Parallel Algorithms on the Rotation-Exchange Network – A Trivalent Variant of the Star Graph

Chi-Hsiang Yeh and Emmanouel A. Varvarigos
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106-9560, USA

Abstract

We investigate a trivalent Cayley graph, which we call the rotation-exchange (RE) network, and present communication algorithms to perform one-to-one routing, single-node broadcasting, multinode broadcasting, and total exchange in it. The RE network can be viewed as a star-graph counterpart to the hypercubic shuffle-exchange network, with the important difference that the RE network is regular and symmetric. We show that RE networks can efficiently embed and emulate star graphs, meshes, hypercubes, cube connected cycles (CCC), pancake graphs, bubble-sort graphs, complete transposition graphs, and the shuffle-exchange permutation graphs. We also show that the performance of RE networks can be significantly improved for a variety of applications if the transmission rate of on-chip links is considerably higher than that of off-chip links.

1. Introduction

A variety of topologies have been proposed and analyzed in the literature [2, 16, 23, 25, 29, 33] for the interconnection of processors in parallel computing systems, under several assumptions on the communication model used. Among them, the star graph [2, 3] has received a great deal of attention as an attractive alternative to the hypercube for building parallel computers. Star graphs belong to the class of Cayley graphs [3], are symmetric and strongly hierarchical, and have diameter, average distance, and node degree that are superior to those of similar-sized hypercubes. Also, many important algorithms can be efficiently performed on the star graph [4, 6, 7, 10, 22, 24].

Even though the hypercube and the star graph have many desirable topological and algorithmic properties, their node degrees increase with the size of the network. Several constant-degree networks, such as the cube connected cycles (CCC) [23], the shuffle-exchange (SE) networks, the de Bruijn graphs [20], the star connected cycles (SCC) [17], the shuffle-exchange permutation (SEP) graphs [18], and the cyclic networks [30], have been proposed as alternatives to the hypercube and the star graph topologies. Since the SCC graph inherits some important properties from the star graph, and the star graph has been shown to be superior to the hypercube in several aspects, the SCC graph has some important advantages over the CCC network under certain assumptions [17]. The well-known shuffle-exchange (SE) network, which is another hypercubic network, has diameter that is somewhat smaller than that of a similar-sized CCC, and can emulate a hypercube of the same size with simpler and faster algorithms than a CCC [20]. The SE network, however, is neither symmetric nor regular.

The trivalent Cayley graph to be studied in this paper can be viewed as a star-graph counterpart to the hypercubic SE network, and will be referred to as the *rotation-exchange (RE) network*. The RE network first appeared as an example of group graphs in [1], but its topological and algorithmic properties have not been explored in the literature before. We show that, as is the case with the SCC graph, the RE network inherits many desirable properties from the star graph, and is therefore in many respects superior to the CCC and SE networks under certain assumptions. Since the relationship between the RE network and the star graph is similar to that between the SE network and the hypercube, the RE network can embed and emulate a star graph of the same size as well as a variety of other network topologies with faster and considerably simpler algorithms than the corresponding embeddings and emulation for an SCC graph. In contrast to the SE network, the RE network is both regular and vertex-symmetric.

We present efficient algorithms to perform one-to-one routing, single-node broadcasting, multinode broadcasting, and total exchange in RE networks. We also derive simple and efficient embeddings and emulation of star graphs [2, 3], meshes, hypercubes, CCC [23], pancake graphs [3], bubble-sort graphs [3], complete transposition graphs [19, 20], and shuffle-exchange permutation graphs [18], under a variety of assumptions on the communication model.

We assume that several processors of the RE network are placed on the same module (e.g., chip, board, wafer, or multi-chip module (MCM)) and look at the case where

the transmission rate of on-module links is different (larger) than the transmission rate of off-module links. We find the time required to perform (unicast) routing, single-node broadcasting, and total exchange when on-module transmission rates are large enough so that the off-module bandwidth is the main communication bottleneck. We show that when the transmission rate of on-module links is considerably higher than that of off-module links, the performance of RE networks can also be significantly improved for a variety of other applications, including the embeddings and emulation of star graphs [2, 3], meshes, hypercubes, pancake graphs [3], and complete transposition graphs [19, 20].

The remainder of this paper is organized as follows. In Section 2, we formally define the rotation-exchange network and give some related notation. In Section 3, we derive a variety of embeddings and emulation algorithms for RE networks. In Section 4, we consider RE networks that have fast on-module links. In Section 5, we present algorithms to execute several prototype communication tasks in RE networks. Finally, in Section 6, we conclude the paper.

2. Rotation-exchange (RE) networks

The rotation-exchange (RE) network was first mentioned in [1] as an example of a Cayley graph, but it has not been investigated in detail. In this section, we introduce the definition of the RE network and some related notation.

A permutation of k distinct symbols in the set $\{1,2,...,k\}$ is represented by $U = u_{1:k} = u_1u_2\cdots u_k$, where $u_i \in \{1,2,...,k\}$ and $u_i \neq u_j$ for $i \neq j$, $1 \leq i,j \leq k$. A k-dimensional RE network is an undirected regular graph with $N = k!$ nodes, each corresponding to a distinct permutation of the set $\{1,2,...,k\}$. Two nodes are directly connected if and only if the label (permutation) of one node can be obtained from the label of the other by one of the following operations:

- Swapping the first two symbols (the leftmost symbol is ranked as first).
- Shifting the last $k-1$ symbols cyclically to the left (or right) by one position.

