A Comparison Study on Data Vortex Packet Switched Networks with Redundant Buffers and with Inter-cylinder Paths

Qimin Yang
Harvey Mudd College, Engineering Department
Claremont, California, USA
E-mail: Qimin_yang@hmc.edu

***Abstract*-Optical switching fabric networks become essential systems in high capacity communication and computing systems. This paper focuses on Data Vortex network architecture with two alternative implementations for improved performance. Either a buffer is added within the routing node or inter cylinder paths are provided for enhanced routing performance. Since the extra hardware required for both implementations are the same, the network with better routing performance provides a better solution. The comparison study has demonstrated that networks with inter cylinder paths provide significantly lower latency and better throughput, therefore this approach provides more effective sharing of the routing resources within the network compared with the node buffering implementation. The difference in performance is also shown to be more significant under higher load conditions and for larger networks.**

Keywords-data vortex network; packet switched network; optical; network; buffering.

I. INTRODUCTION

Switching fabric networks are important subsystems in high capacity communication networks and computing systems. Typical space switch uses rich connectivity to handle dynamic traffic coming from a large number of I/O ports while maintaining a high data throughput and small latencies. In high end multi-processor computing applications, the number of I/O ports or processors can reach ~1000 and each could run at tens of Gbit/s data rate, and at the same time low latency (tens or hundreds of μs) must be maintained through such networks. Multistage self-routing network architectures often provide better system scalability with distributed routing nodes incorporating relatively simple routing logics which lead to cost-effective implementation and shorter processing delay. In order to provide higher data throughput, such networks can be implemented using optical fibre or optical switching technology.

Many recent researches have focused on developing optical switching fabric networks and network testbeds [1][2][3][4]. While it is relatively easy to achieve higher transmission bandwidth with Wavelength Division Multiplexing (WDM) within a single fibre, the routing logics as well as the handlings of traffic contention are hard to manage within the optical domain. In particular, Data Vortex packet switched network architecture is developed for the ease of photonics implementation, and such network is highly scalable to support a large number of I/O ports where each runs at high data rate and the network maintains a small routing latency [4][5]. The combination of its high spatial connectivity and an electronic traffic control mechanism among routing nodes lead to bufferless operation and a much simpler routing logic within the nodes. Even though it uses deflection based routing, the special connectivity avoids large deflection penalty and overall probability of deflection; therefore, it is advantageous compared with other commonly used interconnection architectures.

Previous researches have shown that with sufficient network redundancy, Data Vortex network scales to support a large number of I/O ports while achieving high throughput and low latency performance. On the other hand, at extremely high load conditions, and less redundant network conditions, the throughput tends to be limited by traffic backpressure in the deflection based routing. There have been several approaches suggested to enhance the routing performance of the Data Vortex networks, especially for these less ideal operating conditions [6][7][8][9]. In general, these performance enhancement methods require additional routing paths or routing resources, thus detailed cost and performance analysis must be carried out in comparison with the original networks. There is no comparison between different enhancement method, so in this paper, we emphasize such comparison of two methods using buffering and using extra inter-cylinder paths. These two methods are of particular interests because of they share the same cost with reasonable hardware increase in comparison to the original network and their easy implementation. The performance will be compared to each other as well as to the original Data Vortex networks.

The paper is organized as follows: in Section II, the original Data Vortex network architecture will be explained in details. In Section III, two previously proposed enhancement methods, the nodal buffering method as well as inter-cylinder path method are illustrated and compared in details. The routing performance comparison will be provided in Section IV for various network conditions, and the conclusion is given in Section V.

 ISBN: 978-1-61208-174-8

II. DATA VORTEX ARCHITECTURE

The Data Vortex architecture arranges its routing nodes in three dimensional multiple stage configuration as shown in Fig. 1. While the cylindrical levels (c=0 at the outermost cylinder to $c = \log_2 H$ at the innermost cylinder) provide the multiple levels in the routing stages, the angular dimension with repeated connection patterns provides multiple open paths to the destination therefore results in a much smaller latency penalty as deflection occurs. Inter-cylinder paths are not shown for a better view, and they are simply parallel links that maintain the height position of the packets when they propagate from outer to inner cylinders.

