices representing the 2^m distinct binary m-tuples. A directed edge leads from vertex v_i to v_j in G_m, if the m-tuple v_j is a successor of the m-tuple v_i, i.e., v_j can be obtained from v_i by a single cyclic shift operation. For example, the m-tuple $a_1 a_2 \ldots a_m$ has two successors, $a_2 \ldots a_m 0$ and $a_2 \ldots a_m 1$ and two predecessors, $0a_1 \ldots a_{m-1}$ and $1a_1 \ldots a_{m-1}$. The

next lemma is a direct consequence of the definitions of the ISE-network of order $n = 2^m$ and the Good–de Bruijn graph of order m.

Lemma 8. *The β-graph of the ISE-network of order $n = 2^m$ is isomorphic with the Good–de Bruijn graph G_m of order m.*

A shift-register sequence of maximum length 2^m corresponds to a Hamiltonian circuit in the Good–de Bruijn graph G_m. Good has proven the existence of such maximum-length sequences [4]. Combining this fact with Lemma 8 we obtain the following results.

Lemma 9. *The β-graph of the ISE-network contains a Hamiltonian circuit.*

Lemmas 7 and 9 together confirm that the top and bottom β-elements are the only critical β-elements in the ISE-network of order n. In order to make the network 1-FT the two self-loops associated with these β-elements must be removed.

Sowrirajan and Reddy have recently investigated the design of a class of fault-tolerant β-network, called C_2-networks with the maintenance of rearrangeability as the fault-tolerance criterion [4]. They showed that by adding one redundant β-element a C_2 network can be made 1-FT with respect to rearrangeability. Employing a similar approach we can easily make the ISE-network 1-FT with respect to the DFA by adding a redundant β-element. Figure 12 illustrates a 1-FT version of the ISE-network of order four. During fault-free operation the redundant β-elements b_r is set to the T-state which makes the modified network isomorphic to the original network. When either β-element 00 or 11 is stuck-at-T, b_r can be set to the X-state to maintain DFA. With the introduction of a redundant β-element, additional hardware and delay are also introduced.

We next describe another modification method for ISE-networks which makes the networks 1-FT but uses no redundant β-element and incurs no additional delay penalty. The *modified ISE-network*, or MISE-network, of order n is an ISE-network of order n with two of its links altered as follows. The top output from β-element $00\ldots0$ is connected to link $11\ldots1$ instead of to link $00\ldots0$. Similarly, the bottom output of β-element $11\ldots1$ is connected to link $00\ldots0$ instead of to link $11\ldots1$. Basically in the MISE-network, the destinations of the two original self-loop links are exchanged. Figure 13 illustrates the MISE-network of order four obtained from Fig. 11(a). We now show that the MISE-network of order n is 1-FT and still possesses the communication delay $d = \log_2 n + 1$ of the original ISE-network.

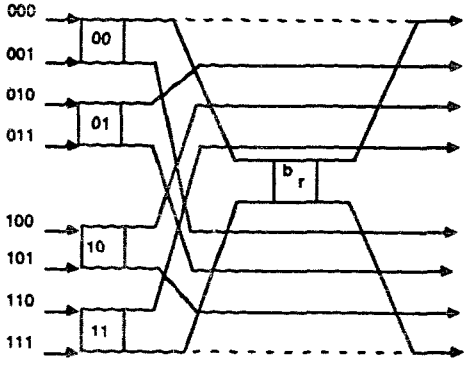

Fig. 12. The 1-FT redundant ISE-network of order four based on [14].

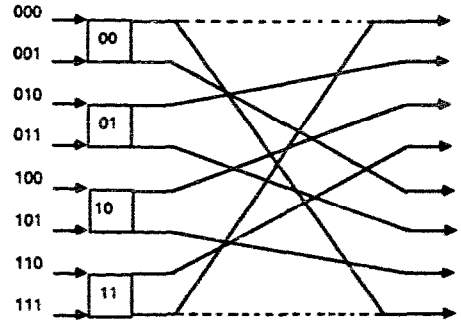

Fig. 13. The MISE-network of order four.

Lemma 10. *Every MISE-network has FT parameter* $k = 1$.

Proof. The MISE-network is obtained by deleting the two original self-loops from the ISE-network. These self-loops could not have been part of a Hamiltonian circuit in the ISE-network. Hence any such Hamiltonian circuit must still exist in the MISE-network. The MISE-network thus contains a Hamiltonian circuit and has no self-loops, therefore, by Lemma 6 is is 1-FT. □

Lemma 11. *The MISE-network of order* $n = 2^m$, *for some integer* m, *has CD parameter* $d = \log_2 n + 1$.