A 4-RE network is shown in Fig. 1. The following two types of generators will be useful in formally describing the RE network topology.

Definition 2.1 (Transposition Generator T_i) :
Given a permutation $U = u_{1:k}$, we define the *dimension-i transposition generator* T_i, $i = 2,3,...,k$, as the permutation that interchanges symbol u_i with symbol u_1 in $u_{1:k}$.

In other words, for $i = 2,3,...,k$,

$$T_i(u_{1:k}) = u_iu_{2:i-1}u_1u_{i+1:k},$$

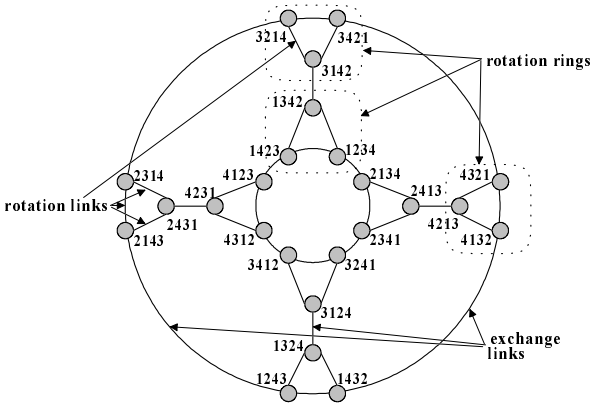

Figure 1. The structure of a 4-RE network.

where the notation $u_{j_1:j_2}$, $j_1 \leq j_2$, denotes the sequence $u_{j_1}u_{j_1+1}\cdots u_{j_2}$. For example, for the permutation $I =$ 123456789 we have

$$T_2(I) = 213456789;\ T_5(I) = 523416789;\ T_8(I) = 823456719.$$

Definition 2.2 ($(k-1)$-Cycle Generator C^i) :
Given a permutation $U = u_{1:k}$, we define the $(k-1)$-*cycle generator* C^i as the permutation that cyclically shifts the sequence of symbols $u_{2:k}$ by i positions to the left.

That is,

$$C^i(u_{1:k}) = u_1u_{i+2:k}u_{2:i+1}.$$

For example, for $I = 1$ 23456789, we have

$$C^1(I) = 1\ 34567892;\ C^2(I) = 1\ 45678923.$$

For any integer i, C^i is equivalent to the generator $C^{i \bmod k-1}$, which is equivalent to the sequence of generators $\underbrace{C^1C^1\cdots C^1}_{i \bmod k-1}$, where k is the number of symbols in the permutation. It can be seen that the sequence of generators, $C^iT_2C^{-i}$, which stands for the chain function

$$UC^iT_2C^{-i} = C^{-i}(T_2(C^i(U))),$$

is equivalent to the transposition generator T_{i+2} for $i = 1,2,3,...,k-2$.

The *k-dimensional rotation-exchange network,* abbreviated k-RE, is a trivalent Cayley graph that has $k!$ nodes, each represented by a permutation of the symbols in $\{1,2,...,k\}$, and is defined as follows.

Definition 2.3 (Rotation-Exchange (RE) Networks) :
The k-dimensional rotation-exchange network (abbreviated k-RE) is the graph (V,E), where

$$V = \{u_{1:k} | u_i,u_j \in \{1,2,...,k\},\ u_i \neq u_j \text{ for } i \neq j,\ 1 \leq i,j \leq k\}$$

is the set of vertices, and

$$E = \{(U,V) | U,V \in V \text{ satisfying } U = T_2(V) \text{ or } U = C^1(V) \text{ or } U = C^{k-2}(V)\}$$

is the set of edges.

A k-RE is a $k!$-node Cayley graph based on the generator set $\{T_2,\ C^1,\ C^{k-2}\}$. The RE network is vertex-symmetric and regular and has degree equal to 3. In this paper, the integer "k" is exclusively used to represent the number of symbols in the permutation representing a node. We will sometimes use T, C, and C^{-1} to signify T_2, C^1, and C^{k-2}, respectively. The link connecting nodes U and $G(U)$ is called the *link* G of node U, where generator $G \in \{T, C, C^{-1}\}$. Note that links C and links C^{-1} correspond to a left or a right cyclic shift of the node label, respectively. Links C and C^{-1} will be collectively referred to as the *rotation links*, while link T will also be referred to as the *exchange link* of a node. By removing all exchange links from a $k!$-node k-RE, we obtain $k \cdot (k-2)!$ disconnected $(k-1)$-node rings, each of which will be called a *rotation ring*.

In [18], Latifi and Srimani proposed an interesting degree-3 Cayley graph, called the *shuffle-exchange permutation (SEP) graph*, which has generators similar to those of an RE network. Instead of using $(k-1)$-cycle generators, the SEP graph uses *k-cycle generators*, which shift the k symbols cyclically to the left or right by one position. As we will show later, the RE network is more efficient in emulating several graphs based on permutation groups. Also, if we assume that local links (or on-module links) are faster than remote links (or off-module links) by a factor of $\Theta(k)$, the performance of the RE network for many problems is better than that of the SEP graph by a factor of $\Theta(k)$.

