Optimal circulant graphs as low-latency network topologies

Xiaolong Huang[1] · **Alexandre F. Ramos**[2] · **Yuefan Deng**[1,3]

Abstract

Communication latency has become one of the determining factors for the performance of parallel clusters. To design low-latency network topologies for high-performance computing clusters, we optimize the diameters, mean path lengths, and bisection widths of circulant topologies. We obtain a series of optimal circulant topologies of size 2^5 through 2^{10} and compare them with torus and hypercube of the same sizes and degrees. We further benchmark on a broad variety of applications including effective bandwidth, FFTE, Graph 500 and NAS parallel benchmarks to compare the optimal circulant topologies and Cartesian products of optimal circulant topologies and fully connected topologies with corresponding torus and hypercube. Simulation results demonstrate superior potentials of the optimal circulant topologies for communication-intensive applications. We also find the strengths of the Cartesian products in exploiting global communication with data traffic patterns of specific applications or internal algorithms.

Keywords Network topology · Circulant graph · Latency · Optimization · Benchmarks

✉ Yuefan Deng
    yuefan.deng@stonybrook.edu

    Xiaolong Huang
    xiaolong.huang@stonybrook.edu

    Alexandre F. Ramos
    alex.ramos@usp.br

1   Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794, USA

2   Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Avenida Arlindo Béttio 1000, São Paulo, SP CEP 03828-000, Brazil

3   Mathematics, Division of Science, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

1 Introduction

The development of supercomputers has been advancing rapidly for the past five decades with the most recent supercomputer Fugaku culminating the Top 500 list in November 2021 at a peak speed of 537 PFlops and 7,630,848 cores [1]. Along with the growth of number of faster processors – as suggested by the Moore's law [2], the interconnection network that can efficiently hold and connect such massive systems has been constantly on demand. Due to the power wall from the Moore's law [2], however, the increase of processor's clock speed has hit its barrier. Consequently, the development of ever-expanding supercomputers with ever-increasing speed leans on the design of optimally interconnected networks topologies. Generally, ideal interconnection networks should satisfy small network diameter, large bisection width, topological simplicity, symmetry, engineering feasibility, and modular, expandable design, to provide maximal connectivity, scalable performance, minimal engineering complexity and least monetary cost [3]. Meshes [4], tori [5–10], hypercubes [11], fat trees [12, 13] and multistage switched fabrics [4, 14] have been the mainstream networks for years. Meshes have simple layout but the large diameter slows down node-to-node communications. Fat trees can realize maximum bisection width, but the diameter grows with the number of levels of switches. The torus and its variation k-ary n-cube [15] have relatively smaller diameter and average distance. 3D torus is applied in Cray SeaStar [5] and IBM Blue Gene/P [6], while modified 3D torus formulates Cray Gemini interconnect [7]. 5D torus is incorporated in IBM Blue Gene/Q [8]. Hybrid 6D mesh/torus Tofu interconnect is integrated in K and post-K supercomputers [9, 10]. Recently, Dragonfly topology [16] among hierarchical topologies with small diameter has been deployed into Aries and Slingshot interconnects [17, 18].

In this manuscript, we propose the optimal circulant topologies with minimal diameter and mean path length (MPL) and maximum bisection width as promising low-latency network topologies by a home-grown and highly efficient parallel exhaustive search algorithm to generate and optimize circulant topologies. We obtain a set of optimal circulant topologies of size 2^5 through 2^{10}. By comparing with corresponding torus and hypercube of the same sizes and degrees, we show remarkable improvement in graph properties by the optimal circulant topologies and the Cartesian products of optimal circulant topologies and fully connected topologies. The benchmarking results by simulation also demonstrate significant performance enhancement for communication-intensive applications. In the meantime, the Cartesian products of optimal circulant topologies and fully connected topologies show potential in balancing the influence of global communication with data traffic patterns of specific applications or internal algorithms. We discuss related work in Sect. 2 and present the parallel optimization algorithm for circulant topologies in Sect. 3. Section 4 examines graph properties of optimal circulant topologies and Cartesian products by comparative analysis with torus and hypercube. Section 5 shows simulated benchmark performance and detailed analysis. We conclude the manuscript in Sect. 6.

2 Related work

The optimization of interconnection networks is essentially to reduce the node-to-node communication latency, which is directly indicated by the diameter and mean path length (MPL) of the topology. The idea of optimizing MPL can be dated back to Cerf et al., who first calculated the lower bound of MPL given arbitrary regular graph size and degree [19]. Zhang et al. modified torus by adding bypass links to construct interlaced bypass torus (iBT) [20, 21], designed efficient broadcast algorithms on iBT [22] and evaluated its performance by building a re-configurable cluster and system [23]. Feng et al. further explored 6D mesh/iBT on custom routing algorithms [24] and performance evaluation by simulation [25]. Deng et al. implemented parallel exhaustive search of regular graphs [26] and random search with rotational symmetry to obtain small and large (near) optimal topologies and evaluated the performance by benchmarking both on real clusters and simulation platform [27]. More small optimal topologies with symmetries and other special properties were produced and presented in a structured table [28]. Distributed shortcut networks targeting the diameter and cable length trade-off [29] and host-switch graphs designed by minimizing diameter and/or MPL using randomized heuristics [30] were also introduced and benchmarked by simulation.

In addition, network topologies with additional symmetry properties provide advantages in routing and load balancing [4] and also in design and engineering simplicity. Recent research on constructing supercomputer network utilizing Lie algebra symmetry [3, 31] and optimizing MPL by random or heuristic algorithms with rotational symmetry constraints [27, 32] have demonstrated the importance of the symmetry perspective in network design. The total number of symmetries of a graph calculated as its automorphism group size has also been included in further optimization of smaller topologies [28]. Vertex-symmetric topologies are one class of highly symmetric topologies, where any two vertices are equivalent under self mappings that preserve adjacency [4]. Torus consisting of equal-sized rings, hypercube and dragonfly are notable examples of vertex-symmetric topologies that have been widely applied in current mainstream high-performance computing networks as introduced in Sect. 1.

Among vertex-symmetric topologies, circulant topologies, also known as multi-loop networks, have been extensively explored in both theory and applications as summarized in three major surveys [33–35]. The lower bounds of diameter and MPL for circulant topologies have been estimated in theory [34]. For degree-4 (double-loop) circulants, such lower bounds can be achieved [36, 37] and their layouts [38], routing [39, 40] and broadcast algorithms [41] have been studied. However, for arbitrary size and degree, the existence of such optimal circulant topologies that reach theoretical lower bounds of diameter and MPL still remains open.

