A First Approach to King Topologies for On-Chip Networks

Esteban Stafford, Jose L. Bosque, Carmen Martínez, Fernando Vallejo,
Ramon Beivide, and Cristobal Camarero

Electronics and Computers Department, University of Cantabria, Faculty of Sciences,
Avda. Los Castros s/n, 39006 Santander, Spain
esteban.stafford@gestion.unican.es,
{joseluis.bosque,carmen.martinez,fernando.vallejo,
ramon.beivide}@unican.es, cristobal.camarero@alumnos.unican.es

Abstract. In this paper we propose two new topologies for on-chip networks that we have denoted as king mesh and king torus. These are a higher degree evolution of the classical mesh and torus topologies. In a king network packets can traverse the networks using orthogonal and diagonal movements like the king on a chess board. First we present a topological study addressing distance properties, bisection bandwidth and path diversity as well as a folding scheme. Second we analyze different routing mechanisms. Ranging from minimal distance routings to missrouting techniques which exploit the topological richness of these networks. Finally we make an exhaustive performance evaluation comparing the new king topologies with their classical counterparts. The experimental results show a performance improvement, that allow us to present these new topologies as better alternative to classical topologies.

1 Introduction

Although a lot of research on interconnection networks has been conducted in the last decades, constant technological changes demand new insights about this key component in modern computers. Nowadays, networks are critical for managing both off-chip and on-chip communications.

Some recent and interesting papers advocate for networks with high-radix routers for large-scale supercomputers[1][2]. The advent of economical optical signalling enables this kind of topologies that use long global wires. Although the design scenario is very different, on-chip networks with higher degree than traditional 2D meshes or tori have also been recently explored[3]. Such networks entail the use of long wires in which repeaters and channel pipelining are needed. Nevertheless, with current VLSI technology, the planar substrate in which the network is going to be deployed suggests the use of 2D mesh-like topologies. This has been the case of Tilera[4] and the Intel's Teraflop research chip[5], with 64 and 80 cores arranged in a 2D mesh respectively. Forthcoming technologies such as on-chip high-speed signalling and optical communications could favor the use of higher degree on-chip networks.

P. D'Ambra, M. Guarracino, and D. Talia (Eds.): Euro-Par 2010, Part II, LNCS 6272, pp. 428–439, 2010.

In this paper, we explore an intermediate solution. We analyze networks whose degrees double the radix of a traditional 2D mesh while still preserving an attractive layout for planar VLSI design. We study meshes and tori of degree eight in which a packet located in any node can travel in one hop to any of its eight neighbours just like the king on a chessboard. For this reason, we denote these networks *king meshes* and *king tori*. In this way, we adopt a more conservative evolution towards higher radix networks trying to exploit their advantages while avoiding the use of long wires. The simplicity and topological properties of these networks offer tantalising features for future on-chip architectures: higher throughput, smaller latency, trivial partitioning in smaller networks, good scalability and high fault-tolerance.

The use of diagonal topologies has been considered in the past, in the fields of VLSI[6], FPGA[7] and interconnection networks[8]. Also mesh and toroidal topologies with added diagonals have been considered, both with degree six[9] and eight[10].The king lattice has been previously studied in several papers of Information Theory[11].

The goal of this paper is to explore the suitability of king topologies to constitute the communication substrate of forthcoming on-chip parallel systems. With this idea in mind, we present the foundations of king networks and a first attempt to unleash their potential. The main contributions of our research are the following:

i) An in-depth analysis of the topological characteristics of king tori and king meshes.
ii) The introduction and evaluation of king tori, not considered previously in the technical literature.
iii) A folding scheme that ensures king tori scalability.
iv) An adaptive and deadlock-free routing algorithm for king topologies.
v) A first performance evaluation of king networks based on synthetic traffic.

