Date:	27 Feb 2025
To:	"Samira Hossein Ghorban" s.hosseinghorban@ipm.ir
From:	"J. of Interconnection Networks (JOIN)" join@wspc.com
Subject:	Your Submission: JOIN-D-24-00180R1

Ref.: Ms. No. JOIN-D-24-00180R1
A Cost-Performance Scalability Measure for Interconnection Networks and a Novel Scalable Cube-Based Topology
Journal of Interconnection Networks

Dear Dr. Samira Hossein Ghorban,

I am pleased to inform you that your work has now been accepted for publication in Journal of Interconnection Networks.

It was accepted on 27 Feb 2025.

Journal of Interconnection Networks encourages authors to consider publishing their papers in Open Access. OA ensures that copyright of the article is fully retained by the authors, and that all readers can view your full article freely online. This strongly enhances the discoverability of your article, and raises the possibility of increasing its download and citation count. OA papers will be processed upon acceptance as a priority.

World Scientific has various Open Access publishing arrangements with higher education institutions, and developing and transition economy countries.

Such arrangements include Publish and Read agreement with JISC for participating UK Higher education institutions, and discounts or waivers of Article Processing charges (APC) for some lower-income economies. Please visit the journal’s Open Access page to find out more: https://www.worldscientific.com/JOIN/openaccess.

Should you wish to publish your paper on an Open Access basis, please inform us of your choice at join@wspc.com.

Comments from the Editor and Reviewers can be found below.

Thank you for submitting your work to this journal.

IMPORTANT NOTICE

World Scientific Publishing will only issue an official invoice for the payment of an Article-Processing Charge if you opt for Gold Open Access, and after your paper has been accepted. Before responding to any payment request, we advise you to check on your submission's status at Editorial Manager website, or contact the Editor-in-Chief/Managing Editor, to ensure your submission has officially been “Accepted” for publication. Please disregard any payment request if otherwise. When in doubt, contact us at World Scientific (editor@wspc.com).

With kind regards

Eddie Cheng
Editor-in-Chief
Journal of Interconnection Networks

Comments from the Editors and Reviewers:

The authors have addressed the minor issues.

Journal of Interconnection Networks

A Cost-Performance Scalability Measure for Interconnection Networks and a Novel Scalable Cube-Based Topology

--Manuscript Draft--

Manuscript Number:	JOIN-D-24-00180R1
Full Title:	A Cost-Performance Scalability Measure for Interconnection Networks and a Novel Scalable Cube-Based Topology
Article Type:	Research Paper
Keywords:	Interconnection networks; Network scalability; Hypercubes; Overlapped cubes
Corresponding Author:	Samira Hossein Ghorban Institute for Research in Fundamental Sciences School of Computer Science IRAN (ISLAMIC REPUBLIC OF)
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Institute for Research in Fundamental Sciences School of Computer Science
Corresponding Author's Secondary Institution:	
First Author:	Samira Hossein Ghorban
First Author Secondary Information:	
Order of Authors:	Samira Hossein Ghorban
	Bardyaa Hesaam
	Hamid Sarbazi-Azad
Order of Authors Secondary Information:	
Abstract:	Employing proper interconnection networks is essential for enhancing the overall performance of multiprocessor systems. Efficient size scalability is a critical consideration when designing interconnection networks for cost-effective implementation and management of communication overhead. Size scalability indicates how well systems can scale capacity when expanding to permitted larger configurations based on the network topology constraints. In this paper, we first explore establishing a formulation for evaluating cost-performance scalability. Then, a comparison among hypercube, as a prominent and widely used interconnection network topology, and its famous varieties is conducted given their scalability. We then propose a novel variation of the hypercube, named Overlapped Cube (or OCube, for short) on m, n, k $\in$ N such that $m \geq n > k$ and denoted by OQm,n,k and examine its topological properties. Comparing the scalability of some traditional networks and known variants of hypercubes demonstrates that OCube has the potential to be an effective interconnection network with desirable scaling behavior.
Response to Reviewers:	We have considered all the comments made by the reviewers and applied them all. Note that, for the first comment of reviewer #2 about the appearance of references, the current visual form of references (as super scripts) is implied by the ws-join class in the JOIN template. As we used the standard \cite macro without any modifications, they should appear in the correct form within the finalized form of the manuscript.

Journal of Interconnection Networks

A Cost-Performance Scalability Measure for Interconnection Networks and a Novel Scalable Cube-Based Topology

SAMIRA HOSSEIN GHORBAN* and BARDYAA HESAAM

School of Computer Science
Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
bardyaa@ipm.ir, s.hosseinghorban@ipm.ir

HAMID SARBAZI-AZAD

Department of Computer Engineering
Sharif University of Technology, Tehran, Iran
School of Computer Science
Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
azad@sharif.edu, azad@ipm.ir

Employing proper interconnection networks is essential for enhancing the overall performance of multiprocessor systems. Efficient size scalability is a critical consideration when designing interconnection networks for cost-effective implementation and management of communication overhead. Size scalability indicates how well systems can scale capacity when expanding to permitted larger configurations based on the network topology constraints. In this paper, we first explore establishing a formulation for evaluating cost-performance scalability. Then, a comparison among hypercube, as a prominent and widely used interconnection network topology, and its famous varieties is conducted given their scalability. We then propose a novel variation of the hypercube, named Overlapped Cube (or OCube, for short) on $m, n, k \in \mathbb{N}$ such that $m \geq n > k$ and denoted by $OQ_{m,n,k}$ and examine its topological properties. Comparing the scalability of some traditional networks and known variants of hypercubes demonstrates that OCube has the potential to be an effective interconnection network with desirable scaling behavior.

Keywords: Interconnection networks; Network scalability; Hypercubes; Overlapped cubes.

AMS Subject Classification: 68M10, 68R10

1. Introduction

The performance of a multiprocessor system is mainly affected by its underlying interconnection network, which is responsible for inter-node data communication. ⊗

*Corresponding author.

The topology of a multiprocessor system (and thus its interconnection network) is defined by a graph where vertices show nodes and edges indicate the inter-node channels or links. As the total cost of a machine is important, constructing sparse graphs with small diameters is a key focus in the study of interconnection networks. However, a small diameter implies that the maximum number of nodes required for a message to pass between any two nodes in the network is low. This implements low-latency inter-node communication, which is desirable for efficient data exchange, and hence, results in higher overall system performance. Consequently, this makes it easier to scale the network to a larger number of nodes while keeping the communication overheads manageable. One element of manageability is the ability to efficiently scale system size, called size scalability [23].

A multitude of network architectures have been investigated for application in parallel computers [22,36,13,21]. The network configurations most typically employed in practical applications include trees, cycles, grids, tori, meshes, and hypercubes [30]. Of the various topologies, hypercube-based networks have received substantial research interest. Hypercubes (also called n-cubes or binary n-cubes) are popular interconnection networks used in parallel computer architectures. An n-dimensional hypercube is comprised of $N = 2^n$ nodes placed at the corners of an n-dimensional cube [11]. In exact words, the nodes of a hypercube network are labeled from 0 to $N - 1$; two nodes are connected if and only if the binary representations of their labels differ by one bit, which means their Hamming distance is 1. Consequently, the degree of each node of an n-dimensional hypercube and its diameter is n. Harary et al. [10] reported a comprehensive survey of the fundamental properties associated with hypercubes including distance, coloring, domination, and Hamiltonian cycles. Furthermore, different properties of hypercubes have been extensively studied in the literature, as evidenced by a number of publications, including [29,6,3,5,24,19].