The synthesis of general arbitrary-degree circulant topologies with small diameter and/or MPL has been conducted by both theoretical construction and computer search [35]. Population-based heuristic algorithms [42] have been analyzed

when optimizing MPL for circulants with given size, but may only obtain near-optimal graph properties comparing with exhaustive search. Techniques such as tree-like search guided by girth heuristic [43] and direct product construction [44, 45] have been developed to solve the related degree/diameter problem, which is to maximize the size of circulant topologies given fixed diameter and degree. To the best of our knowledge, there are only a few parallel algorithms for the synthesis and optimization of arbitrary-degree circulant topologies, and even less work on the performance evaluation of circulant topologies as communication networks. Our optimization method of efficient parallel exhaustive search provides another approach in finding all the definitive optimal circulant topologies given graph size and degree. We further select the optimal circulant topology with the largest bisection width to add to its potential in performance enhancement. Moreover, our simulation results bring insight into the actual performance of the optimal circulant topologies for computational applications.

3 Discovery of optimal circulant topologies

An N-vertex circulant graph [46] is defined by a jump set $S = \{s | 1 \leq s \leq \lfloor N/2 \rfloor\}$, where each vertex $i = 0, 1, ..., N-1$ is connected to vertices $i \pm s \pmod{N}$ for every jump s. A degree-k circulant graph has $\lceil k/2 \rceil$ number of jumps. In particular, the rings and complete graphs are both circulant, when $S = \{1\}$ and $S = \{1, 2, ..., \lfloor N/2 \rfloor\}$ respectively. Examples of 16-vertex circulant graphs are shown in Fig. 1, which are optimal as we discovered.

We search optimal circulant graphs of minimal diameter (D), minimal mean path length (MPL), and maximal bisection width (BW). To exhaustively search through all N-vertex degree-k circulant graphs, denoted by (N, k), is equivalently to compute all combinations of choosing $\lceil k/2 \rceil$ jumps from 1 to $\lfloor (N-1)/2 \rfloor$. Jump $N/2$ is

(a) $(16, 4)$ Optimal Circulant, Jumps $= 1, 6$ (b) $(16, 5)$ Optimal Circulant, Jumps $= 1, 3, 8$
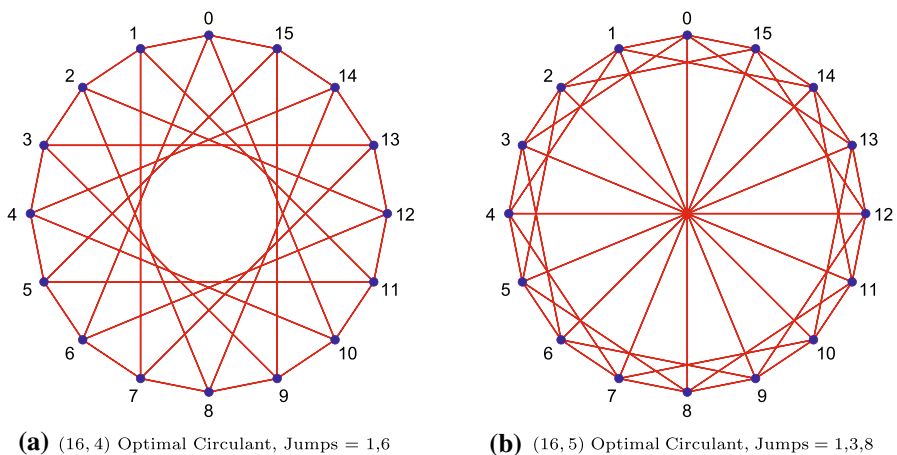
Fig. 1 16-vertex optimal circulant graphs

excluded as the jump set S contains $N/2$ if and only if k is odd. Since we require the circulant graphs to be connected, by theorem of Boesch and Tindell [47], a circulant graph is connected if and only if $\gcd(N, S) = 1$. Thus, when N is a power of 2, there must be an odd jump. Such an odd jump is also relatively prime to N, which can be mapped to jump 1 in an equivalent jump set by Adam isomorphism of the circulant graph [47]. As a result, for N equal power of 2, we can set one of the jumps to 1 and reduce the exhaustive search to choosing $\lfloor k/2 \rfloor - 1$ jumps from 2 to $\lfloor (N-1)/2 \rfloor$. For instance, the numbers of (1024, 10) and (1024, 15) circulant graphs to optimize can be considerably reduced from $\sim 3 \cdot 10^{11}$ to $\sim 3 \cdot 10^{9}$ and from $\sim 2 \cdot 10^{15}$ to $\sim 2 \cdot 10^{13}$ respectively.

To effectively perform the exhaustive search of optimal circulant graphs, we adapt the co-lexicographic ordering algorithm of combinations in the FXT library, which is a highly efficiently library in combinatorial generation [48, 49]. One key feature of the algorithm is that the generation of each combination only depends on the previous combination in co-lexicographic order. Therefore, the generation process can be reduced to a minimum number of operations which makes it as fast as possible. Moreover, such feature enables the parallelization of the total combinatorial generation by incorporating with an unranking algorithm [50]. In the co-lexicographic ordering, each combination corresponds to a unique ordinal number which is called the rank of the combination. The unranking algorithm maps a rank back to its corresponding combination in co-lexicographic order, which allows the generation to start and end at arbitrary ranks.

Utilizing the above combinatorial techniques including the reduction of search space, the co-lexicographic ordering algorithm and the unranking algorithm, we build our own parallel exhaustive search program for circulant graphs. Given (N, k) as input, we design our parallel exhaustive search algorithm to start by counting the total number of (N, k) circulant graphs to optimize, and divide the number evenly among all processors. Then each processor unranks its start rank, generates circulant graphs in the format of jump sets by enumerating through combinations in co-lexicographic order while filtering by finding the ones with minimal diameter and then minimal MPL, and stops at its end rank. For the calculation of diameter and MPL, we run breadth-first search only once on each graph since circulant graphs are vertex-symmetric. For N equal power of 2, we use bit arrays and bit shifts to mark traversed vertices in the implementation of breadth-first search to further improve its speed. Finally, all processors communicate only once to find the overall minimal diameter and MPL. Then each processor filters once more to obtain the optimal circulant graphs. Hence our parallel exhaustive search algorithm has perfectly balanced workloads, minimal communication and synchronization delay, and highly efficient implementations of both combinatorial generation and calculation of diameter and MPL. We execute the parallel exhaustive search on the SeaWulf supercomputer at Stony Brook University. The processing speed on compute nodes with Intel Xeon Gold 6148 CPUs can reach $\sim 10^{5}$ graphs per second per core, which makes it practical to optimize large high-degree circulant graphs even up to (1024, 15) by exhaustive search.

When there are multiple optimal circulant graphs with minimal diameter and MPL, we filter further by optimizing bisection width. To compute the bisection

width of a graph, we use the KaHIP program which can achieve strictly balanced bisection as required [51]. We also eliminate isomorphic optimal circulant graphs using the nauty and Traces programs [52]. We present the discovered optimal circulant graphs (in Table 1) which have minimal diameters and MPLs and maximum bisection widths filtered in such specific order. We believe they serve as promising candidates for interconnection network topologies.