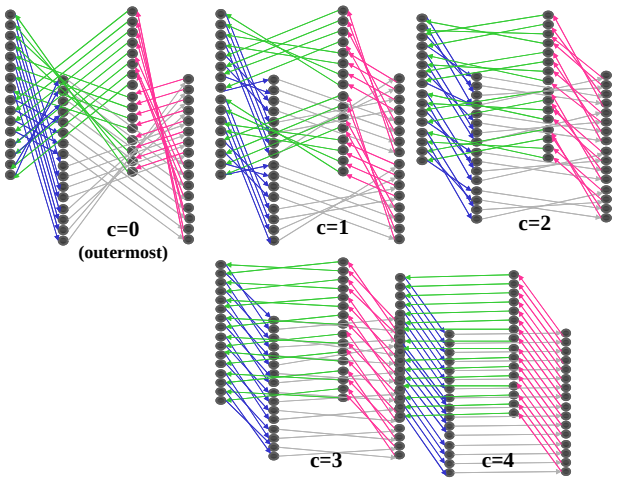

Figure 1. Data Vortex network with Angle=4, Height=16 and Cylinder=5 and its layout of routing node at different cylinders

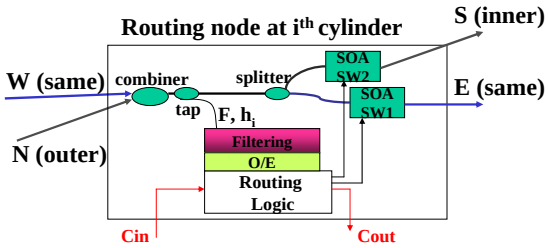

Figure 2. Routing node implementation

Packet's destination height is encoded in binary, and in the physical layer each of these binary bits is modulated onto a distinct wavelength, so that simple passive wavelength filtering can be used to extract and decode the single header bit h_i at the ith cylinder level. This is shown within the node structure in Fig. 2. Only a small amount of optical power will be tapped and used for header decoding purpose so that packets stay in optical domain as they travel through the network. Each node accepts two input paths *West* (*W*) for same cylinder input or *North* (*N*) for outer cylinder input or new injection for the outermost cylinder. Only a single input can be present at the same time, and it is routed to *East* (*E*) to the same cylinder or *South* (*S*) path to the inner cylinder by turning on the proper Semiconductor Optical Amplifier (SOA) switch (SW). Each provides power amplification to balance the power loss at the node due to tap and 3-dB power splitter between *E* and *S* paths.

Data Vortex network combines a traffic control mechanism with deflection routing. Control signals stay in the electronic domain for a simple implementation. As seen in the routing node in Fig. 2, a control signal C_{in} dictates whether *South* path to the inner cylinder is indeed "open" or "blocking". The routing node also generates a proper C_{out} to inform the current cylinder's traffic condition for its outer cylinder node. These distributed control signals allow the neighbouring nodes to coordinate properly for the single packet processing condition for each routing node within the network. Every time a packet is to stay at the current cylinder or to the *East* path, it creates a "blocking" control C_{out} for its outer cylinder contender. In the case the outside traffic receives a "blocking" control, the packet that is intended for *South* path will be deflected by staying on its current outer cylinder and wait for the next open path in two hops. The single packet routing arrangement eliminates optical buffers within the node as the network serves as a virtual buffer when the packet travels on the cylinders.