3. Emulation and embeddings in rotation-exchange networks

In this section, we show how to emulate algorithms developed for several Cayley graphs on an RE network and derive efficient embeddings of several important topologies in RE networks.

3.1. Emulating Cayley graphs in RE networks

In this subection, we show how to emulate algorithms developed for star, bubble-sort, complete transposition, and pancake graphs as well as any Cayley graph in RE networks. In the embedding that we propose, a node in the guest Cayley graph is one-to-one mapped to the node that has the same address in the host k-RE network.

We first assume the *single-dimension communication* (SDC) model [31, 32], where the nodes are allowed to use only links of the same dimension at any given time. Many algorithms developed for the star graph fall into this category [22].

Theorem 3.1 *Any step of an algorithm in a k-star under the SDC model can be emulated on a k-RE in at most $2\lfloor (k-1)/2 \rfloor + 1$ steps under the SDC model, out of which 1 step involves transmissions over exchange links and at most $2\lfloor (k-1)/2 \rfloor$ steps involve transmissions over rotation links.*

Proof: Transmission on link T_i, $i = 2,3,4,\ldots,k$, in a k-star is equivalent to the sequence of generators C^{i-2}, T, C^{2-i} and can be emulated by transmission on either the sequence of links

$$\underbrace{CC\cdots C}_{i-2} T \underbrace{C^{-1}C^{-1}\cdots C^{-1}}_{i-2}$$

or the sequence of links

$$\underbrace{C^{-1}C^{-1}\cdots C^{-1}}_{k-i+1} T \underbrace{CC\cdots C}_{k-i+1}$$

on a k-RE. Since

$$\min(i-2, k-i+1) \leq ((i-2)+(k-i+1))/2 = (k-1)/2,$$

any step of a k-star algorithm under the SDC model can be emulated on a k-RE using 1 step involving transmissions over exchange links and at most $2\lfloor (k-1)/2 \rfloor$ steps involving transmissions over rotation links. □

By emulating routing algorithms developed for star graphs, we can obtain routing algorithms that requires at most $\lfloor 3k/2 - 3 \rfloor$ transmissions over exchange links and $O(k^2)$ transmissions over rotation links.

It is well known that *normal hypercube algorithms* can be emulated with constant slowdown on several hypercubic networks, such as the shuffle-exchange network [20] and cube-connected cycles (CCC) [23]. We now show that normal star-graph algorithms, where the dimensions of the star-graph links used are (cyclically) consecutive, can be emulated with a slowdown factor of 2 on RE networks.

Lemma 3.2 *An algorithm in a k-star that uses links of (cyclically) consecutive dimensions in s consecutive steps can be emulated on a k-RE network in at most $2s - 1 + 2\lfloor (k-1)/2 \rfloor$ steps, out of which s steps corresponds to transmissions over exchange links and at most $s - 1 + 2\lfloor (k-1)/2 \rfloor$ correspond to transmissions over rotation links.*

Proof: From the proof of Theorem 3.1, we know that the sequence of generators $T_j T_{j+1}$ in a star graph is equivalent to the sequence of generators

$$C^{j-2}, T, C^{2-j}, C^{j-1}, T, C^{1-j}$$

Clearly, the third and fourth generators in the last sequence are collectively equivalent to a single generator $C^{(2-j)+(j-1)} = C^1$. As a result, transmissions over dimensions $d_1, d_2, d_3, \ldots, d_s$ in a star graph can be emulated by the action of generators

$$C^{d_1-2}, \underbrace{T, C^{i_2}, T, C^{i_3}, \cdots, T, C^{i_{s-1}}}_{s-1}, T, C^{2-d_s},$$

where

$$i_j = \begin{cases} 1, & \text{if } d_j = d_{j-1}+1 \text{ or } (d_{j-1} = k \text{ and } d_j = 2), \\ -1, & \text{otherwise.} \end{cases}$$

Since either C^{d_1-2} or C^{2-d_s} require at most $\lfloor (k-1)/2 \rfloor$ rotation-link transmissions, the upper bound follows. □

Theorem 3.3 *Any normal k-star algorithm can be emulated on a k-RE with a slowdown factor of 2, assuming that the final locations of the outputs are not specified.*

Proof: When the locations of final destinations are not specified, the last $\lfloor (k-1)/2 \rfloor$ transmissions of Lemma 3.2, which take place over rotation links are not required. If the first step in the emulated normal star-graph algorithm requires transmissions over dimension j, we can rename the nodes of the k-RE by mapping node X to node $C^{j-2}(X)$ so that the first $\lfloor (k-1)/2 \rfloor$ rotation-link transmissions are no longer required. The total number of steps required for emulating s steps in a normal k-star algorithm is then equal to $2s-1$. □

Similar to Theorem 3.3, we can show that any normal k-star algorithm can be emulated on a k-dimensional SEP graph with a slowdown factor of 3, assuming that the final locations of the outputs are not specified.