4 Graph analysis of optimal circulant topologies

We analyze the graph properties of optimal circulant topologies by comparing with torus and hypercube as well as Cartesian products of one smaller optimal circulant topology and the 4-vertex fully connected topology. We first calculate and present the diameter, MPL and bisection width of the three categories of topologies in Table 2. For each N from 2^5 to 2^{10}, single optimal circulant topologies are denoted as Optimal Circulant, while Cartesian products of optimal circulant topologies are denoted as Optimal Circulant Product. For tori and Cartesian products, the product components are placed from small to large starting on the right. An example of Cartesian product of optimal circulant topologies is shown in Fig. 2. We also calculate the number of edges $|E| = Nk/2$ for all topologies in Table 2.

Moreover, for a given graph size we further normalize the diameter, MPL and bisection width in Table 2 by comparing all topologies with corresponding torus and compute the graph property ratios. For diameter and MPL, the ratios are inverted as $D_{Torus}/D_{Topology}$ and $MPL_{Torus}/MPL_{Topology}$. For bisection width, the ratio stays as $BW_{Topology}/BW_{Torus}$. All (inverse) ratios are shown in Fig. 4b–d by bar plots, such that higher bars always represent better graph properties. In the end, for each graph category, we show the average (inverse) ratios of graph properties across size

Table 1 Optimal circulant graphs discovered by parallel exhaustive search

N	k	Jumps
32	4	1,7
	5	1,6,16
64	4	1,14
	5	1,7,32
	6	1,4,25
128	6	1,8,54
	7	1,12,30,64
256	6	1,47,122
	7	1,13,33,128
	8	1,9,74,103
512	8	1,15,56,149
	9	1,23,31,119,256
1024	8	1,144,258,276
	10	1,38,70,393,481

Fig. 2 (32, 5) Optimal Circulant Product (8,2)×(4,3)
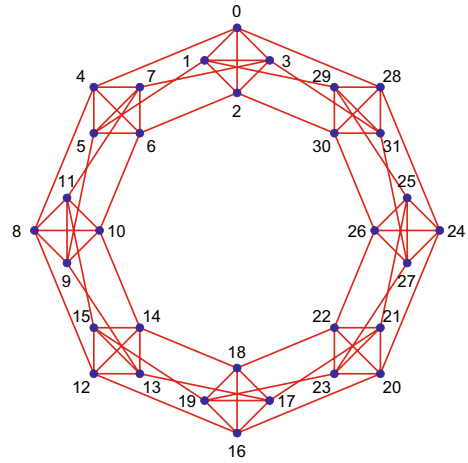
N in Fig. 4a. We use low-degree and high-degree to denote the two different graph degrees for each N as in Table 2, and use consistent color scheme as in Fig. 3 to represent different topology categories in all bar plots Figs. 4, 5, 6, 7, 8, 9 and 10.

Table 2 and Fig. 4 show that the optimal circulant topologies and Cartesian products always have smaller diameter, MPL and larger bisection width than corresponding torus. The low-degree optimal circulant topologies and Cartesian products have average diameter inverse ratios 1.58/1.68 and average MPL inverse ratios 1.18/1.24 respectively comparing with torus, indicating 37%/40% smaller diameter and 15%/19% smaller MPL in average. Meanwhile, the high-degree optimal circulant topologies have further smaller diameter and MPL, 52% and 30% decrease in average comparing with torus. For bisection width, the low-degree optimal circulant topologies and Cartesian products have clear advantages over torus, and almost as good as hypercube, while the high-degree optimal circulant topologies can achieve up to 235% average increase. We also observe that for the optimal circulant topologies and Cartesian products, both the diameter inverse ratios and bisection width ratios are increasing along with the topology size N with slight fluctuation, while the MPL inverse ratios mostly stay around its average.

5 Benchmarks of optimal circulant topologies

5.1 Simulation platform and network routing

To investigate the performance of optimal circulant topologies, we simulate the topologies listed in Table 2 on SimGrid (version 3.26) [53]. SimGrid offers flexible and accurate parallel simulation platform, especially the SMPI interface which enables simulation of MPI applications [53]. We configure the cluster parameters in SimGrid by setting 100 GFlops computational speed per core of single-core nodes, 10 Gbps network bandwidth and 40 ns latency per link. In the cluster the computing

Table 2 Graph properties of optimal circulant (product) topologies and tori/hypercubes

N	k	Topology	D	MPL	BW	|E|
32	4	Torus 2D (8,2)×(4,2)	6	3.10	8	64
		Optimal Circulant	4	2.71	16	
	5	Optimal Circulant	3	2.29	20	80
		Optimal Circulant Product (8,2)×(4,3)	5	2.84	8	
		Hypercube 5D	5	2.58	16	
64	4	Torus 2D (8,2)×(8,2)	8	4.06	16	128
		Optimal Circulant	6	3.78	22	
	6	Optimal Circulant	4	2.60	48	192
		Optimal Circulant Product (16,3)×(4,3)	5	3.24	16	
		Hypercube 6D	6	3.05	32	
128	6	Torus 3D (8,2)×(4,2)×(4,2)	8	4.03	32	384
		Optimal Circulant	5	3.34	76	
	7	Optimal Circulant	4	3.02	100	448
		Optimal Circulant Product (32,4)×(4,3)	5	3.40	64	
		Hypercube 7D	7	3.53	64	
256	6	Torus 3D (8,2)×(8,2)×(4,2)	10	5.02	64	768
		Optimal Circulant	6	4.25	120	
	8	Optimal Circulant	5	3.33	234	1024
		Optimal Circulant Product (64,5)×(4,3)	5	3.78	128	
		Hypercube 8D	8	4.02	128	
512	8	Torus 4D (8,2)×(4,2)×(4,2)×(4,2)	10	5.01	128	2048
		Optimal Circulant	6	4.04	392	
	9	Optimal Circulant	5	3.73	472	2304
		Optimal Circulant Product (128,6)×(4,3)	6	4.07	304	
		Hypercube 9D	9	4.51	256	
1024	8	Torus 4D (8,2)×(8,2)×(4,2)×(4,2)	12	6.01	256	4096
		Optimal Circulant	7	4.88	624	
	10	Optimal Circulant	5	4.07	1036	5120
		Optimal Circulant Product (256,7)×(4,3)	6	4.52	624	
		Hypercube 10D	10	5.00	512	

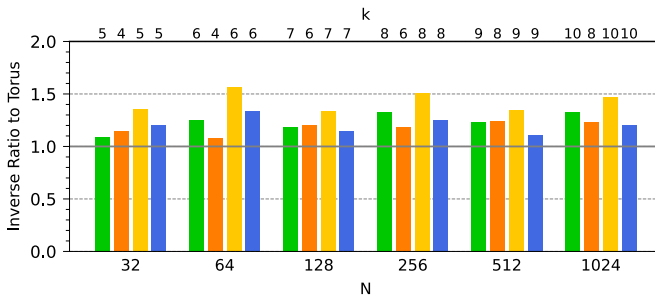**Fig. 3** Color scheme to represent different topology categories in all bar plots