The remainder of this paper is organized as follows. Section 2 is devoted to define the network topologies considered in this paper. The most relevant distance parameters and the bisection bandwidth are computed for each network and a folding method is considered for networks with wrap-around links. Section 3 tackles the task of finding routing algorithms to unlock the networks' potential high performance, starting with simple minimum-distance algorithms and evolving to more elaborate missrouting and load balancing techniques. Section 4 presents a first performance evaluation of these networks. Finally, Section 5 concludes the paper highlighting its most important findings.

2 Description of the Topologies

In this Section we define and analyze distance properties of the network topologies considered in this paper: square meshes, square king meshes, square tori and square king tori. Then, we obtain expressions for significant distance parameters as well as the bisection bandwidth. Finally, we consider lay-out possibilities minimizing wire length for those topologies with wrap-around edges.

As usual, networks are modeled by graphs, where graph vertices represent processors and edges represent the communication links among them. In this paper we will only consider square networks, as sometimes networks with sides of different length result in an unbalanced use of the links in each dimension[12]. Therefore, in the following we will obviate the adjective "square". Hence, for any of the networks considered here the number of nodes will be $n = s^2$, for any integer $s > 1$.

By M_s we will denote the usual mesh of side s. This is a very well-known topology which has been deeply studied. A mesh based network of degree eight can be obtained by adding new links such that, any packet not only can travel in orthogonal directions, but also can use diagonal movements. Will denote by KM_s the *king mesh network*, which is obtained by adding diagonal links (just for non-peripheral nodes) to M_s.

Note that both networks are neither regular nor vertex-symmetric. The way to make this kind of network regular and vertex-symmetric is to add wrap-around links in order to make that every node has the same number of neighbors. We will denote as T_s the usual torus network of side s. The torus is obviously the four degree regular counterpart of the mesh. Then, KT_s will denote the *king torus network*, that is, a king mesh with new wrap-around links in order to obtain an eight degree regular network. Another way to see this network is as a torus with extra diagonal links that turn the four degree torus into an eight degree network. In Figure 1 an example of each network is shown.

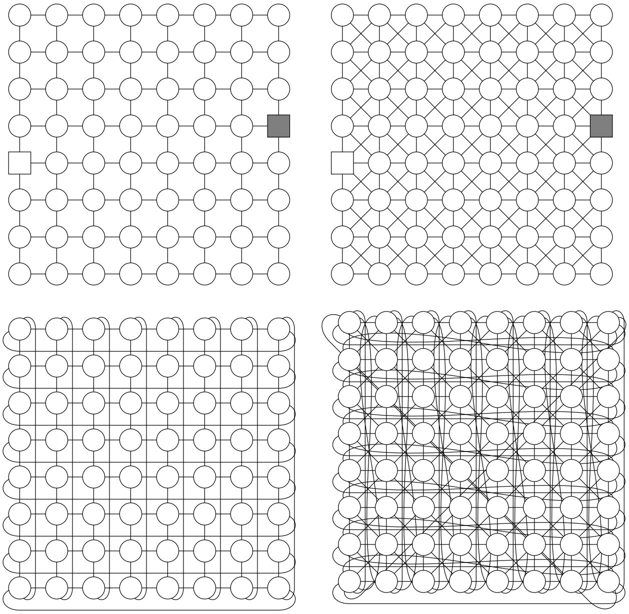

Fig. 1. Examples of Mesh, King Mesh, Torus and King Torus Networks

In an ideal system, transmission delays in the network can be inferred from its topological properties. The maximum packet delay is given by the diameter of the graph. It is the maximum length over all minimum paths between any pair of nodes. The average delay is proportional to the average distance, which is computed as the average length of all minimum paths connecting every pair of nodes of the network. In Table 1 we record these parameters of the four networks considered. The diameter and average distance of mesh and torus are well-known values, [13]. The distance properties of king torus were presented in [14].