While hypercube architectures exhibit numerous beneficial topological, algorithmic, and fault-tolerance properties, they present some restrictions in practice. Several hypercube variants have been proposed in the literature to address the limitations of the original hypercube structure based on different approaches [27,1,10,24,7,35]. A significant drawback of hypercube topology is that the network size (number of nodes) must be a power of two which is a severe restriction on the feasible network sizes that can be constructed using hypercubes. Stated differently, hypercube networks cannot be constructed for any arbitrary number of nodes. Consequently, the conventional feature of hypercube disregards size scalability. To address this restriction, Katseff et al. [16] proposed the incomplete hypercubes which may be obtained by pruning unnecessary nodes of a hypercube. While the structure inherits many nice features of the hypercube structure, fault tolerance of the network is a significant limitation of incomplete hypercubes. This limitation is particularly severe in incomplete hypercube networks where even a single node failure can cause the network to become disconnected. Sen et al. [31] attempted to address the fault tolerance issue by introducing a new network topology, called Supercube which enjoys the same structure of incomplete hypercube enriched by some extra links. This modification results in improved fault tolerance at the expense of losing the property of being embeddable in a bigger hypercube structure. Topological properties of supercubes are studied in [37,33,34].

Structurally, different hypercube variants can be classified into two categories. The first category includes those that can be regarded as edge-induced subgraphs of a hypercube, such as incomplete hypercubes [16] and the connected cubic network graphs [30]. The second category comprises variants of network topologies that are based on the hypercube, such as the supercube [31,34], the necklace-hypercube [26], the Stretched-Hypercube [32], the Linear Crossed Cube [17], the Crossed Cube [4,7,20], the Hierarchical cubic networks [9,14,15], the twisted cube [38], the Folded Hypercube and the k-ary n-cube [8,2,12]. ⊗

When designing interconnection networks, hardware scalability is critical for the ⊗ cost-effective implementation and operation management of a multicomputer system. It indicates how gracefully the system scales to the next available and possible size according to the definition of its network topology. Consequently, the primary concern in designing interconnection networks is to evaluate size scalability which has been discussed, descriptively. Note that performance scalability has been formulated [23,28], but size scalability is not investigated quantitatively. This paper proposes a new quantitative approach to model size scalability for interconnection networks and then studies the size scalability of hypercube and its famous variants. Finally, we introduce a new hypercube-based interconnection network topology by merging two hypercubes with dimensions m and n with respect to a hypercube of dimension k, where $m \geq n > k$, called Overlapped hypercube. This variant is an induced edge subgraph of the hypercube and possesses good size scalability while maintaining other desirable hypercubic properties.

The rest of the paper is organized as follows. Section 2 gives some definitions and propositions. In Section 3, we investigate the scalability of interconnection networks and derive a hybrid cost-performance scalability measure. Section 4 introduces a novel scalable hypercube-based interconnection network topology, named overlapped cube, and examines its topological properties. Finally, Section 5 concludes the paper.

2. Preliminaries

In this section, we briefly give some definitions, notations, and propositions which will be used later. From now on, we write $X(Y)$ for the property X of the object Y, and we may omit Y if the object is known from context. Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

We begin with some familiar notions of graph theory.

Definition 2.1. A network (or graph) G is a tuple (V, E) where V is a non-empty set of vertices and E is a set of two-element subsets of V or edges.

So, if edge $e \in E$ then there are some distinct $x, y \in V$ such that $e = \{x, y\}$. Sometimes, we may write $e = xy$.

Definition 2.2. Basic characteristics: Let $G = (V, E)$ represent a network.

- Each element of V is called a node (or a vertex), and each element of E is called a link or channel (or an edge). The number of nodes is called size of the network. We usually denote it by N.
- With $V = \{v_1, \ldots, v_N\}$, the adjacency matrix of G is a zero-one square matrix of order N whose (i, j) entry is 1 if $v_iv_j \in E$ and 0 otherwise. Note that the adjacency matrix is not unique, as it strictly depends on the chosen ordering (or label) of nodes.
- For a node x, a node y is called its neighbor if $\{x, y\} \in E$; i.e., there is a link in G between x and y. The number of all neighbors of x is called the node degree of x. The maximum node degree in the network is called network degree, denoted by M.
- For two nodes x and y of G, a path between x and y is a sequence $x_1x_2 \ldots x_k$ of nodes where $x_1 = x$, $x_k = y$, and x_{i+1} is a neighbor of x_i for each $1 \leq i < k$. The length of such a path is defined to be $k-1$; i.e., it is the number of links that one passes from x to y along the path.
- For two nodes x and y of G, the length of the shortest path connecting x and y is called the distance of x and y, denoted by $d(x, y)$. For the network itself, the length of the longest shortest path is called diameter, denoted by D, and the average distance is defined to be

$$\bar{d} = \frac{\sum_{x,y \in V} d(x, y)}{N(N-1)}. \tag{2.1}$$

- Two nodes x and y are called connected if there is a path between them.
- G is said to be connected if each two nodes of G are connected. It is called disconnected if it is not connected.
- A network H is called a subgraph of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. H is called an induced subgraph of G if $E(H) = \{e \in E(G) \mid e \subseteq V(H)\}$; i.e., all nodes of H that are neighbors in G are also neighbors in H.
- For a natural number f, G is said to be f-fault-tolerant if after removing any arbitrary f nodes, the remaining induced subgraph is still connected; i.e., it is guaranteed the network remains connected even if f nodes of the network become faulty [18].

Many types of interconnection network topologies have been introduced and studied. Here we present some well-known topologies.

Definition 2.3. Popular networks: Well-known networks of size $N \in \mathbb{N}$ and node set $V = \{0, \ldots, N-1\}$ include:

- Path (or array) is the network $P_N = (V, E)$, where $E = \big\{\{i-1, i\} \mid 1 \le i < N\big\}$; i.e., line the nodes up in a row and link each node to its next and previous neighbors (except for the nodes at the two ends).
- Star is the network $S_N = (V, E)$ where $E = \big\{\{i, 0\} \mid 1 \le i < N\big\}$; i.e., choose a node and link all others only to it.
- Complete graph is the network $K_N = (V, E)$ where $E = \big\{\{i, j\} \mid 1 \le i < j < N\big\}$; i.e., link each nodes to all other nodes.
- Hypercube, which can be defined only when $N = 2^n$ for a non-negative integer n, is the network $Q_n = (V, E)$ where E consists of all pairs $\{x, y\}$, $x, y \in V$, where the binary representations of x and y as sequences of length n are the same except exactly in one place. One may imagine Q_n as two identical copies of Q_{n-1}, with extra links between "same" nodes in two copies — namely, between those nodes that have the same binary representation except for their leftmost bit. For an arbitrary node x, we denote the corresponding node in the other copy with x' and call it the project of x.
- Incomplete hypercube is the network $\mathrm{IH}_N = (V, E)$ as the induced subgraph of Q_n where $n = \lceil \log_2 N \rceil$ [16].
- Supercube, SC_N, is an extension of incomplete hypercube IH_N where $n = \lceil \log_2 N \rceil$, with all links of the form $\{r, s\}$ for $r < 2^{n-1} \le s$ and $r' \ge n$ and $\{r', s\}$ in corresponding Q_n [34].

Proposition 2.1. Let G be a network, $u \ge 3$ be a natural number, and $v \in \mathbb{N}_0$.