To emulate star graph algorithms with all-port communication, we simply perform single-dimension emulation for all dimensions $T_i(X)$, $i = 2,3,\ldots,k-1$, either one after the other or simultaneously with appropriate scheduling, leading to the following theorem.

Theorem 3.4 *Any step of an algorithm in a k-star under the all-port communication model can be emulated on a k-RE in at most $\lfloor (k-1)^2/2 \rfloor + k - 1$ steps under the SDC model, out of which $k-1$ steps involve transmissions over exchange links and $\lfloor (k-1)^2/2 \rfloor$ steps involve transmissions over rotation links.*

Theorem 3.5 *Any step of an algorithm in a k-dimensional bubble-sort graph under the SDC model can be emulated on a k-RE in at most $2\lfloor k/2 \rfloor + 3$ steps under the SDC model, out of which at most 3 steps involve transmissions over exchange links and at most $2\lfloor k/2 \rfloor$ steps involve transmissions over rotation links.*

Proof: A k-dimensional bubble-sort graph has $k-1$ generators $(i, i+1)$, $i = 1,2,3,\ldots,k-1$, which exchange the i^{th} and $(i+1)^{th}$ symbols. Generator $(i,i+1)$, $i = 2,3,4,\ldots,k$, is equivalent to the sequences of generators C^{i-2},T,C,T, C^{-1},T,C^{2-i} or $C^{i-1},T,C^{-1},T,C,T,C^{1-i}$. There exists at least a sequence of generators consisting of at most 3 exchange generators T and at most $2\lfloor k/2 \rfloor$ $(k-1)$-cycle generators C or C^{-1} that is equivalent to the preceding sequence of generators by translating C^j into C or C^{-1}. □

A k-dimensional complete transposition graph k-CT [19, 20] is a Cayley graph defined with a generator set consisting of all the generators that interchange any two of the k symbols in the label of a node. A k-CT graph has $k!$ nodes, degree $k(k-1)/2$, and diameter $k-1$. It contains a k-star or a k-dimensional bubble-sort graph [3] as a subgraph and has been shown to be a rich topology that can efficiently embed many other popular topologies, including hypercubes, meshes, and trees. The following theorem provides efficient embedding of complete transposition graphs in RE networks.

Theorem 3.6 *Any step of an algorithm in a k-dimensional complete transposition graph under the SDC model can be emulated on a k-RE in at most*

$$2\left\lceil \frac{k-\lfloor (k-1)/2 \rfloor}{2} \right\rceil + 2\left\lfloor \frac{k-1}{2} \right\rfloor + 3 \approx 1.5k+3$$

steps under the SDC model, out of which 3 steps involve transmissions over exchange links and at most

$$2\left\lceil \frac{k-\lfloor (k-1)/2 \rfloor}{2} \right\rceil + 2\left\lfloor \frac{k-1}{2} \right\rfloor \approx 1.5k$$

steps involve transmissions over rotation links.

Proof: Generator (i,j), which swaps the i^{th} and the j^{th} symbols, $1 \le i,j \le k$, is equivalent to the sequence of generators

$$C^{i-2},T,C^{j-i},T,C^{i-j},T,C^{2-i}$$

or

$$C^{j-2},T,C^{i-j},T,C^{j-i},T,C^{2-j},$$

so the result follows. □

Theorem 3.7 *Any step of an algorithm in a k-dimensional pancake graph under the SDC model can be emulated on a k-RE in at most $n^2/4 + O(n)$ steps under the SDC model, out of which at most $3\lfloor k/2 \rfloor - 2$ steps involves transmissions over exchange links.*

We can generalize these emulation results to any Cayley graph defined with at most k symbols with slowdown factor upper bounded by the corresponding routing distance in an RE network (between two neighboring nodes in the guest Cayley graph).

3.2. Embeddings of trees, meshes, hyprcubes, and CCC

In this subsection, we present efficient embeddings and packings of trees, meshes, hypercubes, and cube connected cycles (CCC) [23] in RE networks.

A variety of embedding results are available for star graphs, bubble-sort graphs, and complete transposition graphs [9, 19, 21, 14]. These results, when combined with Theorems 3.1, 3.5, and 3.6 give rise to a variety of efficient embeddings for RE networks.

Theorem 3.8 *There exists a dilation-$(2\lfloor(k-1)/2\rfloor+1)$ embedding of the complete binary tree of height $2k-5$ into a k-RE network for $k=5$ or 6, or height at least equal to $(1/2+o(1))k\log_2 k$ into a k-RE network for $k\geq 7$, where an embedded edge consists of one exchange link and at most $2\lfloor(k-1)/2\rfloor$ rotation links.*

Proof: In [9], it has been shown that for $k=5$ or 6 there exists a dilation-1 embedding of the complete binary tree of height $2k-5$ into the k-star. For $k\geq 7$, there exists a dilation-1 embedding of the complete binary tree of height at least equal to $(1/2+o(1))k\log_2 k$ into the k-star. The rest of the proof follows from Theorem 3.1. □

Theorem 3.9 *There exists a dilation-$O(k)$ embedding of the d-dimensional hypercube into a k-RE network, assuming $d\leq k\log_2 k-\frac{3k}{2}+o(k)$, where an embedded edge consists of $O(1)$ exchange links and $O(k)$ rotation links.*