Table 1. Topological Parameters

Network	M_s	KM_S	T_s	KT_s
Diameter	$2s$	s	s	$\lfloor \frac{s}{2} \rfloor$
Average Distance	$\approx \frac{2}{3}s$	$\approx \frac{7}{15}s$	$\approx \frac{s}{2}$	$\approx \frac{s}{3}$
Bisection Bandwidth	$2s$	$6s$	$4s$	$12s$

An specially important metric of interconnection networks is the throughput, the maximum data rate the network can deliver. In the case of uniform traffic, that is, nodes send packets to random nodes with uniform probability, the throughput is bounded by the bisection. According to the study in [13], in networks with homogeneous channel bandwidth, as the ones considered here, the bisection bandwidth is proportional to the channel count across the smallest cut that divides the network into two equal halves. This value represents an upper bound in the throughput under uniform traffic.

In Table 1, values for the bisection for mesh and torus are shown, see [13]. The obtention of the bisection bandwidth in king mesh and torus is straightforward. Note that a king network doubles the number of links of its orthogonal counterpart but has three times the bisection bandwidth.

In a more technological level, physical implementation of computer networks usually requires that the length of the links is similar, if not constant. In the context of networks-on-chip, mesh implementation is fairly straightforward. A regular mesh can be lade out with a single metal layer. Due to the crossing diagonal links, the king mesh requires two metal layers.

However tori have wrap-around links whose length depend on the size of the network. To overcome this problem, a well known technique is graph folding. A standard torus can be implemented with two metal layers. Our approach to folding king tori is based on the former but because of the diagonal links four metal layers are required. As a consequence of the folding, the length of the links is between two and $\sqrt{8}$ in king tori. This seems to be the optimal solution for this kind of networks. Figure 2 shows a 8×8 folded king torus. For the sake of clarity, the folded graph is shown with the orthogonal and diagonal links separated.

Now, if we compare king meshes with tori, we observe that the cost of doubling the number of links gives great returns. Bisection bandwidth is 50% larger,

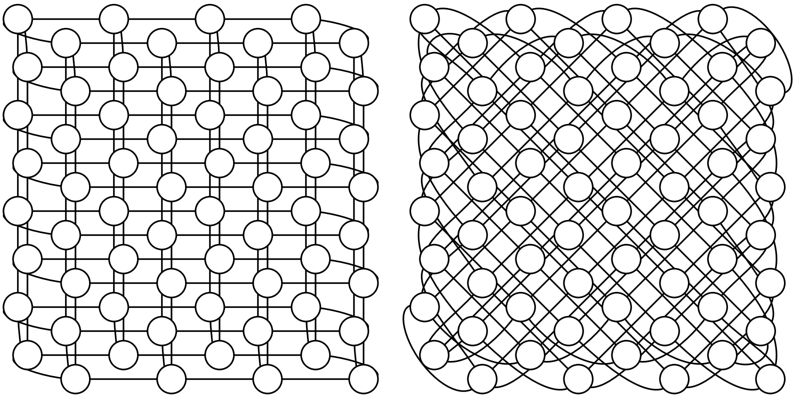

Fig. 2. Folding of King Torus Network. For the sake of clarity, the orthogonal and diagonal links are shown in separates graphs.

average distance is almost 5% less and diameter remains the same. In addition, implementation of a king mesh on a network-on-chip is simpler, as it does not need to be folded and fits in two metal layers just like a folded torus.

3 Routing

This section explores different routing techniques trying to take full advantage of the king networks. For simplicity it focuses on toroidal networks assuming that meshes will have a similar behaviour. Our development starts with the most simple minimum distance routing continuing through to more elaborate load balancing schemes capable of giving high performance in both benign and adverse traffic situations.

Enabling packets to reach their destination in direct networks is traditionally done with source routing. This means that at the source node, when the packet is injected, a routing record is calculated based on source and destination using a *routing function.* This routing record is a vector whose integer components are the number of jumps the packet must make in each dimension in order to reach its destination.

In 2D networks routing records have two components, Δ_X and Δ_Y. These components could be used to route packets in king networks, but the diagonal links, that can be thought as shortcuts, would never be used. Then it is necessary to increase the number of components in the routing record to account for the greater degree of these new networks. Thus we will broaden the definition of routing record as a vector whose components are the number of jumps a packet must make in each *direction*, not dimension. Thus, king networks will have four directions, namely X and Y as the horizontal and vertical, Z for the diagonal $y = x$ and T for the diagonal $y = -x$.