(1) For $G = Q_v$, we have $N = 2^v$, $D = v$, and $M = v$.
(2) For $G = \mathrm{IH}_u$, we have $N = u$, $D = v$, and $M = v$ where $v = \lceil \log_2 u \rceil$.
(3) For $G = \mathrm{SC}_u$, we have $N = u$, $v - 1 \le D \le v$, and $v - 1 \le M \le 2(v-1)$ where $v = \lceil \log_2 u \rceil$.

Proof.

(1) For each node, we have a binary sequence of length v bits. Each neighbor has the same sequence with exactly one different bit; so each node is exactly connected to v other nodes and hence $M = v$. For $x, y \in V$, there may be v different bits to change at most; so $D \le v$, and to convert each string to its 'negate', we need v changes; so $D = v$. Note that, doing these change operations from left bits toward right bits, one finds that for $x > y$, there is a path connecting x to y of length $d(x, y)$ in which all contributing nodes are less than x.
(2) Since $2^{v-1} \le u < 2^v$, we have $\{2^0, 2^1, \ldots, 2^{v-1}\} \subseteq V(\mathrm{IH}_u)$, and each of these nodes has a link to 0; so $M \ge v$. On the other hand, IH_u is induced subgraph of Q_v; so $M \le v$.

To calculate D, let $x, y \in V(\mathrm{IH}_u)$. If $x, y < 2^{v-1}$, then by (i) it is known that $d(x, y) \leq v - 1$. If $x, y \geq 2^{v-1}$, then the same argument works for connecting $x', y' < 2^{v-1}$, and projecting the path gives a path of length at most $v - 1$ between x, y. If $x < 2^{v-1} \leq y$ then $d(x, y) \leq d(x, x') + d(x', y) \leq 1 + v - 1 = v$. Finally, u connects to its negate (namely $2^v - 1 - u$) via a path of length v.

(3) Let V_1 be the set of those nodes in SC_u that has 1 as the leftmost bit, and take an arbitrary node r in V.

If $r \in V_1$, then for each s that $\{r, s\}$ is a link in Q_v, either $s < u$ so $\{r, s\}$ is a link in SC_u, or $s \geq u$ so $\{r, s'\}$ is a link in SC_u; hence r is linked to v other nodes.

For $r \notin V_1$, then $r < 2^{v-1}$ and it has $v - 1$ links to the nodes that are out of V_1. There are again two cases:

- If $r' \in V_1$, then $\{r, s\}$ is a link in SC_u if and only if $s = r'$; so r is linked to $v - 1 + 1 = v$ other nodes.
- If $r' \notin V_1$, then all nodes $s \in V_1$ with $\{r', s\}$ is a link in Q_v would be linked to r in SC_u; so 0 (when no node in V_1 is linked to r' in Q_v: consider 011 in SC_5) to $v - 1$ (when $v - 1$ nodes in V_1 are linked to r' in Q_v: consider 011 in SC_7) nodes of V_1 are linked to r.

Therefore, $v - 1 \leq M \leq 2(v - 1)$.

The diameter is v when $u = 2^v$, and is $v - 1$ when $u < 2^v$ [31] . □

We end up this section with some elementary concepts from mathematical analysis [25].

Definition 2.4. Let S be a non-empty subset of $\mathbb{R}$.

- A $u \in \mathbb{R}$ is called an upper bound of S if $s \leq u$ for all $s \in S$. If S has an upper bound, then the least upper bound of S is called the supremum of S and is denoted by $\sup S$. That is, $\alpha = \sup S$ if and only if α is an upper bound for S and no $\alpha' < \alpha$ has such property.
- An $l \in \mathbb{R}$ is called a lower bound of S if $l \leq s$ for all $s \in S$. If S has a lower bound, then the greatest lower bound of S is called the infimum of S and is denoted by $\inf S$. That is, $\beta = \inf S$ if and only if β is a lower bound for S and no $\beta' > \beta$ has such property.
- S is said to be bounded if has both lower bound and upper bound.

Note that, existence of the least upper bound and greatest lower bound is an important property of the real line, sometimes called completeness of $\mathbb{R}$.

Definition 2.5. Consider a bounded sequence $a_1, a_2, \ldots$.

- Define $a_n^{(u)} = \sup\{a_n, a_{n+1}, \ldots\}$ for each $n \in \mathbb{N}$. The number $\inf\{a_1^{(u)}, a_2^{(u)}, \ldots\}$

is called upper limit (or limit supremum) of the sequence, denoted by $\overline{\lim}_{n\to\infty} a_n$ (or $\limsup_{n\to\infty} a_n$).

- Define $a_n^{(l)} = \inf\{a_n, a_{n+1}, \ldots\}$ for each $n \in \mathbb{N}$. The number $\sup\{a_1^{(l)}, a_2^{(l)}, \ldots\}$ is called lower limit (or limit infimum) of the sequence, denoted by $\underline{\lim}_{n\to\infty} a_n$ (or $\liminf_{n\to\infty} a_n$).

Note that, since every bounded monotone sequence converges to a real number (its supremum or infimum for ascending or descending sequence respectively), upper and lower limits exist for a bounded sequence, even if sequence does not converge itself. Trivially, lower limit lies below upper limit, and for a convergent sequence, both lower and upper limits are equal to the limit of sequence. Moreover, both limits preserve order, and we have the following version of the well-known 'Sandwich' ⊗ theorem.

Proposition 2.2. Consider sequences a, b, and c, where a and c are convergent and $a_n \leq b_n \leq c_n$ for all $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} a_n \leq \underline{\lim}_{n\to\infty} b_n \leq \overline{\lim}_{n\to\infty} b_n \leq \lim_{n\to\infty} c_n. \tag{2.2}$$

3. Cost-efficient Scalability

Usually, one must choose networks from a pre-determined 'family,' and so the ability of networks to scale gracefully in response to demands for more nodes is a crucial issue. Scalability of a network family refers to its ability to accommodate additional nodes in order to maintain both cost and performance in a consistent way, as adding even one single node may lead to the selection of a 'big' new network with many unnecessary nodes.

Restricting cost to the number of nodes, the ratio of a requested network size over the least possible network size provides a measure of cost-efficiency for the hardware scalability of an interconnection network. For the sake of mathematical notation, let $\mathcal{G} = \{G_i\}_{i=1}^{\infty}$ represent a family of possible networks with a strictly increasing network size sequence $\{a_i\}_{i=1}^{\infty}$ of natural numbers, and let $a_0 = 0$ if necessary. For a requested network size r, define $A(r)$ to be the least possible network size that is necessary to contain r nodes; i.e.

$$A(r) = \min_{a_n \geq r} a_n. \tag{3.1}$$

Then, we may define partial cost-efficient scalability of $\mathcal{G}$ as the average of the ratios $r/A(r)$ up to m requested nodes; i.e.

$$S_m^{(p)} = \frac{1}{m} \sum_{r=1}^{m} \frac{r}{A(r)}. \tag{3.2}$$

Here we added the superscript (p) for this 'plain' formulation of scalability. Clearly, a linear array and a complete network have a full value of scalability 1, since for every requested network size r, we have $A(r) = r$. The value of $S_m^{(p)}$ for $1 \leq m \leq 2^{25}$ for hypercubes are depicted in Fig. (1). It may be seen that there is no limit value to which $\{S_m^{(p)}\}_{m=1}^{\infty}$ converges while there is a natural number $t > 2^5$ such that values of $S_m^{(p)}$ for $m > t$ exhibit "alternating" behavior.