■ Low-Degree Optimal Circulant
■ High-Degree Optimal Circulant
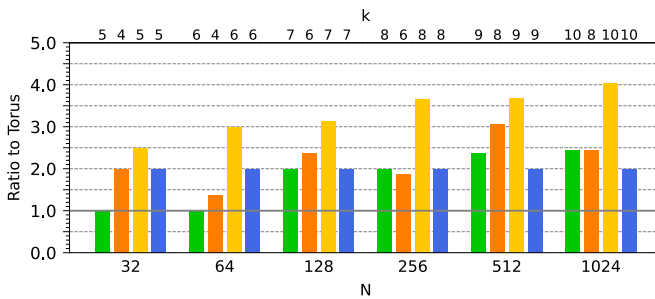■ High-Degree Optimal Circulant Product
■ Hypercube
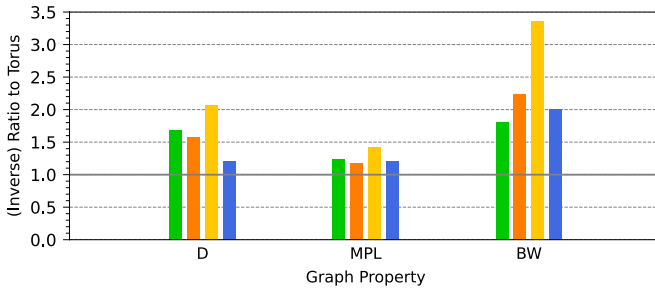
(a) Graph properties average (inverse) ratios
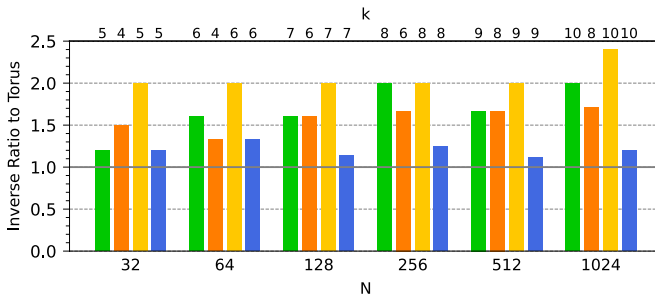
(b) D inverse ratios

(c) MPL inverse ratios

(d) BW ratios

Fig. 4 Graph properties measured in (inverse) ratios to torus

nodes are set to be directly connected via input topologies. We utilize the built-in implementation of MVAPICH2 in SimGrid as the MPI library. All the simulations are run on the SeaWulf supercomputer at Stony Brook University.

For network routing, we use static shortest-path routing with full routing table for all simulated topologies with a custom-designed vertex-symmetric routing for the optimal circulant topologies. To determine the routing table, we first apply breadth-first search once and find the initial routes from node 0 to all the other nodes. Then we add $i \pmod{N}$ to all the nodes on the initial routes to form the routes from node i to all the other nodes. In this way we can construct the full routing table for the optimal circulant topologies. For torus, hypercube, and Cartesian products, we implement the widely used dimension-order routing which achieves shortest distances between nodes in Cartesian product topologies [4].

5.2 Benchmark results and analysis

The following benchmark programs are simulated to examine and compare the performance of network topologies: effective bandwidth (b_eff) [54, 55], FFTE [56, 57], Graph 500 [58, 59] and the NAS Parallel Benchmarks (NPB) [60, 61] consisting of FT, IS, CG, MG, LU, BT and SP. We run each benchmark on all topologies listed in Table 2 and evaluate the performance by ratio of processing speeds to torus of the same size. The performance ratio is calculated as $S_{\text{Topology}}/S_{\text{Torus}}$, where S is the average speed over multiple runs reported by the benchmark. All performance ratios are demonstrated by bar plots in Figs. 5, 6, 7, 8, 9 and 10. The terms low-degree and high-degree are used to denote the two different graph degrees for each N. The same color scheme as in Fig. 3 is used to represent different topology categories in all bar plots. Like the comparative analysis of graph properties in Sect. 4, we also present in Fig. 5 the average performance ratios across topology size N for each benchmark.

Figure 5 shows that comparing with torus, the low-degree and high-degree optimal circulant topologies and Cartesian products achieve average performance gains of 15%, 52% and 15% respectively over all simulated benchmarks. The performance ratios are much higher especially on effective bandwidth, FFTE, NPB FT and Graph 500, while reaching higher or equal average performance for the other
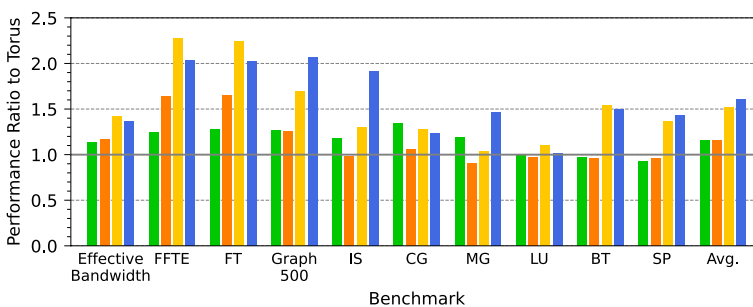

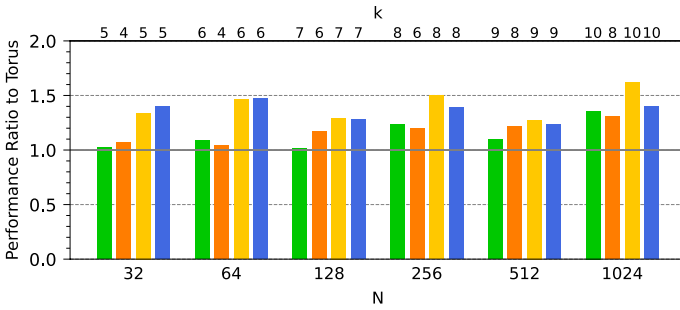**Fig. 5** Average performance measured in ratios to torus

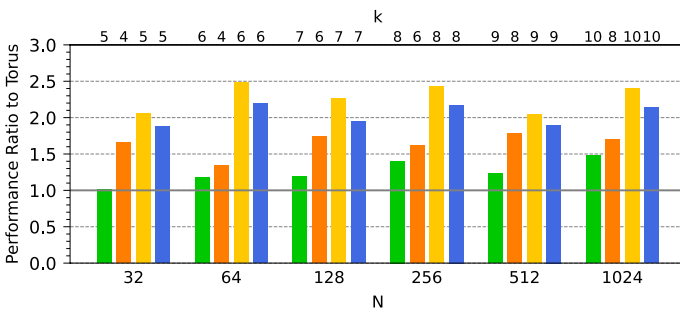**Fig. 6** Performance ratios to torus on effective bandwidth