3.1 Minimal Routing

To efficiently route packets in a king network, we need a routing function that takes source and destination nodes and gives a routing record that makes the packet reach its destination in the minimum number of jumps. Starting with the 2D routing record, it is easy to derive a naive king routing record that is minimal(*Knaive*). From the four components of the routing record, this routing function will not use two of them. Hence, routing records will have, at most, two non-zero components, one is orthogonal and the other is diagonal. The algorithm is simple, consider (Δ_X, Δ_Y) where $\Delta_X > \Delta_Y > 0$. The corresponding king routing record would be $(\delta_X, \delta_Y, \delta_Z, \delta_T) = (\Delta_X - \Delta_Y, 0, \Delta_Y, 0)$. The rest of the cases are calculated in a similar fashion.

In addition to being minimal, this algorithm balances the use of all directions under uniform traffic, a key aspect in order to achieve maximum throughput. The drawback, however, is that it does not exploit all the path diversity available in the network. Path diversity is defined as the number of minimal paths between a pair of nodes a, b of a network. For mesh and tori will denote it as $|R_{ab}|$.

$$|R_{ab}| = \binom{|\Delta_x| + |\Delta_y|}{|\Delta_x|}.$$

Similarly, in king mesh and tori the path diversity is:

$$|RK_{ab}| = \binom{|\Delta_x|}{|\Delta_y|}_2 \text{ where } \binom{n}{k}_2 = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \binom{2n-2j}{n-k-j}$$

Thus, the path diversity for king networks is overwhelmingly higher than in meshes and tori. Take for example $\Delta_x = 7, \Delta_y = 1$, this is the routing record to go from the white box to the gray box in Figure 1. In a mesh the path diversity would be $R_{ab} = 8$ while in a king mesh $RK_{ab} = 357$.

Now, the corresponding Knaive routing record is $(\delta_X, \delta_Y, \delta_Z, \delta_T) = (6, 0, 1, 0)$. This yields only 7 alternative paths, so 350 path are ignored, this is even less than the 2d torus. This is not a problem under uniform and other benign traffic patterns but on adverse situations a diminished performance is observed. For instance, see the performance of 16 × 16 torus with 1-phit packets in Figure 3. The throughput in uniform traffic of the Knaive algorithm is 2.4 times higher than that of a standard torus, which is a good gain for the cost of doubling network resources. However, in shuffle traffic, the throughput is only double and under other traffic patterns even less.

A way of improving this is increasing the path diversity by using routing records with three non-zero components. This can be done by applying the notion that two jumps in one orthogonal direction can be replaced by a jump in Z plus one in T without altering the path's length. Based on our experiments we have found that the best performance is obtained when using transformations similar to the following.

$$(\delta_X, 0, \delta_Z, 0) \rightarrow (\left\lfloor \frac{\delta_X}{3} \right\rfloor, 0, \delta_Z + \left\lfloor \frac{\delta_X}{3} \right\rfloor, \left\lfloor \frac{\delta_X}{3} \right\rfloor)$$

Being this an enhancement of the Knaive algorithm we denote it *EKnaive*. It is important to note that it is still minimum-distance and gives more path diversity but not all that is available. Continuing with our example, this algorithm will give us 210 of the total 357 paths (See Table 2).

As can be seen in Figure 3, the *EKnaive* routing record improves the throughput in some adverse traffic patterns due to its larger path diversity. However this comes at a cost. The inherent balance in the link utilization of the Knaive algorithm is lost, thus giving worse performance under uniform traffic.