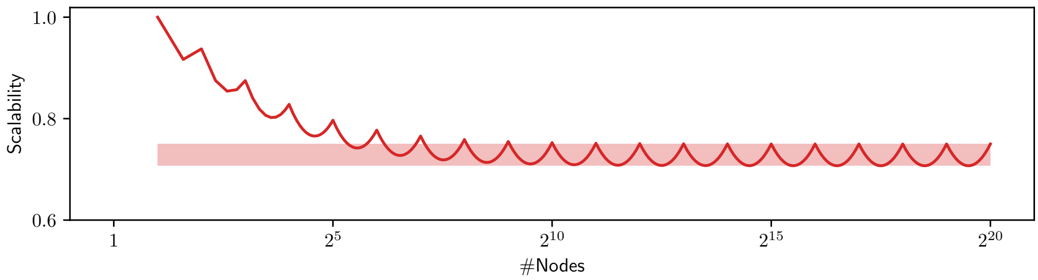

Fig. 1. Scalability versus nodes for hypercubes. Transparent region shows the ultimate boundaries.

Since the sequence $\{S_m^{(p)}\}_{m\geq 1}$ does not converge, we try to use the same broad concept of "limit" that is defined in Definition (2.5) to provide a suitable tool to compare two families from the "scalability" viewpoint: for two families $\mathcal{G}$ and $\mathcal{H}$, we say that $\mathcal{G}$ enjoys strictly more scalability than $\mathcal{H}$ if the lower limit of $S^{(p)}$ for $\mathcal{G}$ lies above the upper limit of $S^{(p)}$ for $\mathcal{H}$; i.e.

$$\varliminf_{m\to\infty} S_m^{(p)}(\mathcal{G}) > \varlimsup_{m\to\infty} S_m^{(p)}(\mathcal{H}). \tag{3.3}$$

This definition gives a very tight condition, where a similar idea may be taken to give more flexible variants, each with its own benefits. Some elaborate yet straightforward computations show that if

$$\lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \rho, \tag{3.4}$$

then

$$\varliminf_{m\to\infty} S_m^{(p)} = \sqrt{\rho} \quad \text{and} \quad \varlimsup_{m\to\infty} S_m^{(p)} = \frac{1+\rho}{2}. \tag{3.5}$$

In Table (1) values of ρ and the upper and lower limits for some typical values of a_n are shown.

Despite its nice yet simple formalization, the current approach to measure scalability gives good values for bad families, as the number of nodes itself is not a sufficient measure to prefer a network family to another, and more points should be considered. Among many parameters, it seems that the role of network edge

a_n	n	$an+b$	$O(n^k)$	2^n	3^n	$\Theta(k^n)$	$n!$	2^{2^n}
ρ	1	1	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{k}$	0	0
$\overline{\lim}_{m\to\infty} S_m^{(p)}$	1	1	1	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{k+1}{2k}$	$\frac{1}{2}$	$\frac{1}{2}$
$\underline{\lim}_{m\to\infty} S_m^{(p)}$	1	1	1	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{3}$	$\frac{\sqrt{k}}{k}$	0	0

set size and node types are more important: a family with too few links (like arrays) should be considered inappropriate since it implies a big diameter and also lacks fault tolerance, while a family with too many links (like complete networks) is also inappropriate as it forces high costs on connection backbone and needs heavy changes to nodes type when adding a new node. In Fig. (2) values of this “plain” scalability are shown: it assigns the highest possible scalability value to families that are trivially inappropriate, as array (too few links, low fault tolerability, high diameter), complete (too many links), and star (too few links, low fault tolerability, one hub node against many leaves). The two cube variants, incomplete hypercube and supercube, enjoy the same scalability value since they have the same order sequence as those three, but with a more plausible size and diameter.

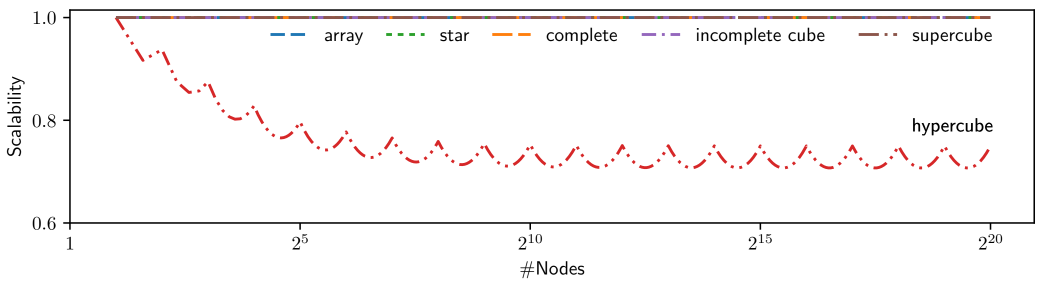

Fig. 2. ‘Plain’ scalability where ‘cost’ reduced to network order, for some known families of networks.

To give a lower scalability value to networks with too few or too many links, we may take the effect of degrees and diameter into account: a low value of maximum degree causes lower values for fault-tolerance together with a high diameter, whilst a high value may imply a big degree. As the values of diameter and maximum degree typically show a complementary behavior, we may consider their sum and compare it with a ‘fair’ magnitude. Although there is no perfect interpretation, it seems that the logarithm of the network size may give such a fair magnitude as it lies between constants and linear polynomials, with many well-known good properties, as seen in cube-based networks.

In general, we may think of a network as a logarithmic network if the average of its diameter and maximum degree is proportional to the logarithm of network size. The similarity between a designed network and a logarithmic network may be evaluated by comparing the average of the network's diameter and maximum degree with the logarithm of the network size. Denoting the diameter, maximum degree, and logarithm of the network size as D, M, and L, respectively, such a comparison may be done in different ways. A simple idea is to consider the difference squared, normalized by L: $\left(\frac{D+M}{2L}-1\right)^2$ measures the deviation from the logarithmic structure. One may consider a constant weight parameter $w_l \geq 0$ to control the impact of this parameter, and the whole coefficient

$$c(N) = \sqrt{1 + w_l \times \left(\frac{D+M}{2L} - 1\right)^2} \tag{3.7}$$

for the network with size N in the family of networks is a suitable magnification factor for the denominator $A(r)$ in $r/A(r)$. Then, we may apply the coefficients on the summands in $S_m^{(p)}$ to obtain

$$S_m^{(\ell)} = \frac{1}{m} \sum_{r=1}^{m} \frac{r}{A(r)} \times \frac{1}{c\big(A(r)\big)}. \tag{3.8}$$

In Fig. (3), the new values are depicted, where scalability values for non-logarithmic networks (array, complete, and star) tending to zero, while the logarithmic network still remains more or less in its previous level of scalability. It worth mentioning that there are many choices for such a coefficient with no difference in essence.