Proof. If G is the β-graph of an ISE-network and G' is the β-graph of the corresponding MISE-network, then we must show that the edge diameter of G' is the same as that of G. Let edges a' and b' in G' be the modified edges of the self-loops a and b in G. Let v_0 and v_{2^m-1} denote the source vertices of a' and b' respectively. Let the other four edges adjacent to v_0 and v_{2^m-1} be as denoted in Fig. 14. G' differs from G only in the edges a' and b'. We know that the delay in G is $\log_2 n + 1$. To show that the delay in G' is also $\log_2 n + 1$, all we need to show is that a' and b' can reach and be reached by all other edges in G' within the distance $\log_2 n + 1$.

Any edge in G must reach a and b via d and e, respectively. If a and b are reachable from any other edge within the distance $\log_2 n + 1$, then a' and b' must also be reachable from any other edge within the distance $\log_2 n + 1$. Edges b and a can reach any other edge of G via edges f and c, respectively, within the distance $\log_2 n + 1$. Hence a' (b') in G' must be able to do the same via edge f (c). Consequently the CD parameter of the MISE-network of order n is the same as that of the ISE-network of order n, namely $d = \log_2 n + 1$. □

Theorem 6. *The FT and CD parameters of the MISE-network of order* $n = 2^m$ *for some integer* m, *are* $k = 1$ *and* $d = \log_2 n + 1$, *respectively.*

MISE-networks are fault-tolerant β-networks with the minimal communication delay. We have thus synthesized a fault-tolerant β-network by modifying a non-fault-tolerant β-network. This was accomplished without adding extra β-elements or increasing communication delay. The simple control algorithm used for ISE-networks needs to be modified only very slightly for the MISE-networks [13].

5. Optimal networks

The networks presented in Sections 3 and 4 show that the bounds on the fault-tolerance and the communication delay parameters k and d, respectively, given by inequalities (1) and (2) of

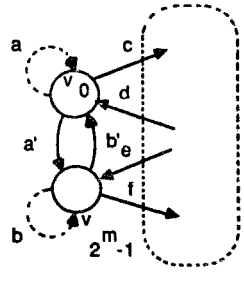

Fig. 14. A portion of the β-graph of a MISE-network.

Theorem 1 are tight. This section considers the tightness of inequality (3) of Theorem 1, which relates k and d. β-networks having $k = d - 1$ will be referred to as *optimal*, because they achieve the best fault-tolerance for a given maximum communication delay or, equivalently, the minimum value of d for a prescribed level of fault tolerance. Since the optimal networks are defined by a single relationship between k and d, many different networks fall in this class. In general, it is possible to have optimal networks with the minimum communication $d = 1$, optimal networks with the maximum $k = n - 1$, and optimal networks with intermediate values of k and d. The selection of a particular optimal network depends on the relative importance given to fault tolerance and performance. Figure 15 illustrates the feasible design space for β-networks as dictated by the three inequalities of Theorem 1. The optimal networks are those which are on the diagonal boundary.

The DPR network introduced in Section 3 is an example of an optimal network, since in this case $k = n - 1$ and $d = n$. Hence, optimal networks with maximal fault tolerance exist; moreover, it is possible to obtain an optimal network with maximal fault tolerance for every value of n. Unfortunately, this does not hold for networks with the minimal delay. In fact, excluding the trivial case of a network with a β-graph composed of two vertices, it is impossible to find an optimal network with a β-graph having 4 vertices and $d = 3$. This may be shown by exhaustively testing all possible 4-vertex β-graphs each of which is either a DPR β-graph, $d = 4$, or else a β-graph with a cycle of length 2, which implies $k \leqslant 1 < \log_2 4 \leqslant d - 1$. By applying classical graph methods [3] and the analysis techniques introduced in [12], it is possible to verify that the 8-vertex β-graph shown in Fig. 16 corresponds to an optimal network with minimum communication delay, since $k = 3$ and $d = \lfloor \log_2 8 \rfloor + 1 = 4$. It is not known whether optimal networks with minimal delay exist for values of n different from 2, 4 and 8. The foregoing results suggest that it may not be possible to find an optimal network with minimal delay for every value of n, although such networks exist for some specific values of n.