Proof: In [21], it has been shown that there exists a dilation-$O(1)$ embedding of the d-dimensional hypercube into a k-star, provided that $d\leq k\log_2 k-(3/2+o(1))k$. This, combined with Theorem 3.1, completes the proof. □

Theorem 3.10 *There exists an embedding of load 1, expansion 1, and dilation*

$$2\left\lceil\frac{k-\lfloor(k-1)/2\rfloor}{2}\right\rceil+2\left\lfloor\frac{k-1}{2}\right\rfloor+3\approx 1.5k+3$$

of the $M_1\times M_2$ mesh onto a k-RE network, where an embedded edge consists of at most 3 exchange link and at most $2\lceil\frac{k-\lfloor(k-1)/2\rfloor}{2}\rceil+2\lfloor\frac{k-1}{2}\rfloor\approx 1.5k$ rotation links and $M_1\times M_2=k!$.

Proof: It follows from Theorem 3.6 and the fact that there exists a dilation-1 expansion-1 embedding of $M_1\times M_2$ mesh into a k-CT graph, where $M_1\times M_2=k!$ [19]. □

Theorem 3.11 *There exists a load-1, expansion-1, and dilation-$(6\lfloor(k-1)/2\rfloor+3)$ embedding of the $2\times 3\times 4\times\cdots\times(k-1)\times k$ mesh into a k-RE network, where an embedded edge consists of at most 3 exchange link and at most $6\lfloor(k-1)/2\rfloor$ rotation links.*

Proof: In [14] it has been shown that there exists a dilation-3 expansion-1 embedding of the $2\times 3\times 4\times\cdots\times(k-1)\times k$ mesh into a k-star. This, combined with Theorem 3.1, completes the proof. □

The minimum possible dilation for a constant-degree network to embed a network of degree $\Omega\left(\frac{\log N}{\log\log N}\right)$ is $\Omega(\log\log N)$. It is interesting to note that by implementing the rotation links with faster local links (see Section 4), the delay for emulating an edge of several degree-$\Omega\left(\frac{\log N}{\log\log N}\right)$ guest graphs, such as the hypercube, k-D mesh, star graph, and complete transposition graph, in an RE network is only a small constant (which has similar effect as having a constant-dilation embedding).

Theorem 3.12 *A k-RE network can pack $\frac{(k-1)!}{2^{k/2-1}}$ k-dimensional CCC with load 1, expansion 1, dilation 4, and congestion 5 when k is even, or $\frac{k(k-2)!}{2^{k/2-1.5}}$ k-dimensional CCC with load 1, expansion 1, dilation 5, and congestion 3 when k is odd.*

Proof: When k is even, a $\frac{k}{2}$-dimensional CCC can be defined as a Cayley graph that has generators T, S^2, and S^{-2}, where $S^2=SS$ and $S^{-2}=S^{-1}S^{-1}$. Since $S^2=TCTC$ and $S^{-2}=C^{-1}TC^{-1}T$, the packings of $\frac{(k-1)!}{2^{k/2-1}}$ $\frac{k}{2}$-dimensional CCC with even k follows.

When k is odd, a $\frac{k-1}{2}$-dimensional CCC can be defined as a Cayley graph that has generators $(2,3), C^2$, and C^{-2} where $(2,3)$ swaps the 2nd and 3rd symbols of a node label and $(2,3)=TCTC^{-1}T$. Therefore, the packings with odd k follows. □

Theorem 3.13 *Any algorithm in a k-dimensional SEP graph (or RE network) under the SDC model can be emulated on a k-dimensional RE network (or SEP graph, respectively) under the SDC model with a slowdown factor of 2.*

Proof: This can be shown by noting that the generator S (or S^{-1}) of the SEP graph is equivalent to the sequence of generators TC (or TC^{-1}, respectively) of the RE network, and the generator C (or C^{-1}) of the RE network is equivalent to the sequence of generators TS (or $S^{-1}T$, respectively) of the SEP graph. □

We can also show that the computation powers of the SEP graph and RE network are equivalent within a small constant under a variety of communication models. More details will be reported in the near future.

4. RE networks using links of different transmission rate

With the rapid advances in VLSI technology, the number of transistors per chip and the number of processors that can be put onto a chip are expected to grow significantly. Since the processor-memory bandwidth is one of the major bottlenecks on the performance of current and future parallel systems, implementing processors in memory (PIM) or

computational RAM [34] is believed to be a promising approach for the construction of future massively parallel computers. EXECUBE [15], hypernets [12], recursively connected complete (RCC) networks [11], and macro-star networks [28, 32], are some of the well-known parallel architectures and networks that use such structures or similar assumptions. In what follows, we consider the case where several nodes (including processors, routers, and their memory banks) of the RE network are implemented on a single chip (or a board, a multi-chip or multi-board module).

To find a good strategy for partitioning the nodes of an RE network into chips (or, in general, modules), note that the number of transmissions over exchange links is considerably smaller than the number of transmissions over rotation links for most of the important algorithms described in Section 3 (and also for the algorithms that will be describe in Section 5). Therefore, if we place nodes belonging to the same rotation ring (and their rotation links) onto the same chip, the traffic will be largely confined within chips when executing these algorithms, and the off-chip traffic will be small. Also, since each node on an RE network has one exchange link and two 2 rotation links, by putting all nodes of a rotation ring onto the same chip, only one off-chip link will be required per node.