Table 2. Alternative routing records for (6,0,1,0) with corresponding path diversity

Routing Record $(\delta_X, \delta_Y, \delta_Z, \delta_T)$	Path Diversity
(6,0,1,0)	7
(4,0,2,1)	105
(2,0,3,2)	210
(0,0,4,3)	35
theoretical	357

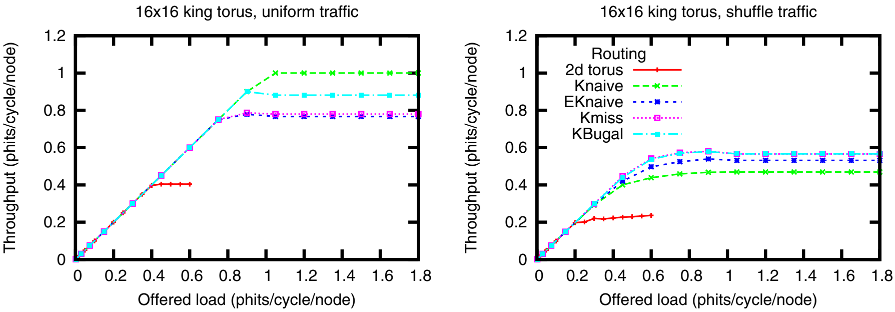

Fig. 3. Throughput comparison of the various routing algorithms in 16 × 16 toroidal networks

3.2 Misrouting

In the light of the previous experiences, we find that direction balancing is key. But is it important enough to relax the minimum distance requirement? In response to this question, we have developed a new routing function whose routing record may have four non-zero components. Forcing packets to use all directions will cause missrouting as the minimum paths will no longer be used. Thus we name this approach *Kmiss*.

Ideally, to achieve direction balance, the four components would be as close as possible. However this would cause path lengths to be unreasonable long. A compromise must be reached between the path length and component similarity. With Kmiss, the routing record is extracted from a table indexed by the 2D routing record. The table is constructed so that the components of the routing records do not differ more than 3.

The new function improves the load balance regardless of the traffic pattern and provides packets with more means to avoid local congestion. In addition it increases the path diversity.

Experimental results as those shown in Section 4 show that this algorithm gives improved throughput in adverse traffic patterns but the misrouting diminishes its performance in benign situations. Figure 3 shows that Kmiss is still poor in uniform traffic, but gives the highest throughput under shuffle.

3.3 Routing Composition

In essence, we have a collection of routing algorithms. Some are very good in benign traffic but perform badly under adverse traffic, while others are reasonably good in the latter but disappointing in the former. Ideally, we would like to choose which algorithm to use depending on the situation. Better yet would be that the network switches from one to another by its self. This is achieved to a certain extent in Universal Globally Adaptive Load-balancing (UGAL)[15]. In a nutshell what this algorithm does is routing algorithm composition. Based on local traffic information, each node decides whether a packet is sent using a minimal routing or the non-minimal Valiant's routing [16], composing a better algorithm that should have the benefits of both of the simple ones.

As we show next, *KBugal* is an adaptation of UGAL to king networks and bubble routing with two major improvements. On one hand, for the non-minimal routing, instead of Valiant's algorithm, we use Kmiss routing. This approach takes advantage of the topology's path diversity without significantly increasing latency and it has a simpler implementation. On the other hand, the philosophy behind UGAL resides in estimating the transmission time of a packet at the source node based on local information. Thus selecting the shortest output queue length among all profitable channels both for the minimal and the non-minimal routings. In the best scenario, the performance of KBugal is the best out of the two individual algorithms, as can be seen in Figure 3.

The use of bubble routing allows deadlock-free operation with only two virtual channels per physical channel in contrast to the three used by original UGAL. In order to get a better estimation, KBugal takes into account the occupation of both virtual channels together for each profitable physical channel. The reason behind this is fairly simple. Considering that all virtual channels share the same physical channel, the latency is determined by the occupation of all virtual channels, not only the one it is injected in.