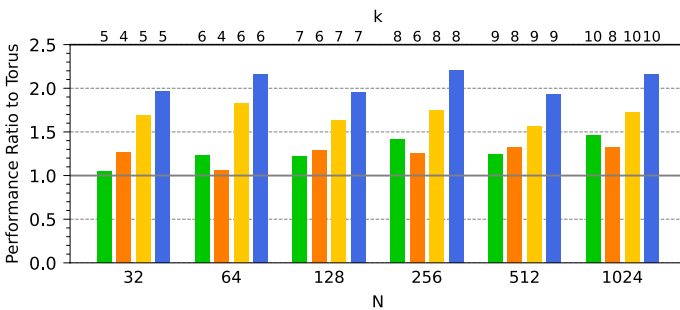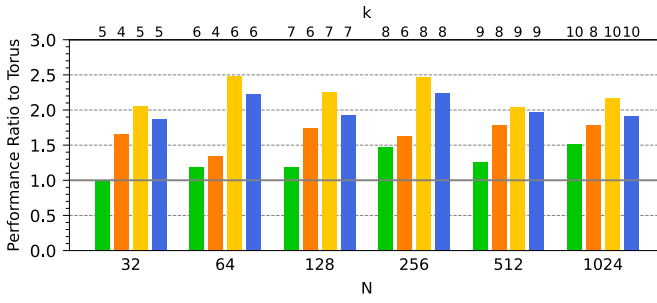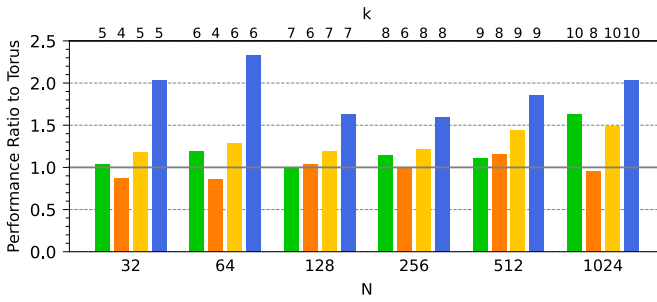
Fig. 7 Performance ratios to torus on 1D FFTE

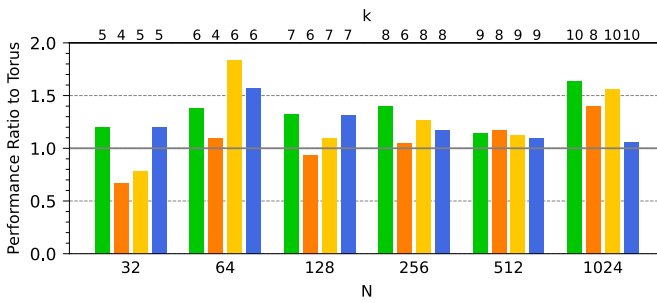**Fig. 8** Performance ratios to torus on Graph 500

NPB benchmarks. In particular, the high-degree optimal circulant topologies are capable of enhancing the performance up to 128% and 124% in average on FFTE and NPB FT respectively, even 24% and 21% higher than hypercube. We present more detailed analysis on the benchmarks and simulation results in the rest of this section.
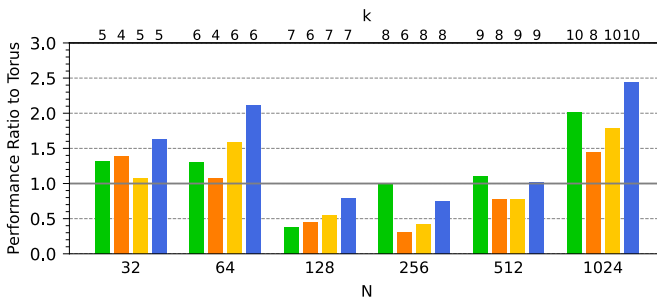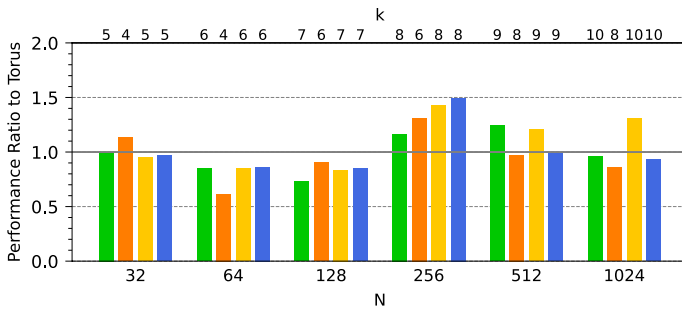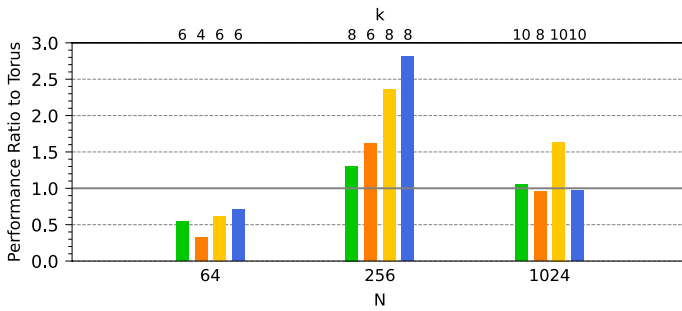
(a) NPB FT

(b) NPB IS

(c) NPB CG

(d) NPB MG

Fig. 9 Performance ratios to torus on NPB kernels

(a) NPB LU

(b) NPB BT
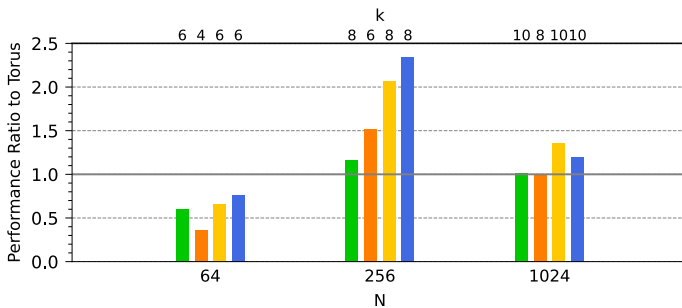
(c) NPB SP

Fig. 10 Performance ratios to torus on NPB pseudo applications

5.2.1 Effective bandwidth

Effective bandwidth (b_eff, version 3.6.0.1) [54, 55] measures the total network bandwidth over multiple ring and random communication patterns. It also compares three different communication methods: MPI_Sendrecv, MPI_Alltoallv and non-blocking MPI_Irecv and MPI_Isend with MPI_Waitall, and reports the maximum bandwidth of all methods. To fairly compare the performance of optimal circulant

topologies with torus, we modify the benchmark to run over random communication patterns only and collect the reported bandwidth at maximum message size of 1 MB per processor.

The effective bandwidth performance ratios to torus are plotted in Fig. 6. The high-degree optimal circulant topologies substantiate 42% average performance gain, reaching 62% at $N = 1024$ which is even 22% higher that hypercube. The low-degree optimal circulant topologies and Cartesian products show higher performance gain starting with larger topology size from $N = 256$, with 17%/14% in average and as high as 31%/34% respectively at $N = 1024$, almost equal to hypercube. Since effective bandwidth essentially measures cross-network traffic, optimal circulant topologies with small diameter and MPL and large bisection width therefore can significantly improve the performance.

5.2.2 FFTE

FFTE (version 7.0) [56, 57] is simulated, which requires global all-to-all data transpositions across the network. We perform the parallel 1D FFTE routine with transform array lengths ranging from 2^{10} to 2^{31}, equal to total transform array sizes from 16 KB to 32 GB. Then we collect the sustained computational speed at the maximum transform array length.