In most papers on routing in interconnection networks, it is assumed that the transmission and propagation delay is the same (equal to 1 unit of time) for all network links. Since, however, on-chip links are significantly shorter than off-chip links and do not need extra delay to drive off-chip pins, they can be implemented using a considerably higher clock rate. Moreover, since the cost for an on-chip connection is much smaller than that of an off-chip connection, the channel width of an on-chip link can be increased, if required, without significantly increasing the hardware cost. Thus, a realistic model for message-passing parallel architectures is to assume that on-chip connections have (considerably) larger transmission rate than off-chip connections. A similar model that uses unequal transmission rates for on-chip and off-chip links was assumed in [5, 6, 17] for SCC graphs. This model is also implied in several other papers concerning interconnection networks [11, 12, 32].

By increasing the transmission rate of on-module links, the delay required for node-to-node communication can be significantly reduced.

Theorem 4.1 *Packet routing can be performed in an N-node RE network in $O(\log N/\log\log N)$ time if the transmission rate of on-module links is $\Omega(\log N/\log\log N)$, the transmission rate of off-module links is $\Omega(1)$, and nodes belonging to the same rotation ring are placed on the same module.*

It is also easy to show the following emulation results.

Corollary 4.2 *A k-star, k-dimensional bubble-sort, or complete transposition graph can be emulated in a k-RE network under the SDC model with $O(1)$ slowdown if the transmission rate of on-module links is $\Omega(k)$, the transmission rate of off-module links is $\Omega(1)$, and nodes belonging to the same rotation ring are placed on the same module.*

Corollary 4.3 *Transmission over an embedded edge of*

- *(a) a complete binary tree of height $2k-5$ for $k=5$ or 6,*
- *(b) a complete binary tree of height at least equal to $(1/2+o(1))k\log_2 k$ for $k\geq 7$,*
- *(c) a hypercube of dimension $d\leq k\log_2 k-\frac{3k}{2}+o(k)$,*
- *(d) an $M_1\times M_2$ mesh with $M_1\times M_2=k!$, or*
- *(e) a $2\times 3\times 4\times\cdots\times(k-1)\times k$ mesh*

in a k-RE network can be performed in $O(1)$ time if the transmission rate of on-module links is $\Omega(k)$, the transmission rate of off-module links is $\Omega(1)$, and nodes belonging to the same rotation ring are placed on the same module.

These results show that many important topologies can be emulated by an RE network with constant slowdown under a communication model that takes into account the large difference between the speed of on-module and off-module links.

5. Communication algorithms for RE networks

In this section, we present algorithms to execute certain communication tasks in RE networks.

Two prototype communication tasks that arise often in applications are the multinode broadcast (MNB) and the total exchange (TE) [8, 13, 26, 27]. In the MNB each node has to broadcast a packet to all the other nodes of the network, while in the TE each node has to send a different (personalized) packet to every other node of the network. Mišić and Jovanović [22] have proposed strictly optimal algorithms to execute both tasks in time $k!-1$ and $(k+1)!+o((k+1)!)$[‡], respectively, in a k-star with single-dimension communication. Using Theorem 3.1, the algorithms in [22] give rise to corresponding algorithms for the k-RE network.

Corollary 5.1 *The total exchange task can be optimally executed in a k-RE network under the SDC model in $O((k+2)!)=O\left(\frac{N\log^2 N}{(\log\log N)^2}\right)$ steps, where $(k+1)!+o((k+1)!)=\frac{N\log_2 N}{\log_2\log_2 N}+o\left(\frac{N\log N}{\log\log N}\right)$ steps involve transmissions over exchange links and $O\left(\frac{N\log^2 N}{(\log\log N)^2}\right)$ steps involve transmissions over rotation links, by emulating any optimal TE algorithm developed for the star graph under the SDC or the all-port communication model, where $N=k!$ is the size of the k-RE network.*

[‡]The notation $f(N)=o(g(N))$ means that $\lim_{N\to\infty} f(N)/g(N)=0$.

Proof: The proof follows from Theorems 3.1 and 3.4. and the emulation of any optimal TE algorithm developed for star graphs under the SDC (e.g., the algorithm given in [22]) or the all-port communication model (e.g., the algorithm given in [10]). Since the diameter of a k-RE network is $\Theta(k^2)$, it is straightforward to show that the required number of steps $O((k+2)!) = O\left(\frac{N\log^2 N}{(\log\log N)^2}\right)$ is of the optimal order of magnitude. □

Note that since each node of an RE network has degree equal to 3, the all-port communication model (where all incident links can be used for packet transmission and reception at the same time) and the single-dimension communication model are equally powerful within a constant of 3.

The slowdown for emulating MNB algorithms developed for star graphs under the all-port communication model is considerably smaller than the upper bound given in Theorem 3.4.