The last cylinder is typically added for exit buffering purpose so packets are looping around in the last cylinder without changing height positions. As a result, the total number of cylinders is given by $C = \log_2 H + 1$. Note that inter-cylinder paths and intra-cylinder paths are slightly different to allow for the establishment of the control signal. The inner cylinder nodes always make the routing decision slightly earlier than their outer neighbour to inform the traffic condition, so by making the inter-cylinder travel slightly shorter, packets can arrive the same cylinder node at the same time frame regardless of their origins. Detailed traffic control and routing performance have been reported in earlier studies [4-5], and it is shown that Data Vortex network's overall routing performance is very reasonable even as the network scales up to thousands of I/O ports. In addition, many physical layer limitations have also been studied and addressed in these studies.

III. MODIFIED DATA VORTEX IMPLEMENTATION

As Data Vortex networks run at high load conditions or less redundant configurations, i.e. more input angles are attached to I/O ports for incoming traffic, the traffic backpressure could build up between the cylinders, so it takes longer to go through the network and the overall throughput also drops significantly. Due to the physical degradation of the optical signal through each node, reduction of the latency is highly desired as well as maintaining the high data throughput. There have been several approaches suggested to enhance the routing performance of the Data Vortex networks with additional hardware. The detailed analysis of cost and

 ISBN: 978-1-61208-174-8

performance comparison to the original network has been reported earlier in these studies. This paper emphasizes comparison of the two methods using buffering and using extra inter-cylinder paths respectively. Because the increase in hardware in the two methods is reasonably low and the costs are close to each other, a comparison of the two methods under the same network operation conditions will be of great interests.

A. *Buffering*

The original Data Vortex network is attractive for its bufferless operation. However, for enhanced performance, separate buffers can be added within the routing nodes with slightly more complicated routing logic. This allows for less deflection when the packets wait in the buffer instead of circulating around the cylinders. As shown in Fig. 3, an additional switch (SW3) is used to provide the third routing path to the buffer unit. However, to inform the presence of the traffic within the buffer path so that other traffic is not allowed to enter the node during the same time slot, the buffer unit must have at least two slot delays. Even though previous studies also show that two simultaneous packets routing scheme are possible and it provides much better performances, the required hardware is significantly more [6]. So this study only focuses on the buffer implementation that maintains a single packet routing principle through a two hop delay buffer unit.

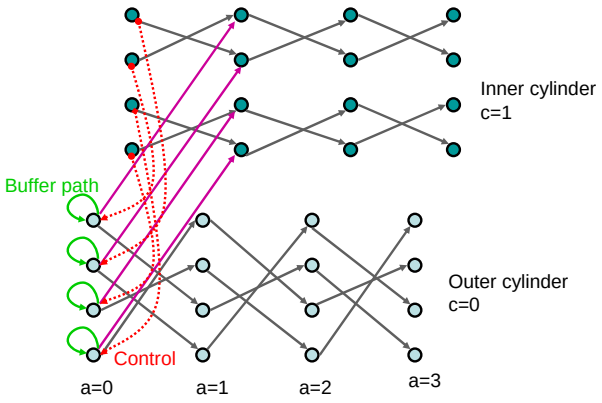

Figure 3. Data Vortex network with buffers within node shown at a=0.

The detailed node implementation is shown in Fig. 4. This implementation requires the network to have roughly 50% more hardware in number of switches and in routing paths compared to that in the original network. There is slight modification of routing logic within the node. If a packet is not able to reach S output, it will be sent to the buffer unit and be routed within the same node in two time slots. If the buffer packet is entering the node, it will not accept W or N input from the other nodes to maintain the single packet routing rule. Overall the priority is given to the packet within the buffer, and if there is no buffer traffic, then the same cylinder traffic gets the priority over the outer cylinder traffic as that in the original Data Vortex network. The additional control signal has to inform both the same cylinder neighbour and the outer cylinder neighbour to avoid contention.