Figure 7 shows the 1D FFTE performance ratios to torus. The low-degree and high-degree optimal circulant topologies and Cartesian products have slightly fluctuating performance ratios, maintaining average performance gains of 64%, 128% and 25% respectively. Moreover, the high-degree optimal circulant topologies outperform hypercube for every topology size, adding to 24% higher average performance gain. The high performance gains result from the performance dependence of 1D FFTE on global communication, which relies heavily on the optimization of circulant network topology.

5.2.3 Graph 500

The Graph 500 (version 2.1.4) [58, 59] tests graph search and shortest path algorithms on an tremendously large undirected graph distributed among all processors. It evaluates large-scale data-intensive performance for supercomputers. The processing speed is reported as mean TEPS, i.e. traversed edges per second. Due to implementation issues with SimGrid, we use the previous version 2.1.4 and the replicated version of parallel breadth-first search. We choose the scale of 29 with edge factor 4, generating an initial unweighted graph of 36 GB.

The performance ratios to torus for Graph 500 are shown in Fig. 8. The low-degree and high-degree optimal circulant topologies and Cartesian products can achieve average performance gains of 25%, 70% and 27% respectively. The global communication involved in Graph 500 makes the optimal circulant topologies suitable for enhancing the performance. We also note that hypercube brings top performance for Graph 500, which shows that the extremely symmetric topology structure coupled with proper routing may compensate its other disadvantages by matching with traffic patterns and internal MPI algorithms.

5.2.4 NAS parallel benchmarks (NPB)

The NAS Parallel Benchmarks (version 3.4.1) [60, 61] consist of programs derived from computational fluid dynamics (CFD) applications. We benchmark four kernel programs: discrete 3D FFT (FT), integer sort (IS), conjugate gradient method (CG) to calculate the smallest eigenvalue and multi-grid solver (MG) on a sequence of meshes, and three pseudo applications: lower-upper (LU) Gauss-Seidel solver, block tri-diagonal (BT) solver and scalar penta-diagonal (SP) solver. FT tests long-distance all-to-all communication; IS uses random memory access and tests both integer computation speed and communication; CG uses irregular memory access and tests unstructured long-distance communication; MG tests highly structured short- and long-distance communication with intensive memory access [60, 61]. We select the standard problem size Class C for each benchmark. Since BT and SP require a square number of processors, we perform these two applications only on topologies of $N = 64$, 256 and 1024.

Figures 9 and 10 show the NPB performance ratios to torus. The performance of FT (Fig. 9a) is similar to 1D FFTE (Fig. 7) in accordance with global communication, in which the low-degree and high-degree optimal circulant topologies and Cartesian products all have high performance gains of 66%, 124% and 27% in average respectively, with high-degree optimal circulant topologies 21% higher than hypercube. For IS (Fig. 9b), the performance is also similar to Graph 500 (Fig. 8) but with relatively lower performance ratios, in which the high-degree optimal circulant topologies and Cartesian products achieve average performance gains of 30% and 18%. The performances of other NPB benchmarks fluctuate around the average (Figs. 9c, d and 10). The high-degree optimal circulant topologies are able to reach higher average performance gains of 28%, 54% and 36% in CG, BT and SP, almost as good as hypercube, while maintaining around 3% and 10% in MG and LU. The Cartesian products achieve higher average performance gains of 35% and 18% in CG and MG, while keeping around equal performance in LU, BT and SP. Meanwhile, the low-degree optimal circulant topologies have almost equal average performance compared with torus in most of the NPB benchmarks apart from FT.

The NPB performance ratios to torus demonstrated multiple perspectives for performance enhancement. On one hand, applications with intensive global communication such as FT are strongly dependent on network topology, where the optimal circulant topologies contribute most to enhancing the performance. On the other hand, other factors such as data traffic patterns, internal algorithms and routing also influence the performance. Limited parallelism in LU [61] may result in its relatively lower performance ratios in general among all topologies. The fluctuation of performance ratios especially for the linear solvers such as MG, BT and SP can be caused by different partitions of the fixed input 3D grid onto topologies of varying sizes, which creates diverse data traffic patterns. In the meantime, some application can reach particularly high performance when its internal traffic pattern matches with the network connections, such as hypercube for IS. The Cartesian products of optimal circulant topologies and fully connected topologies may serve as better candidates for balancing the influence of global communication with specific traffic patterns, such as in CG and MG.

6 Discussion and conclusion

Optimal circulant graphs, obtained by highly efficient parallel exhaustive search algorithm in our study, are recognized as low-latency network topologies for large clusters of computing systems. The optimal circulant graphs and their Cartesian products, obtained from a huge search space of graphs with the same graph size and degree, have minimal diameter and MPL and maximum bisection width. These favorable graph properties prompt us to verify a common belief that they can significantly improve the performance for communication-intensive applications.

Indeed, the second contribution of our work is to demonstrate the enhancement of performance of these optimal circulant topologies, compared with other mainstream topologies including torus and hypercube of the same size, on effective bandwidth, 1D FFTE, NPB FT and Graph 500. These applications not only have high communication-to-computation ratio, but also have dominant dependence on long-distance global communication such as cross-network and all-to-all data traffic. The characterization of such applications is in accordance with our optimization methodology in which diameter, MPL and bisection width are overall global topology properties. As a result, on these applications our newly discovered optimal circulant topologies fulfill their advantages to the most extent. We also observe that, although the optimal circulant topologies and their Cartesian products achieve higher performance than torus, the performance ratios fluctuate around the average on different topology sizes for certain benchmarks such as NPB MG, BT and SP. This exhibits the influence on performance from other factors including specific data traffic patterns, internal algorithms, topology and problem scale, routing methods and memory access. A noticeable phenomenon is the benefit of extremely symmetric hypercube for the performance of Graph 500 and NPB IS. More investigation shows that specific algorithms for MPI collective operations such as recursive doubling and halving [62] match exactly the hypercube network connections. The relation between communication pattern and network topology can be explored to develop topology-aware mapping algorithms and frameworks [63, 64], which may even lead to hardware-software co-design [65, 66]. Additionally, different routing methods including adaptive routing on circulant topologies [67] can be examined and both topology- and routing-aware task mapping [68] may further enhance the performance of optimal circulant topologies.

To extend to the design of larger-scale networks, we propose and evaluate Cartesian products of optimal circulant topologies and fully connected topologies, which have both optimal local and global structures and thus can balance the global communication with specific traffic patterns as shown by our simulations. Various graph products [69, 70] of multiple optimal circulant topologies may also be applied to expand the network size. As another advantage over torus and hypercube, the node degree of circulant topologies can be arbitrary. Therefore, optimal circulant topologies are ideal candidates for global connections in large hierarchical networks [71, 72].