A class of optimal networks, referred to as *Reduced Doubly Twisted Torus* (RDTT) networks, is now introduced. An n-vertex RDTT-network is defined in terms of its β-graph which contains $n = rc - 1$ vertices. An ordered pair of integers (i, j), with $0 \leqslant i < r$ and $0 \leqslant j < c$, is associated with each vertex of the β-graph. The two successors of node (i, j) are conveniently

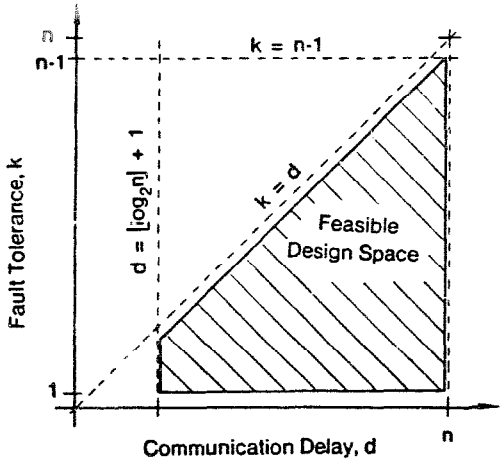

Fig. 15. Feasible design space for β-networks.

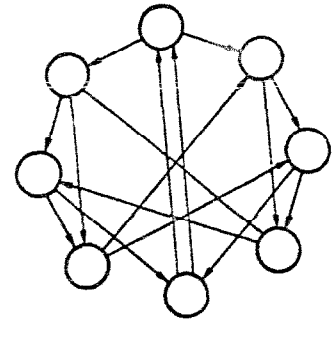

Fig. 16. β-graph of an optimal network containing 8 β-elements.

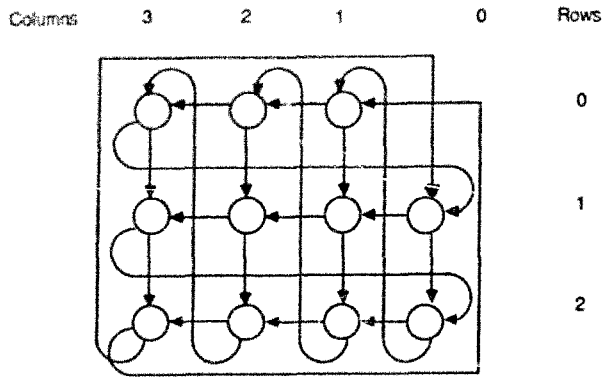

Fig. 17. β-graph of a RDTT network with 11 nodes.

defined by associating counters with i and j. Suppose that i is the current value of a modulo-r counter RC, and j is the current value of a modulo-c counter CC. The two counters are connected so that when a transition from $c-1$ to 0 takes place in the CC counter, the RC counter is incremented, and when the transition from $r-1$ to 0 takes place in the RC counter, the CC counter is incremented from j to $j+1$. Hence, when RC is incremented in the state $(r-1, c-1)$, the new state obtained is $(0, c-1)$. However, since the increment has reset RC, the CC counter needs to be incremented, leading to (0,0). Furthermore, the reset of CC causes an increment of RC, so that the final state (1,0) is reached. Similarly, if CC is incremented in the state $(r-1, c-1)$, then a symmetrical sequence of events is activated, leading to the final state (0,1). As a consequence of this behavior, the state (0,0) can never be obtained, and the corresponding nodes does not appear in the β-graph. Using this representation, one successor of vertex (i, j) is obtained by incrementing only RC, while the other successor is obtained by incrementing only CC. Figure 17 shows the β-graph of the 11-vertex RDTT network with $r=3$ and $c=4$. The β-graph is drawn as an $r \times c$ array with vertex (i, j) placed in row i and column j.

It is worth noting that the β-graph of an RDTT network is a directed Doubly Twisted Torus (DTT) [11], where the vertex in row 0 and column 1 has been replaced by two arcs connecting the vertex $(r-1, c-1)$ to the vertices (1,0) and (0,1). Although the graphs of an RDTT and a DTT network [11] have similar structures, they have quite different interpretations. In the RDTT case the graph of interest is a β-graph, in which each vertex is a β-element and each edge is either an inter-stage link (unlabelled edge) or a processor (labelled edge). In the DDT graph, each vertex represents a processor and the edges are the interprocessor links.