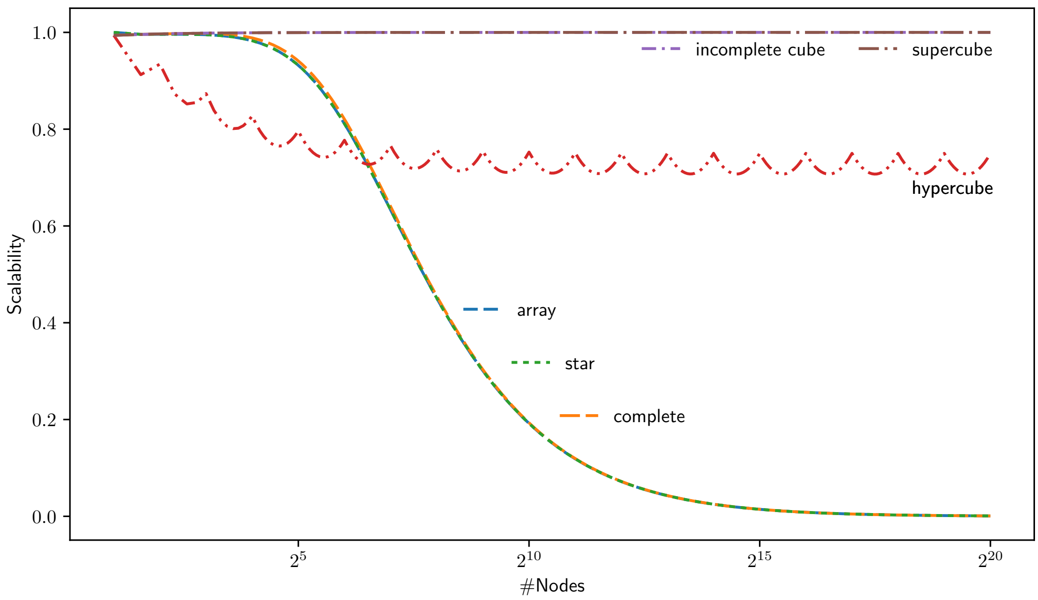

Fig. 3. 'Logarithmic' scalability where terms are adjusted to give better values for logarithmic structures as in Eq. (3.8) with $w_l = 0.1$.

As it may be seen in Fig. (3), Eq. (3.8) filters 'inappropriate' families; but we should keep in mind that even for the remaining logarithmic families, the simplicity of design is indeed a key advantage when it comes to interconnection network architectures and their implementation. In the case of hypercubes, the nodes are identical (regular structure). This uniformity contributes to lower implementation costs since the same type of node can be replicated throughout the network. So, it seems plausible to give better scalability values to regular and near-regular networks.

To address this, let $t(N)$ be the number of node types — i.e. the number of different node degrees in the network with size N, and we propose a penalty function $p(N)$ in terms of $t(N)$ to rewrite Eq. (3.8) as

$$S_m = \frac{1}{m}\sum_{r=1}^{m}\frac{r}{A(r)} \times \frac{1}{c\big(A(r)\big)} \times \frac{1}{p\big(A(r)\big)}. \tag{3.9}$$

Again, there are many choices for such a penalty function; but here we use

$$p(N) = \big(1 + w_t \times t(N)\big)^{t(N)-1} \tag{3.10}$$

for an arbitrarily chosen positive w_t. The effect is shown in Fig. (4), where the values show no change for the hypercube (as it is regular), and big fall-down for both incomplete hypercube and supercube as they both have growing types in terms of the network order.

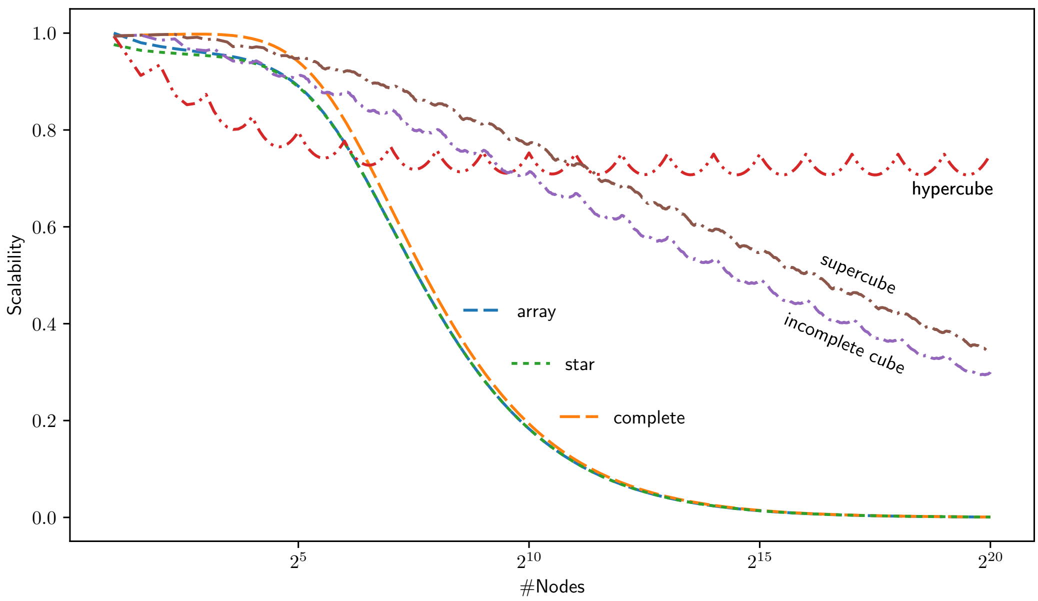

Fig. 4. 'Logarithmic' scalability with types, where a penalty is taken into account for a growing number of types in logarithmic networks as in Eq. (3.9) with $w_l = 0.1$ and $w_t = 0.025$.

4. Overlapped Cubes

In this section, we present the definition of the overlapped cube, $\mathrm{OQ}_{m,n,k}$, and then some characteristics of this family of networks are investigated. Intuitively, an overlapped cube $\mathrm{OQ}_{m,n,k}$ consists of hypercubes Q_m and Q_n which are joined together by a hypercube Q_k as it is depicted in Fig. (5a). As it may be seen in Fig. (5b), it may be considered as an induced subgraph of Q_{m+n-k}: node labels of the Q_m block may be taken as those m-bit labels of Q_m with $n-k$ filling zeros at right, where node labels of the Q_n block may be taken as those n-bit labels of Q_n with $m-k$ filling ones at left. Now, we formulate the definition as follows.

Definition 4.1. Let $m, n, k \in \mathbb{N}_0$, $m \geq n$, and either $n = k = 0$ or $n > k$. Consider the set V of all binary sequences of length $m+n-k$ with ones in all first $m-k$ places or zeroes in all last $n-k$ places. The overlapped cube has the node set $V(\mathrm{OQ}_{m,n,k}) = V$, and two nodes u and v are connected to each other if and only if the corresponding sequences are the same unless just in one place.

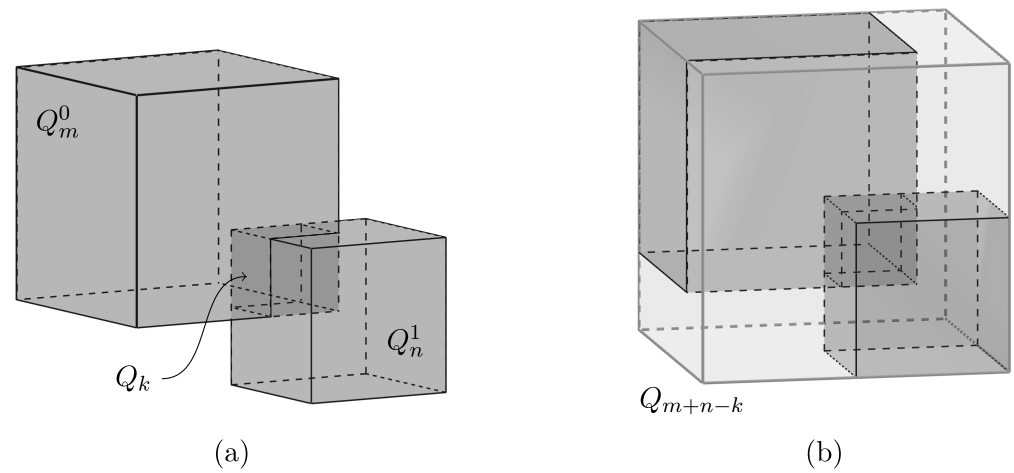

Fig. 5. (a) Intuitive idea behind the concept of overlapped cube: two cubes Q_m and Q_n intersect at a small cube Q_k. (b) The overlapped cube, embedded in a bigger cube Q_{m+n-k}.