References

1. Top 500 supercomputer site (2021) http://www.top500.org
2. Moore GE (1965) Cramming more components onto integrated circuits. Electronics 38(8):114–117
3. Deng Y, Ramos AF, Hornos JEM (2012) Symmetry insights for design of supercomputer network topologies: roots and weights lattices. Int J Mod Phys B 26(31):1250169. https://doi.org/10.1142/s021797921250169x
4. Dally W, Towles B (2003) Principles and practices of interconnection networks. Elsevier Science & Technology, Amsterdam
5. Brightwell R, Pedretti K, Underwood K et al (2006) SeaStar interconnect: balanced bandwidth for scalable performance. IEEE Micro 26(3):41–57. https://doi.org/10.1109/mm.2006.65
6. IBM Blue Gene Team (2008) Overview of the IBM Blue Gene/P project. IBM J Res Dev 52(1.2):199–220. https://doi.org/10.1147/rd.521.0199
7. Alverson R, Roweth D, Kaplan L (2010) The gemini system interconnect. In: (2010) 18th IEEE Symposium on High Performance Interconnects. IEEE, Mountain View, CA, USA,. https://doi.org/10.1109/hoti.2010.23
8. Chen D, Parker JJ, Eisley NA et al (2011) The IBM Blue Gene/Q interconnection network and message unit. In: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis on - SC'11. ACM Press, Seattle, WA, USA, https://doi.org/10.1145/2063384.2063419
9. Ajima Y, Sumimoto S, Shimizu T (2009) Tofu: a 6D mesh/torus interconnect for exascale computers. Computer 42(11):36–40. https://doi.org/10.1109/mc.2009.370
10. Ajima Y, Kawashima T, Okamoto T et al (2018) The tofu interconnect d. In: 2018 IEEE International Conference on Cluster Computing (CLUSTER). IEEE, Belfast, UK, https://doi.org/10.1109/cluster.2018.00090
11. Hayes J, Mudge T (1989) Hypercube supercomputers. Proc IEEE 77(12):1829–1841. https://doi.org/10.1109/5.48826
12. Leiserson CE (1985) Fat-trees: universal networks for hardware-efficient supercomputing. IEEE Trans Comput C 34(10):892–901. https://doi.org/10.1109/tc.1985.6312192
13. Fu H, Liao J, Yang J et al (2016) The sunway TaihuLight supercomputer: system and applications. Sci China Inf Sci. https://doi.org/10.1007/s11432-016-5588-7
14. Wu CL, Feng TY (1980) On a class of multistage interconnection networks. IEEE Trans Comput C 29(8):694–702. https://doi.org/10.1109/tc.1980.1675651
15. Dally W (1990) Performance analysis of k-ary n-cube interconnection networks. IEEE Trans Comput 39(6):775–785. https://doi.org/10.1109/12.53599
16. Kim J, Dally WJ, Scott S, (2008) Technology-driven, highly-scalable dragonfly topology. In: (2008) International Symposium on Computer Architecture. IEEE, Beijing, China. https://doi.org/10.1109/isca.2008.19
17. Faanes G, Bataineh A, Roweth D, (2012) Cray cascade: a scalable HPC system based on a dragonfly network. In: (2012) International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, Salt Lake City, UT, USA. https://doi.org/10.1109/sc.2012.39
18. Sensi DD, Girolamo SD, McMahon KH et al (2020) An in-depth analysis of the slingshot interconnect. In: SC20: International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, Atlanta, GA, USA, https://doi.org/10.1109/sc41405.2020.00039
19. Cerf VG, Cowan DD, Mullin RC et al (1974) A lower bound on the average shortest path length in regular graphs. Networks 4(4):335–342. https://doi.org/10.1002/net.3230040405
20. Zhang P, Powell R, Deng Y (2011) Interlacing bypass rings to torus networks for more efficient networks. IEEE Trans Parallel Distrib Syst 22(2):287–295. https://doi.org/10.1109/tpds.2010.89

21. Zhang P, Deng Y (2012) An analysis of the topological properties of the interlaced bypass torus (iBT) networks. Appl Math Lett 25(12):2147–2155. https://doi.org/10.1016/j.aml.2012.05.013

22. Zhang P, Deng Y (2012) Design and analysis of pipelined broadcast algorithms for the all-port interlaced bypass torus networks. IEEE Trans Parallel Distrib Syst 23(12):2245–2253. https://doi.org/10.1109/tpds.2012.93

23. Zhang P, Deng Y, Feng R et al (2015) Evaluation of various networks configured by adding bypass or torus links. IEEE Trans Parallel Distrib Syst 26(4):984–996. https://doi.org/10.1109/tpds.2014.2315201

24. Feng R, Zhang P, Deng Y (2013) Deadlock-free routing algorithms for 6d mesh/iBT interconnection networks. In: 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. IEEE, Honolulu, HI, USA, https://doi.org/10.1109/snpd.2013.43

25. Feng R, Zhang P, Deng Y (2012) Simulated performance evaluation of a 6d mesh/iBT interconnect. In: 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. IEEE, Kyoto, Japan, https://doi.org/10.1109/snpd.2012.19

26. Xu Z, Huang X, Jimenez F et al (2019) A new record of graph enumeration enabled by parallel processing. Mathematics 7(12):1214. https://doi.org/10.3390/math7121214

27. Deng Y, Guo M, Ramos AF et al (2020) Optimal low-latency network topologies for cluster performance enhancement. J Supercomput 76(12):9558–9584. https://doi.org/10.1007/s11227-020-03216-y

28. Zhang Y, Huang X, Xu Z et al (2019) A structured table of graphs with symmetries and other special properties. Symmetry 12(1):2. https://doi.org/10.3390/sym12010002

29. Truong NT, Fujiwara I, Koibuchi M et al (2017) Distributed shortcut networks: low-latency low-degree non-random topologies targeting the diameter and cable length trade-off. IEEE Trans Parallel Distrib Syst 28(4):989–1001. https://doi.org/10.1109/tpds.2016.2613043

30. Yasudo R, Koibuchi M, Nakano K et al (2019) Designing high-performance interconnection networks with host-switch graphs. IEEE Trans Parallel Distrib Syst 30(2):315–330. https://doi.org/10.1109/tpds.2018.2864286

31. Sabino AU, Vasconcelos MFS, Deng Y et al (2018) Symmetry-guided design of topologies for supercomputer networks. Int J Mod Phys C 29(07):1850048. https://doi.org/10.1142/s0129183118500481

32. Nakao M, Sakai M, Hanada Y et al (2021) Graph optimization algorithm for low-latency interconnection networks. Parallel Comput 106(102):805. https://doi.org/10.1016/j.parco.2021.102805

33. Bermond J, Comellas F, Hsu D (1995) Distributed loop computer-networks: a survey. J Parallel Distrib Comput 24(1):2–10. https://doi.org/10.1006/jpdc.1995.1002

34. Hwang F (2003) A survey on multi-loop networks. Theor Comput Sci 299(1–3):107–121. https://doi.org/10.1016/s0304-3975(01)00341-3

35. Monakhova EA (2012) A survey on undirected circulant graphs. Discret Math Algorithms Appl 04(01):1250002. https://doi.org/10.1142/s1793830912500024