Theorem 7. *An $(rc-1)$-vertex RDTT network has CD parameter $d = r + c - 2$.*

Proof. The length of a path between an edge entering the vertex (a, b) and an edge leaving the vertex (f, e) is one greater than the length of a path between the vertices (a, b) and (f, e). Considering the correspondence with the counters RC and CC introduced earlier, the following four cases are possible:

(1) $f \geqslant a$, $e \geqslant b$: the minimum number of increments for obtaining (f, e) from (a, b) is $(f-a)+(e-b) \leqslant r+c-2$, because $(a, b) = (0,0)$ is not allowed;

(2) $f < a$, $e > b$: the minimum number of increments is $r-(a-f)+(e-b-1) \leqslant r+c-2$;

(3) $f > a$, $e < b$: this case is similar to (2), hence the minimum distance is $c-(e-b)+(f-a-1) \leqslant r+c-2$;

(4) $f \leqslant a$, $e \leqslant b$ (excluding the case $f = a$ and $e = b$ already considered in (1) the minimal length of the path is $r - (a - f) + c - (b - e) - 2 \leqslant r + c - 2$. □

Corollary 1. *The minimum length of an elementary circuit in an RDTT β-graph is* $r + c - 2$.

Proof. An elementary circuit is a closed path containing (a, b) and (f, e) such that $(a, b) = (i, j)$ and $(f, e) = (i, j - 1)$ or $(f, e) = (i - 1, j)$. Hence, from case (4) in the proof of Theorem 7, it follows that the minimum length of an elementary circuit is equal to the maximum communication delay. □

Theorem 7 gives the value of the CD parameter of a RDTT network. In order to prove that these networks are optimal, it is necessary to show that the FT parameter is $k = r + c - 3$. First, it should be noted that it is possible to redraw the β-graph of a RDTT by applying a suitable number of row and column rotations. A row rotation is shown in Fig. 18, where the unbracketed pairs indicate the original positions of the vertices, while the bracketed pairs indicate the final positions after the row rotation. The column rotation is obtained by exchanging the operations performed on the row and column indices. Hence. a vertex (i, j) may be moved to the bottom left corner of the graph by applying $r - 1 - i$ row rotations and $c - 1 - j$ column rotations. It turns out that the properties valid for the vertex in the bottom left corner and its associated edges are also valid for other vertices and their edges.

Lemma 12. *Let* C_1 *be a circuit which includes the bottom left vertex in an RDTT β-graph. There exists another circuit* C_2, *which has at least* $r + c - 2$ *common nodes and no common edge with* C_1. *One of the common vertices should be the bottom left vertex.*

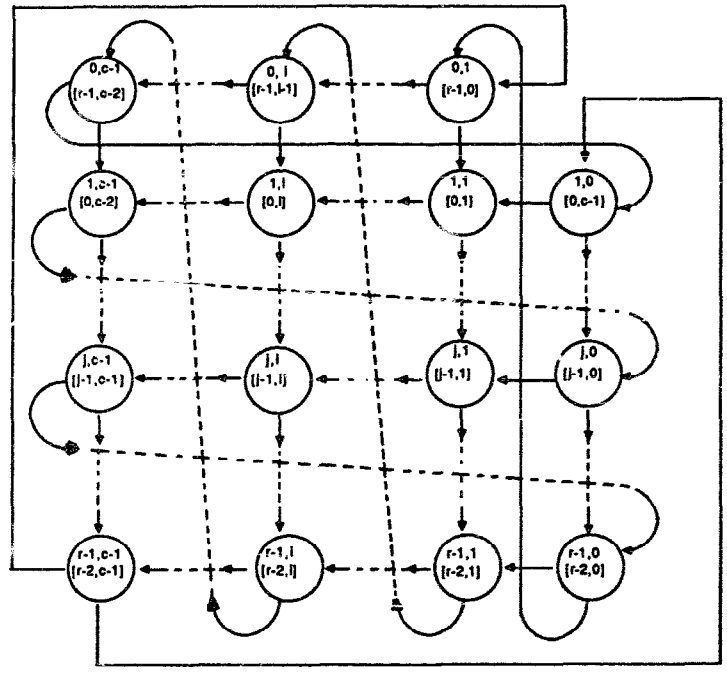

Fig. 18. Effect of the row rotation on the β-graph of an RDTT network

Proof. It is possible to use either of the edges leaving any given vertex of the β-graph (i, j) to construct a path from (i, j) to the vertex in the bottom left corner, it turns out that the same property is valid for every pair of vertices. Hence, the constraint imposed by the condition that C_1 and C_2 should have common vertices and no common edge does not prevent the finding of the required path. The number of common vertices is limited by the minimum length of a cycle, hence it is at least $r + c - 2$. □