We denote these embedded copies of Q_m and Q_n by Q_m^0 and Q_n^1 respectively. There are some examples of overlapped cubes in Fig. (6), and Fig. (7) shows the adjacency matrix of $\mathrm{OQ}_{5,3,2}$, in which a recursive pattern of hypercubes could be seen.

Note that, for a given (m, n, k) with $m, n, k \in \mathbb{N}_0$, $m \geq n > k$ or $n = k = 0$, we have $\left|V(\mathrm{OQ}_{m,n,k})\right| = 2^m + 2^n - 2^k$ and $\left|E(\mathrm{OQ}_{m,n,k})\right| = \frac{1}{2}(m2^m + n2^n - k2^k)$.

We call a natural number of the form $2^m + 2^n - 2^k$ an OQ-number, denoted by $\#\mathrm{OQ}_{m,n,k}$ for $m, n, k \in \mathbb{N}_0$, $m \geq n > k$ or $n = k = 0$. We may list all triples as

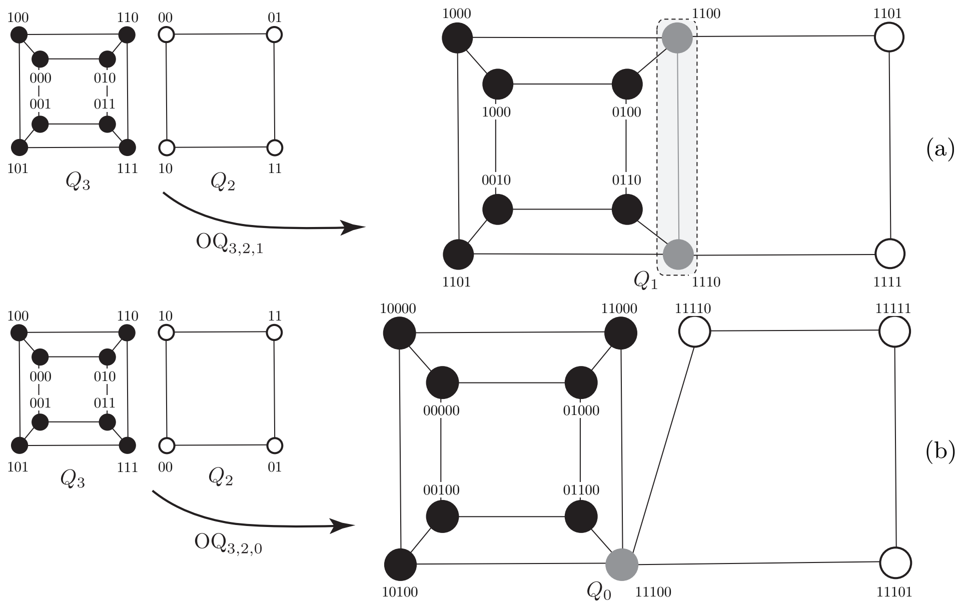

Fig. 6. Examples of hypercubes and overlapped cubes: the 3-d hypercube Q_3 and the 2-d hypercube Q_2 joind together in two different ways to provide (a) $\mathrm{OQ}_{3,2,1}$ and (b) $\mathrm{OQ}_{3,2,0}$.

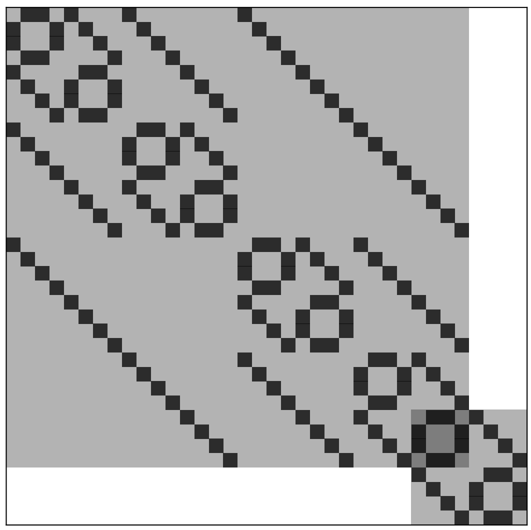

Fig. 7. A visualization of the adjacency matrix of $\mathrm{OQ}_{5,3,2}$. Corresponding adjacency matrices of Q_5 and Q_3 are highlighted, and their shared Q_2 is highlighted in dark gray.

follows: let $T_0 = (0,0,0)$, and for $r \in \mathbb{N}_0$, if $T_r = (m,n,k)$ then

$$T_{r+1} = \begin{cases} (m,n,k-1) & k > 0 \\ (m,n+1,n) & k = 0, m > n \\ (m+1,0,0) & k = 0, m = n \end{cases} \tag{4.1}$$

and set $\mathbb{T}^{oq} = \{T_r \mid r \in \mathbb{N}_0\}$. So, $\mathbb{T}^{oq}$ consists of all "allowed" triples for overlapped cubes. We may denote components of T_r by m_r, n_r, and k_r respectively.

Proposition 4.1. For $r, s \in \mathbb{N}_0$, if $r < s$ then $\#\mathrm{OQ}(T_r) < \#\mathrm{OQ}(T_s)$.

Proof. It is sufficient to show that for $r \in \mathbb{N}_0$, $\#\mathrm{OQ}(T_r) < \#\mathrm{OQ}(T_{r+1})$. Suppose that $T_r = (m, n, k)$. If $k > 0$, then

$$\#\mathrm{OQ}(T_r) = 2^m + 2^n - 2^k < 2^m + 2^n - 2^{k-1} = \#\mathrm{OQ}(T_{r+1}). \tag{4.2}$$

If $k = 0$ and $m > n$, then $2^{n+1} - 2^n = 2^n > 2^n - 1$ and

$$\#\mathrm{OQ}(T_r) = 2^m + 2^n - 2^0 < 2^m + 2^{n+1} - 2^n = \#\mathrm{OQ}(T_{r+1}). \tag{4.3}$$

If $k = 0$ and $m = n$, then

$$\#\mathrm{OQ}(T_r) = 2^m + 2^m - 2^0 < 2^{m+1} + 2^0 - 2^0 = \#\mathrm{OQ}(T_{r+1}). \tag{4.4}$$

□

Therefore, $\#\mathrm{OQ}(T_r)$ as a function of r, provides a strictly increasing one-to-one correspondence between $\mathbb{N}_0$ and the set of all OQ-numbers. It is straightforward to check that there are $\frac{1}{6}(m^3 + 5m)$ OQ-numbers below $\#\mathrm{OQ}_{m,0,0}$. Also, there are $\frac{1}{2}m(m+1)$ overlapped cubes whose size is between sizes of Q_m and Q_{m+1}, which intuitively provide better scalability rather than hypercube networks, as it is seen in Fig. (8).