4 Evaluation

In this section we present the experimental evaluation carried out to verify the better performance and scalability of the proposed networks. This is done by comparing with other networks usually considered for future network-on-chip

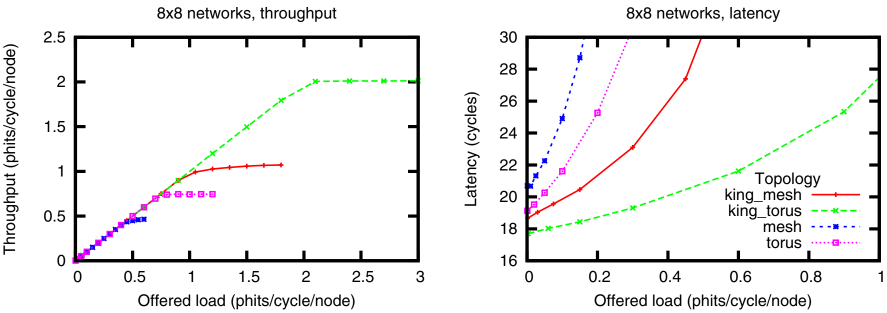

Fig. 4. Throughput and latency of king topologies with Knaive compared to mesh and tori under uniform traffic

architectures, as are the mesh and torus with size 8 × 8. The same study was made with 16 × 16 networks, but due to their similarity to 8 × 8 and lack of space, these results are not shown.

All the experiments have been done on a functional simulator called fsin[17]. The router model is based on the bubble adaptive router presented in [18] with two virtual channels. As we will be comparing networks of different degree, a constant buffer space will be assigned to each router and will be divided among all individual buffers. Another important factor in the evaluation of networks are the traffic patterns. The evaluation has been performed with synthetic workload using typical traffic patterns. According to the effect on load balance, traffic patterns can be classified into benign and adverse. The former naturally balances the use of network resources, like uniform or local, while the latter introduces contention and hotspots that reduce performance, as in complement or butterfly. Due to space limitations, only the results for three traffic patterns are shown as they can represent the behaviour observed on the rest. These are uniform, bit-complement and butterfly.

Figure 4 shows the throughput and latency of king networks using Knaive compared to those of 2d tori and meshes. It proves that the increased degree of the king networks outperforms their baseline counterparts by more than a factor two. The average latency on zero load is reduced according to the average distance theoretical values. Packets are 16-phit long, thus making the latency improvement less obvious in the graphs. Observe that king meshes have significantly better performance than 2d tori, both in throughput and latency.

Figure 5 presents an analysis of the different routing techniques under the three traffic patterns and for 8 × 8 king tori and meshes. Comparing the results of networks with different sizes highlights that the throughput per node is halved. This is due to the well known fact that the number of nodes in square networks grows quadratically with the side while the bisection bandwidth grows linearly.

For benign traffic patterns, the best results are given by Knaive routing. However in adverse traffic, a sensible decrease in performance is observed, caused by the reduced path diversity. As mentioned in Section 3 this limitation is overcome