Theorem 8. *An $(rc - 1)$-vertex RDTT network has FT parameter $k = r + c - 3$.*

Proof. Given $r + c - 3$ faults, it is always possible to find a circuit which is compatible with the faults and includes the edge A leaving the vertex in the bottom left corner. Lemma 12 assures that another circuit exists, such that it includes the other edge B leaving the vertex in the bottom left corner and it has $r + c - 2$ common vertices and no edge in common with the first one. It can be shown that, given $r + c - 3$ faulty vertices, A is reachable from B and vice versa. In fact, since the two circuits including A and B have at least $r + c - 2$ common vertices, and since there are only $r + c - 3$ faulty vertices, at least one common vertex must not be faulty. Hence, the edges in the circuit including A can be reached from the edges in the circuit including B via the fault-free vertex. This property is valid for every vertex and every pair of edges leaving the same vertex. This is sufficient to conclude that the DFA property holds, since it is possible from any edge to reach both edges leaving the successor vertex, even when the later is faulty. Applying this property recursively, it may be shown that every edge is reachable. This leads to the conclusion that $k \geqslant r + c - 3$. Equation (3) of Theorem 1 requires $k \leqslant d - 1 = r + c - 3$, hence $k = r + c - 3$. □

The number n of β-elements in a RDTT network is $rc - 1$, while r and c are integers. In general, several values for r and c may exist for a given value of n. Since $d = r + c - 2$ and $n = rc - 1$, it is possible to prove that the minimum value of d is achieved for $r = c = (n + 1)^{1/2}$, provided that $(n + 1)^{1/2}$ is an integer. The maximum value of d occurs when $r = 1$ and $c = n + 1$, or when $c = 1$ and $r = n + 1$ In the latter case, the RDTT network is also a DPR network. Hence, the RDTT networks can be considered a superset of the DPR networks.

6. Conclusion

In this paper, graph-theoretic techniques are successfully applied to the analysis of the performance and fault tolerance of β-networks. Theoretical bounds for fault tolerance and communication delay are characterized. Several classes of β-networks are analyzed and the feasible design space of β-networks are explored. A class of β-networks exhibiting optimal balance between fault tolerance and communication delay are introduced.

While the FT and CD parameters do provide rudimentary characterizations of the fault tolerance and performance of a β-network, more refined parameters are needed for practical design procedures. We believe this paper provides the foundation for developing an intelligent design procedure for high performance and fault tolerant β-networks.

References

[1] T. Aardenne-Ehrenfest and N.G. de Bruijn, Circuits and trees in oriented linear graphs, *Simon Stevin* **28** (1951) 203–217.

[2] N.G. de Bruijn, A combinatorial problem, *Nederl. Akad. Wetensch. Proc.* **49** (1946) 758–764.

[3] N. Deo, *Graph Theory With Application To Engineering And Computer Science* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[4] I.J. Good, Normal recurring decimals, *J. London Math. Soc.* **21** (1946) 167–169.
[5] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).
[6] D.H. Lawrie, Access and alignment of data in an array processor, *IEEE Trans. Comput.* **24** (1975) 1145–1155.
[7] K.N. Levitt, M.W. Green and J. Goldberg, A study of the data communication problems in a self-repairable multiprocessor, *Proc. AFIPS Conference* **32** (1968) 515–527.
[8] S.F. Lundstrom and G.H. Barnes, A controllable MIMD architecture, *Proc. Parallel Processing Conference* (1980) 19–27.
[9] D.S. Parker, Jr., Notes on shuffle/exchange-type switching networks, *IEEE Trans. Comput.* **29** (1980) 213–222.
[10] M.C. Pease, The indirect binary n-cube microprocessor array, *IEEE Trans. Comput.* **26** (1977) 458–473.
[11] C.H. Sequin, Doubly twisted networks for VLSI processor arrays, *Proc. 8th Annual Symposium on Computer Architecture* (1981) 417–480.
[12] J.P. Shen and J.P. Hayes, Fault tolerance of a class of connecting networks, *Proc. 7th Annual Symposium on Computer Architecture* (1980) 61–71.
[13] J.P. Shen and J.P. Hayes, Fault-tolerance of dynamic-full-access interconnection networks, *IEEE Trans. Comput.* **33** (1984).
[14] S. Sowrirajan and S.M. Reddy, A design for fault-tolerant full connection networks, *Proc. Conference on Information Sciences and Systems*, Princeton (1980).
[15] H.S. Stone, Parallel processing with the perfect shuffle, *IEEE Trans. Comput.* **20** (1971) 153–161.