For a given natural number t, we need an ocube with at least t nodes. To find such an ocube, let $m = \lfloor \log_2 t \rfloor$. If $t = 2^m$, set $n = k = 0$, otherwise set $n = \lfloor \log_2(t - 2^m) \rfloor + 1$, and $k = \lfloor \log_2(2^m + 2^n - t) \rfloor$. Then $\#\mathrm{OQ}_{m,n,k}$ is the least OQ-number greater than t or equal to it.

Now, we investigate paths and diameter of ocubes. From definition (4.1), we know that each node in $\mathrm{OQ}_{m,n,k}$ is a zero-one string of length $m + n - k$. We split such a node x as $x_0 \oplus \bar{x} \oplus x_1$, where x_0, $\bar{x}$, and x_1 are of lengths $n - k$, k, and $m - k$ respectively. For two nodes x, y in $\mathrm{OQ}_{m,n,k}$, we may work on first, second, and third part respectively to find a path between x and y. There are several cases:

- If x, y are both in Q_m^0, then $x_1 = y_1$, at most $m - k$ changes are needed for converting x_0 to y_0, and then at most k changes are needed for converting $\bar{x}$ to $\bar{y}$. So at most $m - k + k = m$ changes are needed for converting x to y.
- If x, y are both in Q_n^1, then $x_0 = y_0$, at most k changes are needed for converting $\bar{x}$ to $\bar{y}$, and then at most $n - k$ changes are needed for converting x_1 to y_1. So at most $k + n - k = n$ changes are needed for converting x to y.

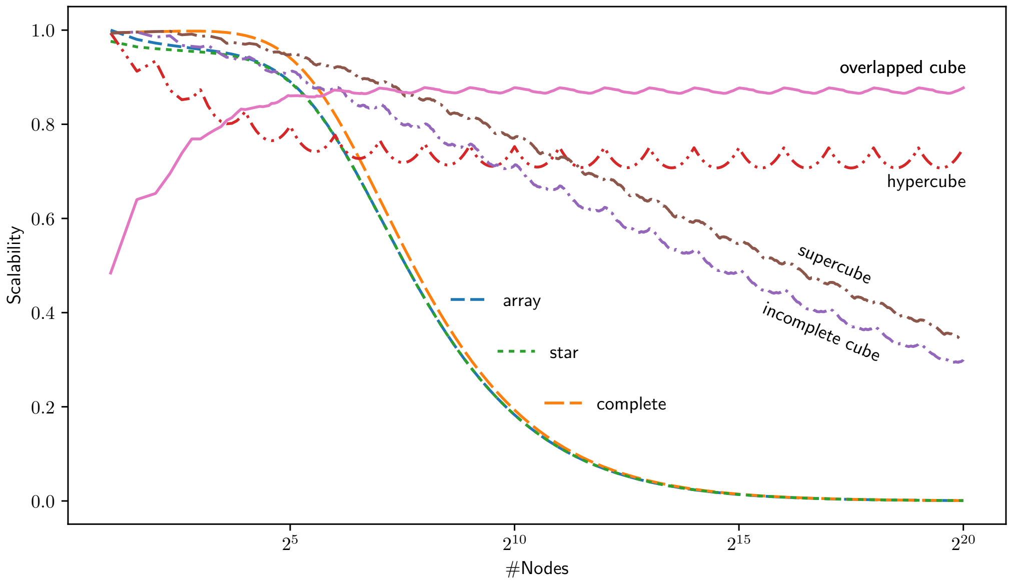

Fig. 8. Scalability with the same parameters as in Fig. 4, with overlapped cubes.

- For x in Q_m^0 and y in Q_n^1, at most $m-k$ changes are needed for converting x_0 to y_0, then k changes are needed for converting $\bar{x}$ to $\bar{y}$, and then at most $n-k$ changes are needed for converting x_1 to y_1. So at most $m-k+k+n-k = m+n-k$ changes are needed for converting x to y.

Note that, in each case, all intermediate nodes remain in $\mathrm{OQ}_{m,n,k}$. The process gives a routing algorithm in $\mathrm{OQ}_{m,n,k}$, and more, the following proposition is proved.

Proposition 4.2. The diameter of $\mathrm{OQ}_{m,n,k}$ is $m+n-k$.

As it may be seen, the above procedure is a natural extension of the same procedure for cubes. Ocubes still inherit more properties from their hypercube ancestors. Note that Q_r is $(r-1)$-fault-tolerant.

Proposition 4.3. $\mathrm{OQ}_{m,n,k}$ is $(n-1)$-fault-tolerant for $k \geq 1$.

Proof. Since Q_n^1 as an induced sub-network of $OQ_{m,n,k}$ is $(n-1)$-fault-tolerant, $OQ_{m,n,k}$ is $(n-1)$-fault-tolerant. □

We close this section with some asymptotical properties of the average shortest path in ocubes.

Proposition 4.4. $\bar{d}(\mathrm{OQ}_{m,n,k})$ is asymptote to $\frac{1}{2}m$ from below. It is asymptote to $\frac{3}{4}m$ from above.

Proof. Trivially, $\bar{d}(\mathrm{OQ}_{m,n,k})$ is decreasing with respect to k, and is increasing with respect to n; so

$$\bar{d}(\mathrm{OQ}_{m,0,0}) \leq \bar{d}(\mathrm{OQ}_{m,n,k}) \leq \bar{d}(\mathrm{OQ}_{m,m,0}). \tag{4.5}$$

Consider a $\mathrm{OQ}_{m,m,0}$ with two Q_m copies Q_m^0 and Q_m^1 connected by a node q. Fixing a node x in Q_m^0, there are $\binom{m}{r}$ nodes y with $d(x,y) = r$. So

$$\begin{aligned} \sum_y d(x,y) &= \sum_{r=1}^{m} r\binom{m}{r} && (4.6)\\ &= \sum_{r=1}^{m} m\binom{m-1}{r-1} && (4.7)\\ &= m \times 2^{m-1} && (4.8) \end{aligned}$$

Now, take x and y in the first and the second copy except q respectively. Since $d(x,y)$ may be split as $d(x,q) + d(q,y)$, the number of such pairs (x,y) with $d(x,y) = l$ is

$$\sum_{\substack{r+s=l\\ r,s\geq 1}} \binom{m}{r}\binom{m}{s} = \binom{2m}{l} - 2\binom{m}{l}. \tag{4.9}$$

Thus the sum of all distances between two copies is equal to

$$\begin{aligned} 2\sum_{l=2}^{2m}\left(l \times \sum_{\substack{r+s=l\\ r,s\geq 1}} \binom{m}{r}\binom{m}{s}\right) &= 2\sum_{l=2}^{2m}\left(l\binom{2m}{l} - 2l\binom{m}{l}\right) && (4.10)\\ &= 2\sum_{l=2}^{2m}\left(2m\binom{2m-1}{l-1} - 2m\binom{m-1}{l-1}\right) && (4.11)\\ &= 2 \times 2m\Big(\big(2^{2m-1} - 1\big) - \big(2^{m-1} - 1\big)\Big) && (4.12)\\ &= 2m\big(2^{2m} - 2^m\big). && (4.13) \end{aligned}$$

So, by Eq. (4.8), we have

$$\begin{aligned} \bar{d}(\mathrm{OQ}_{m,m,0}) &= \frac{2 \times m \times 2^{m-1} \times 2^m + 2m \times \big(2^{2m} - 2^m\big)}{\big(2^{m+1} - 1\big)\big(2^{m+1} - 2\big)} && (4.14)\\ &= \frac{3m \times 2^{2m} - 2m \times 2^m}{\big(2^{m+1} - 1\big)\big(2^{m+1} - 2\big)}. && (4.15) \end{aligned}$$

Therefore, by Proposition (2.2), we have

$$\varliminf_{r\to\infty} \frac{\bar{d}\big(\mathrm{OQ}(T_r)\big)}{m_r} = \lim_{m\to\infty} \frac{\bar{d}(\mathrm{OQ}_{m,0,0})}{m} = \frac{1}{2}, \tag{4.16}$$

and

$$\overline{\lim_{r\to\infty}} \frac{\bar{d}\big(\mathrm{OQ}(T_r)\big)}{m_r} = \lim_{m\to\infty} \frac{\bar{d}(\mathrm{OQ}_{m,m,0})}{m} = \frac{3}{4}. \tag{4.17}$$

□

5. Conclusion

In this paper, we introduced a measure for the cost-performance scalability of interconnection networks considering the network size, diameter and degree, and the variation of node degrees. We used the proposed scalability measure for comparing different popular basic topologies and cube-based network topologies.