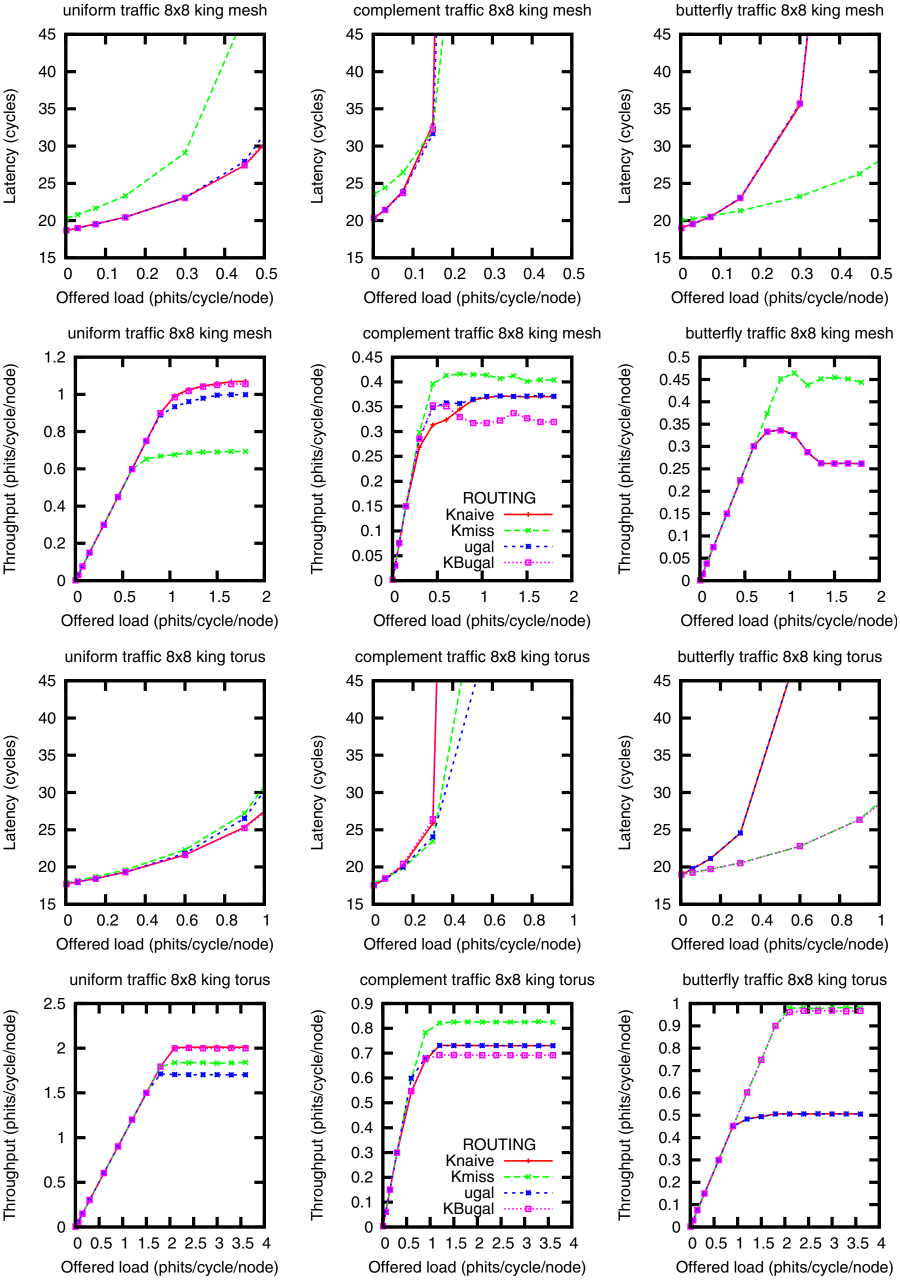

Fig. 5. Throughput and latency of routings on 8×8 king meshes and tori under different traffic patterns

by the Kmiss routing. In fact this routing yields poor performance under benign traffic pattern but very good under the adverse ones.

Our composite routing algorithm KBugal gives the best average performance on all traffic patterns. In the benign situations the throughput is slightly less than Knaive. And under adverse traffic, performance is similar to the Kmiss routing, being even better in some situations. The results show that KBugal gives better performance than its more generic predecessor UGAL. As can be seen, under benign traffic a improvement of 15% is obtained and between 10% (complement) and 90% (butterfly).

5 Conclusion

In this paper we have presented the foundations of king networks. Their topological properties offer tantalising possibilities, positioning them as clear candidates for future network-on-chip systems. Noteworthy are king meshes, which have the implementation simplicity and wire length of a mesh yet better performance than 2d tori. In addition, we have presented a series of routing techniques specific for king networks, that are both adaptive and deadlock free, which allow to exploit their topological richness. A first performance evaluation of these algorithms based on synthetic traffic has been presented in which their properties are highlighted. Further study will be required to take full advantage of these novel topologies that promise higher throughput, smaller latency, trivial partitioning and high fault-tolerance.