Corollary 5.2 *The multinode broadcast task can be executed in a k-RE network under the SDC model or the all-port communication model with $k! + o(k!) = N + o(N)$ steps for transmission over exchange links and $O(k!) = O(N)$ steps for transmission over rotation links, by emulating any optimal MNB algorithms developed for star graphs under the all-port communication model.*

Proof: We emulate optimal MNB algorithms developed for star graphs (e.g., the algorithms given in [22]), and use the technique described in the proof of Theorem 3.4. The TE algorithms for $(k-1)$-node rings used in the proof of Theorem 3.4 are replaced here with MNB algorithms for $(k-1)$-node rings. □

Using similar emulation techniques, we can obtain an efficient algorithm to perform single-node broadcasting algorithm in RE networks.

Corollary 5.3 *The single-node broadcasting task can be executed in a k-RE network under the SDC model or the all-port communication model in at most $O(k^2) = O\left(\frac{\log^2 N}{(\log\log N)^2}\right)$ steps, out of which $\lceil 3(k-1)/2 \rceil$ steps correspond to transmissions over exchange links and $O\left(\frac{\log^2 N}{(\log\log N)^2}\right)$ steps correspond to transmissions over rotation links.*

Proof: We simply emulate any optimal single-node broadcast algorithm developed for k-stars (which requires $k-1$ steps) under the all-port communication model. The proof is similar to those for Theorem 3.4 and Corollary 5.2, but instead of executing TE or MNB tasks in $(k-1)$-node rings, we now execute single-node broadcasts in $(k-1)$-node rings. □

When on-module links have larger transmission rates than off-module links, the execution times of some of the previous tasks can be considerably reduced as the following corollaries indicate.

Corollary 5.4 *Single-node broadcasting can be performed in a k-RE network in $O(k) = O\left(\frac{\log N}{\log\log N}\right)$ time if the transmission rate of on-module links is $\Omega(k) = \Omega\left(\frac{\log N}{\log\log N}\right)$, the transmission rate of off-module links is $\Omega(1)$, and nodes belonging to the same rotation ring are placed on the same module.*

Proof: The proof follows from Theorem 4.1 and Corollary 5.3. □

Corollary 5.5 *The total exchange task can be executed in a k-RE network in $O((k+1)!) = O\left(\frac{N\log N}{\log\log N}\right)$ time if the transmission rate of on-module links is $\Omega(k) = \Omega\left(\frac{\log N}{\log\log N}\right)$, the transmission rate of off-module links is $\Omega(1)$, and nodes belonging to the same rotation ring are placed on the same module.*

Proof: The proof follows from Corollary 5.1. □

From Corollary 5.2, it is evident that higher transmission rates for on-module links do not help reduce the execution time of the multinode broadcast task. This is because for this task on-module links and off-module links are utilized to similar extent, and they both form a bottleneck for communication.

6. Conclusion

We have derived efficient embeddings and emulation of star graphs, meshes, hypercubes, CCC, pancake graphs, bubble-sort graphs, complete transposition graphs, and the shuffle-exchange permutation graphs, under a variety of assumptions on the communication model. We presented efficient algorithms to perform routing, single-node broadcasting, multinode broadcasting, and total exchange, on RE networks. We also showed that the performance of RE networks can be significantly if on-module transmissions are considerably faster than off-module transmissions.

References

[1] Akers, S.B. and B. Krishnamurthy, "Group graphs as interconnection networks," *Proc. Int'l Conf. Fault-Tolerant Computing*, 1984, pp. 422-427.

[2] Akers, S.B., D. Harel, and B. Krishnamurthy, "The star graph: an attractive alternative to the n-cube," *Proc. Int'l Conf. Parallel Processing*, 1987, pp. 393-400.

[3] Akers, S.B. and B. Krishnamurthy, "A group-theoretic model for symmetric interconnection networks," *IEEE Trans. Comput.*, Vol. 38, Apr. 1989, pp. 555-565.

[4] Akl, S.G. and K.A. Lyons, *Parallel Computational Geometry*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[5] Azevedo, M.M., N. Bagherzadeh, and S. Latifi, "Broadcasting algorithms for the star-connected cycles interconnection network," *J. Parallel Distrib. Comput.*, vol. 25, no. 2, Mar. 1995, pp. 209-222.

[6] Azevedo, M.M., "Star-based interconnection networks and fault-tolerant clock synchronization for large multicomputers," Ph.D. dissertation, Dept. Electrical & Comp. Eng., Univ. of California, Irvine, 1997.

[7] Bagherzadeh, N., M. Dowd, and S. Latifi, "A well-behaved enumeration of star graphs," *IEEE Trans. Parallel Distrib. Sys.*, Vol. 6, no. 5, May 1995, pp. 531-535.

[8] Bertsekas, D.P. and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, 1997.

[9] Bouabdallah, A., M.C. Heydemann, J. Opatrny, and D. Sotteau, "Embedding complete binary trees into star networks," *Proc. Int'l Symp. Mathematical Foundations of Computer Science*, 1994, pp. 266-275.

[10] Fragopoulou, P. and S.G. Akl, "Optimal communication algorithms on star graphs using spanning tree constructions," *J. Parallel Distrib. Computing.*, Vol. 24, 1995, pp. 55-71.

[11] Hamdi, M., "A class of recursive interconnection networks: architectural characteristics and hardware cost," *IEEE Trans. Circuits and Sys.–I: Fundamental Theory and Applications*, vol. 41, no. 12, Dec. 1994, pp. 805-816.