We then introduced a new interconnection network topology based on the hypercube, called the Overlapped Cube (OCube), and studied its topological properties and scalability behavior. The OCube is formed by two hypercubes which are joined together by a hypercube.

Our findings showed that OCubes are more cost-performance scalable compared to other cube-based interconnection networks, revealing their potential for better scalability while preserving the main cubical properties.

References

1. G.-B. Chae, E. M. Palmer, and R. W. Robinson. Counting labeled general cubic graphs. Discrete Mathematics, 307(23):2979–2992, 2007.
2. H.-Y. Chang and R.-J. Chen. Incrementally extensible folded hypercube graphs. In Proceedings 1998 International Conference on Parallel and Distributed Systems (Cat. No. 98TB100250), pages 566–571. IEEE, 1998.
3. G.-H. Chen and H.-L. Huang. Cube-connected modules: a family of cubic networks. In Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN), pages 57–64. IEEE, 1994.
4. Q. Dong, X. Yang, J. Zhao, and Y. Y. Tang. Embedding a family of disjoint 3d meshes into a crossed cube. Information Sciences, 178(11):2396–2405, 2008.
5. D.-R. Duh, G.-H. Chen, and D. F. Hsu. Combinatorial properties of generalized hypercube graphs. Information Processing Letters, 57(1):41–45, 1996.
6. K. Efe. A variation on the hypercube with lower diameter. IEEE Transactions on Computers, 40(11):1312–1316, 1991. ⊗
7. K. Efe. The crossed cube architecture for parallel computation. IEEE Transactions on Parallel & Distributed Systems, 3(05):513–524, 1992.
8. A. El-Amawy and S. Latifi. Properties and performance of folded hypercubes. IEEE Transactions on Parallel & Distributed Systems, 2(01):31–42, 1991.
9. K. Ghose and K. R. Desai. Hierarchical cubic networks. IEEE Transactions on Parallel and Distributed Systems, 6(4):427–435, 1995.
10. F. Harary, J. P. Hayes, and H.-J. Wu. A survey of the theory of hypercube graphs. Computers & Mathematics with Applications, 15(4):277–289, 1988.

11. J. P. Hayes, T. N. Mudge, Q. F. Stout, S. Colley, and J. Palmer. Architecture of a hypercube supercomputer. In ICPP, pages 653–660, 1986.
12. S.-Y. Hsieh and T.-J. Lin. Panconnectivity and edge-pancyclicity of k-ary n-cubes. Networks: An International Journal, 54(1):1–11, 2009.
13. L.-H. Hsu and C.-K. Lin. Graph theory and interconnection networks. CRC press, 2008.
14. A. Karci. Hierarchical extended fibonacci cubes. Iranian Journal of Science and Technology Transaction B: Engineering, 29(1):117–125, 2005.
15. A. Karci. Hierarchic graphs based on the fibonacci numbers. IU-Journal of Electrical & Electronics Engineering, 7(1):345–365, 2007.
16. H. P. Katseff. Incomplete hypercubes. IEEE transactions on computers, 37(5):604–608, 1988.
17. Z. A. Khan, J. Siddiqui, and A. Samad. Linear crossed cube (lcq): A new interconnection network topology for massively parallel system. International Journal of Computer Network and Information Security, 7(3):18–25, 2015.
18. M. S. Krishnamoorthy and B. Krishnamurthy. Fault diameter of interconnection networks. Computers & Mathematics with Applications, 13(5-6):577–582, 1987.
19. T.-L. Kung and C.-K. Lin. Cluster connectivity of hypercube-based networks under the super fault-tolerance condition. Discrete Applied Mathematics, 293:143–156, 2021.
20. C.-J. Lai, C.-H. Tsai, H.-C. Hsu, and T.-K. Li. A dynamic programming algorithm for simulation of a multi-dimensional torus in a crossed cube. Information Sciences, 180(24):5090–5100, 2010.
21. F. T. Leighton. Introduction to parallel algorithms and architectures: Arrays · trees · hypercubes. Elsevier, 2014.
22. G. Lemieux and D. Lewis. Design of interconnection networks for programmable logic, volume 22. Springer, 2004.
23. A. Louri and H. Sung. Scalable optical hypercube-based interconnection network for massively parallel computing. Applied optics, 33(32):7588–7598, 1994.
24. W. Mao and D. M. Nicol. On k-ary n-cubes: theory and applications. Discrete applied mathematics, 129(1):171–193, 2003.
25. O. Michael and O. Alexander. Analysis for computer scientists: Foundations, methods, and algorithms, 2018.
26. M. Monemizadeh and H. Sarbazi-Azad. The necklace-hypercube: a well scalable hypercube-based interconnection network for multiprocessors. In Proceedings of the 2005 ACM symposium on Applied computing, pages 729–733, 2005.
27. J. Nieminen, M. Peltola, and P. Ruotsalainen. On graphs like hypercubes. Tsukuba journal of mathematics, 32(1):37–48, 2008.
28. B. Parhami and D.-M. Kwai. Comparing four classes of torus-based parallel architectures: Networkparameters and communication performance. Mathematical and Computer Modelling, 40(7-8):701–720, 2004.
29. Y. Saad and M. H. Schultz. Topological properties of hypercubes. IEEE Transactions on Computers, 37(7):867–872, 1988.
30. B. Selçuk and A. Karci. Connected cubic network graph. Engineering Science and Technology, an International Journal, 20(3):934–943, 2017.
31. A. Sen. Supercube: An optimally fault tolerant network architecture. Acta Informatica, 26:741–748, 1989.

32. P. Shareghi and H. Sarbazi-Azad. The stretched network: Properties, routing, and performance. Journal of Information Science & Engineering, 24(2), 2008.
33. J.-J. Sheu and L.-H. Hsu. Fault diameter for supercubes. Parallel Processing Letters, 9(01):21–30, 1999.
34. J.-J. Sheu, J. J. Tan, and L.-H. Hsu. Routing properties of supercubes. Information Sciences, 131(1-4):107–128, 2001.
35. S. Sur and P. K. Srimani. Ieh graphs: A novel generalization of hypercube graphs. Acta informatica, 32:597–609, 1995.
36. J. Xu. Topological structure and analysis of interconnection networks, volume 7. Springer Science & Business Media, 2013.
37. S.-M. Yuan. Topological properties of supercube. Information Processing Letters, 37(5):241–245, 1991.
38. X. Zhu. A hypercube variant with small diameter. Journal of Graph Theory, 85(3):651–660, 2017. ⊗