36. Boesch F, Wang JF (1985) Reliable circulant networks with minimum transmission delay. IEEE Trans Circuits Syst 32(12):1286–1291. https://doi.org/10.1109/tcs.1985.1085667

37. Beivide R, Herrada E, Balcazar J et al (1991) Optimal distance networks of low degree for parallel computers. IEEE Trans Comput 40(10):1109–1124. https://doi.org/10.1109/12.93744

38. Lau F, Chen G (1996) Optimal layouts of midimew networks. IEEE Trans Parallel Distrib Syst 7(9):954–961. https://doi.org/10.1109/71.536939

39. Mukhopadhyaya K, Sinha B (1995) Fault-tolerant routing in distributed loop networks. IEEE Trans Comput 44(12):1452–1456. https://doi.org/10.1109/12.477250

40. Gómez D, Gutierrez J, Ibeas Á (2007) Optimal routing in double loop networks. Theor Comput Sci 381(1–3):68–85. https://doi.org/10.1016/j.tcs.2007.04.002

41. Obradoviç N, Peters J, Ružić G (2005) Reliable broadcasting in double loop networks. Networks 46(2):88–97. https://doi.org/10.1002/net.20076

42. Monakhov O, Monakhova E, (2019) A comparative analysis of bioinspired algorithms for solving the problem of optimization of circulant and hypercirculant networks. In: (2019) 15th International Asian School-Seminar Optimization Problems of Complex Systems (OPCS), IEEE. https://doi.org/10.1109/opcs.2019.8880247

43. Feria-Purón R, Ryan J, Pérez-Rosés H (2014) Searching for large multi-loop networks. Electron Notes Discret Math 46:233–240. https://doi.org/10.1016/j.endm.2014.08.031

44. Bevan D, Erskine G, Lewis R (2017) Large circulant graphs of fixed diameter and arbitrary degree. Ars Math Contemp 13(2):275–291. https://doi.org/10.26493/1855-3974.969.659,

45. Lewis RR (2018) The degree-diameter problem for circulant graphs of degrees 10 and 11. Discret Math 341(9):2553–2566. https://doi.org/10.1016/j.disc.2018.05.024
46. Gross JL, Yellen J, Zhang P (2013) Handbook of graph theory. CRC Press, Boca Raton
47. Boesch F, Tindell R (1984) Circulants and their connectivities. J Graph Theory 8(4):487–499. https://doi.org/10.1002/jgt.3190080406
48. Arndt J (2010) Matters computational: ideas, algorithms, source code. Springer Science & Business Media, Berlin
49. Fxt: a library of algorithms (2021) https://www.jjj.de/fxt/
50. Ruskey F (2003) Combinatorial generation. Preliminary working draft University of Victoria, Victoria, BC, Canada 11:20
51. Sanders P, Schulz C (2013) Think locally, act globally: highly balanced graph partitioning. In: Experimental Algorithms. Lecture notes in computer science, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 164–175, https://doi.org/10.1007/978-3-642-38527-8_16
52. McKay BD, Piperno A (2014) Practical graph isomorphism, II. J Symb Comput 60:94–112. https://doi.org/10.1016/j.jsc.2013.09.003
53. Casanova H, Giersch A, Legrand A et al (2014) Versatile, scalable, and accurate simulation of distributed applications and platforms. J Parallel Distrib Comput 74(10):2899–2917. https://doi.org/10.1016/j.jpdc.2014.06.008
54. Effective Bandwidth (b_eff) Benchmark (2021) https://fs.hlrs.de/projects/par/mpi/b_eff/
55. Koniges A, Rabenseifner R, Solchenbach K (2001) Benchmark design for characterization of balanced high-performance architectures. In: Proceedings 15th International Parallel and Distributed Processing Symposium. IPDPS 2001. IEEE Comput. Soc, San Francisco, CA, USA. https://doi.org/10.1109/ipdps.2001.925208
56. FFTE : A fast fourier transform package (2021) http://www.ffte.jp/
57. Takahashi D, Kanada Y (2000) High-performance radix-2, 3 and 5 parallel 1-D complex FFT algorithms for distributed-memory parallel computers. J Supercomput 15(2):207–228. https://doi.org/10.1023/a:1008160021085
58. Graph 500 (2021) http://graph500.org/
59. Murphy RC, Wheeler KB, Barrett BW et al (2010) Introducing the graph 500. Cray Users Group (CUG) 19:45–74
60. NPB: NAS parallel benchmarks (2021) http://www.nas.nasa.gov/publications/npb.html
61. Bailey D, Barszcz E, Barton J et al (1991) The NAS parallel benchmarks. Int J Supercomput Appl 5(3):63–73. https://doi.org/10.1177/109434209100500306
62. Thakur R, Rabenseifner R, Gropp W (2005) Optimization of collective communication operations in MPICH. Int J High Perform Comput Appl 19(1):49–66. https://doi.org/10.1177/1094342005051521
63. Deveci M, Kaya K, Ucar B (2015) Fast and high quality topology-aware task mapping. In: (2015) IEEE International Parallel and Distributed Processing Symposium. IEEE. https://doi.org/10.1109/ipdps.2015.93
64. Ma T, Bosilca G, Bouteiller A (2012) HierKNEM: an adaptive framework for kernel-assisted and topology-aware collective communications on many-core clusters. In: (2012) IEEE 26th International Parallel and Distributed Processing Symposium, IEEE. https://doi.org/10.1109/ipdps.2012.91
65. Wolf W (2003) A decade of hardware/ software codesign. Computer 36(4):38–43. https://doi.org/10.1109/mc.2003.1193227
66. Teich J (2012) Hardware/software codesign: the past, the present, and predicting the future. Proc IEEE 100(Special Centennial Issue):1411–1430. https://doi.org/10.1109/jproc.2011.2182009
67. Monakhov OG, Monakhova EA, Romanov AY et al (2021) Adaptive dynamic shortest path search algorithm in networks-on-chip based on circulant topologies. IEEE Access 9:160836–160846. https://doi.org/10.1109/access.2021.3131635
68. Mirsadeghi SH, Afsahi A (2016) PTRAM: a parallel topology-and routing-aware mapping framework for large-scale HPC systems. In: (2016) IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), IEEE. https://doi.org/10.1109/ipdpsw.2016.146
69. Parsonage E, Nguyen HX, Bowden R et al (2011) Generalized graph products for network design and analysis. In: 2011 19th IEEE International Conference on Network Protocols. IEEE, https://doi.org/10.1109/icnp.2011.6089084,
70. Hammack RH, Imrich W, Klavžar S et al (2011) Handbook of product graphs, vol 2. CRC Press, Boca Raton
71. Dandamudi S, Eager D (1990) Hierarchical interconnection networks for multicomputer systems. IEEE Trans Comput 39(6):786–797. https://doi.org/10.1109/12.53600

72. Abd-El-Barr M, Al-Somani TF (2011) Topological properties of hierarchical interconnection networks: a review and comparison. J Electr Comput Eng 2011:1–12. https://doi.org/10.1155/2011/189434