[12] Hwang, K. and J. Ghosh, "Hypernet: a communication efficient architecture for constructing massively parallel computers," *IEEE Trans. Comput.*, vol. 36, no. 12, Dec. 1987, pp. 1450-1466.

[13] Johnsson, S.L. and C.-T. Ho, "Optimum broadcasting and personalized communication in hypercubes," *IEEE Trans. Comput.*, vol. 38, no. 9, Sep. 1989, pp. 1249-1268.

[14] Jwo J.S., S. Lakshmivarahan, and S.K. Dhall, "Embedding of cycles and grids in star graphs," *Proc. IEEE Symp. Parallel and Distributed Processing*, 1990, pp. 540-547.

[15] Kogge, P.M., "EXECUBE - a new architecture for scalable MPPs," *Proc. Int'l Conf. Parallel Processing*, vol. I, 1994, pp. 77-84.

[16] Lakshmivarahan, S., J.-S. Jwo, and S.K. Dhall, "Symmetry in interconnection networks based on Cayley graphs of permutation groups: a survey," *Parallel Computing*, Vol. 19, no. 4, Apr. 1993, pp. 361-407.

[17] Latifi, S., M. Azevedo, and N. Bagherzadeh, "The star connected cycles: a fixed-degree network for parallel processing," *Proc. Int'l Conf. Parallel Processing*, Vol. I, 1993, pp. 91-95.

[18] Latifi, S. and P.K. Srimani, "A new fixed degree regular network for parallel processing," *Proc. IEEE Symp. Parallel and Distributed Processing*, Oct. 1996, pp. 152-159.

[19] Latifi, S. and P.K. Srimani, "Transposition networks as a class of fault-tolerant robust networks," *IEEE Trans. Parallel Distrib. Sys.*, Vol. 45, no. 2, Feb. 1996, pp. 230-238.

[20] Leighton, F.T., *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes*, Morgan-Kaufman, San Mateo, CA, 1992.

[21] Miller, Z., D. Pritikin, and I.H. Sudborough, "Bounded dilation maps of hypercubes into Cayley graphs on the symmetric group," *Math. Sys. Theory*, Vol. 29, no. 6, Springer-Verlag, Nov.-Dec., 1996, pp.551-572.

[22] Mišić, J. and Z. Jovanović, "Communication aspects of the star graph interconnection network," *IEEE Trans. Parallel Distrib. Sys.*, Vol. 5, no. 7, Jul. 1994, pp. 678-687.

[23] Preparata, F.P. and J.E. Vuillemin, "The cube-connected cycles: a versatile network for parallel computation," *Commun. of the ACM*, Vol. 24, no. 5, May 1981, pp. 300-309.

[24] Saikia, D.K. and R.K. Sen, "Two ranking schemes for efficient computation on the star interconnection networks," *IEEE Trans. Parallel Distrib. Sys.*, Vol. 4, Apr. 1996, pp. 321-327.

[25] Scherson, I.D., and A.S. Youssef, *Interconnection Networks for High-Performance Parallel Computers*, IEEE Computer Society Press, 1994.

[26] Varvarigos, E.A., "Static and dynamic communication in parallel computing," Ph.D. dissertation, Dept. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1992.

[27] Varvarigos, E.A. and D.P. Bertsekas, "Multinode broadcast in hypercubes and rings with randomly distributed length of packets," *IEEE Trans. Parallel Distrib. Sys.*, vol. 4, no. 2, Feb. 1993, pp. 144-154.

[28] Yeh, C.-H. and E.A. Varvarigos, "Macro-star networks: efficient low-degree alternatives to star graphs for large-scale parallel architectures," *Proc. Symp. Frontiers of Massively Parallel Computation*, Oct. 1996, pp. 290-297.

[29] Yeh, C.-H. and B. Parhami, "Recursive hierarchical swapped networks: versatile interconnection architectures for highly parallel systems," *Proc. IEEE Symp. Parallel and Distributed Processing*, Oct. 1996, pp. 453-460.

[30] Yeh, C.-H. and B. Parhami, "Cyclic networks – a family of versatile fixed-degree interconnection architectures," *Proc. Int'l Parallel Processing Symp.*, Apr. 1997, 739-743.

[31] Yeh, C.-H., "Efficient low-degree interconnection networks for parallel processing: topologies, algorithms, VLSI layouts, and fault tolerance," Ph.D. dissertation, Dept. Electrical & Computer Engineering, Univ. of California, Santa Barbara, Mar. 1998.

[32] Yeh, C.-H. and E.A. Varvarigos, "Macro-star networks: efficient low-degree alternatives to star graphs," *IEEE Trans. Parallel Distrib. Sys.*, Vol. 9, no. 10, Oct. 1998, pp. 987-1003.

[33] Yeh, C.-H. and B. Parhami, "VLSI layouts of complete graphs and star graphs," *Information Processing Letters*, Vol. 68, Oct. 1998, pp. 39-45.

[34] *Proc. Symp. Frontiers of Massively Parallel Computation*, NASA Scientific and Technical Information Office, Washington, D.C., Oct. 1996.