Acknowledgment

This work has been partially funded by the Spanish Ministry of Education and Science (grant TIN2007-68023-C02-01 and Consolider CSD2007-00050), as well as by the HiPEAC European Network of Excellence.

References

1. Kim, J., Dally, W., Scott, S., Abts, D.: Technology-driven, highly-scalable dragonfly topology. SIGARCH Comput. Archit. News 36(3), 77–88 (2008)
2. Scott, S., Abts, D., Kim, J., Dally, W.: The blackwidow high-radix clos network. SIGARCH Comput. Archit. News 34(2), 16–28 (2006)
3. Kim, J., Balfour, J., Dally, W.: Flattened butterfly topology for on-chip networks. In: MICRO 2007: Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture, pp. 172–182. IEEE Computer Society, Washington (2007)
4. Wentzlaff, D., Griffin, P., Hoffmann, H., Bao, L., Edwards, B., Ramey, C., Mattina, M., Miao, C.C., Bown III, J.F., Agarwal, A.: On-chip interconnection architecture of the tile processor. IEEE Micro 27, 15–31 (2007)
5. Vangal, S., Howard, J., Ruhl, G., Dighe, S., Wilson, H., Tschanz, J., Finan, D., Singh, A., Jacob, T., Jain, S., Erraguntla, V., Roberts, C., Hoskote, Y., Borkar, N., Borkar, S.: An 80-tile sub-100-w teraflops processor in 65-nm cmos. IEEE Journal of Solid-State Circuits 43(1), 29–41 (2008)

6. Igarashi, M., Mitsuhashi, T., Le, A., Kazi, S., Lin, Y., Fujimura, A., Teig, S.: A diagonal interconnect architecture and its application to risc core design. IEIC Technical Report (Institute of Electronics, Information and Communication Engineers) 102(72), 19–23 (2002)
7. Marshall, A., Stansfield, T., Kostarnov, I., Vuillemin, J., Hutchings, B.: A reconfigurable arithmetic array for multimedia applications. In: FPGA 1999: Proceedings of the 1999 ACM/SIGDA seventh international symposium on Field programmable gate arrays, pp. 135–143. ACM, New York (1999)
8. Tang, K., Padubidri, S.: Diagonal and toroidal mesh networks. IEEE Transactions on Computers 43(7), 815–826 (1994)
9. Shin, K., Dykema, G.: A distributed i/o architecture for harts. In: Proceedings of 17th Annual International Symposium on Computer Architecture, pp. 332–342 (1990)
10. Hu, W., Lee, S., Bagherzadeh, N.: Dmesh: a diagonally-linked mesh network-on-chip architecture. nocarc (2008)
11. Honkala, I., Laihonen, T.: Codes for identification in the king lattice. Graphs and Combinatorics 19(4), 505–516 (2003)
12. Camara, J., Moreto, M., Vallejo, E., Beivide, R., Miguel-Alonso, J., Martinez, C., Navaridas, J.: Twisted torus topologies for enhanced interconnection networks. IEEE Transactions on Parallel and Distributed Systems 99 (2010) (PrePrints)
13. Dally, W., Towles, B.: Principles and Practices of Interconnection Networks. Morgan Kaufmann Publishers Inc., San Francisco (2003)
14. Martinez, C., Stafford, E., Beivide, R., Camarero, C., Vallejo, F., Gabidulin, E.: Graph-based metrics over qam constellations. In: IEEE International Symposium on Information Theory, ISIT 2008, pp. 2494–2498 (2008)
15. Singh, A.: Load-Balanced Routing in Interconnection Networks. PhD thesis (2005)
16. Valiant, L.: A scheme for fast parallel communication. SIAM Journal on Computing 11(2), 350–361 (1982)
17. Ridruejo Perez, F., Miguel-Alonso, J.: Insee: An interconnection network simulation and evaluation environment (2005)
18. Puente, V., Izu, C., Beivide, R., Gregorio, J., Vallejo, F., Prellezo, J.: The adaptive bubble router. J. Parallel Distrib. Comput. 61(9), 1180–1208 (2001)