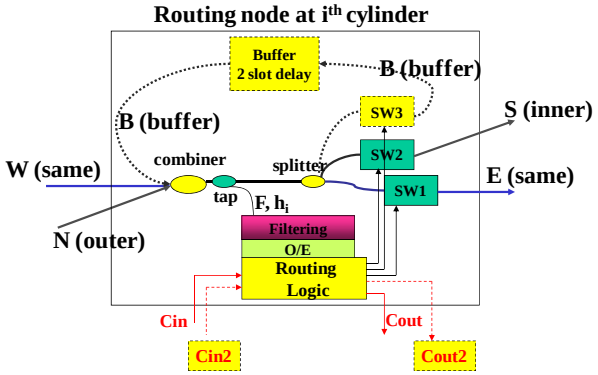

Figure 4. Routing node with buffer implementation: a 2-slot delay for buffer path is necessary to setup the control signal on time and additional controls C_{out2} are used to inform the state of buffer

B. *Inter cylinder paths*

In addition to buffering, there are also proposals for additional routing paths between the cylinders because these paths are critical to channel the traffic through the cylinders as fast as possible [7-8]. Lack of such routing resource would result in deflection thus the building up of the traffic backpressure. Here we allow the packet to be routed to a secondary inter-cylinder path S_2 output if there is no other traffic (from regular West and North path) entering that same node. We will only focus on this inter cylinder paths implementation, which is the same as that reported in [7] because a separate study has shown very similar results for implementations in [7] and [8] under various traffic and network conditions. An additional injection path is provided at each of the injection ports so that packets are less likely to be blocked by the traffic that is already circulating around the outermost cylinder. The setup of extra links and controls are shown in Fig. 5, and a detailed node implementation is shown in Fig. 6. The single packet routing rule is maintained for simplicity and an additional switch (SOA-SW_3) is added to provide the third routing path as shown in the routing node. In this case, an additional control is also necessary to inform the same cylinder traffic so that the traffic that goes to the regular S_1 output obtains the higher priority over the traffic that requires the S_2 output path. The height choice for the secondary inter-cylinder path must maintain the same binary bits for all the previous cylinders as those in the primary inter-cylinder path's height. As an example, for routing at c^{th} cylinder, the secondary height or the height of its S_2 path node can simply inverts the $(c+1)^{th}$ header bit of the current height.

 ISBN: 978-1-61208-174-8

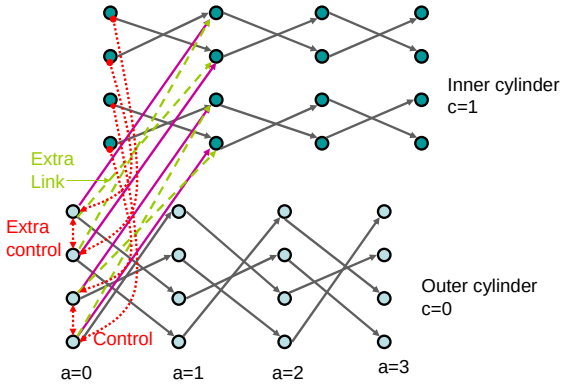

Figure 5. Additional inter cylinder path in Data Vortex network with required extra control

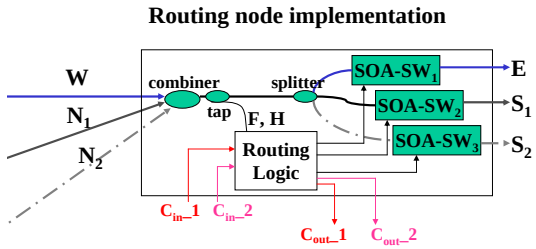

Figure 6. Modified routing node

The inter cylinder paths implementation requires about 50% more hardware in the number of switches and number of routing paths; therefore, it has comparable cost to the buffering implementation.

III. PERFORMANCE EVALUATION

In order to compare the effect of node buffering and the extra inter-cylindrical path for routing, a simulation in C/C++ is written to study the routing performance such as latency and data throughput. The networks under comparison are of the same size and same load conditions. Only random and uniform traffic pattern is studied for the purpose. Latency is measured as the average latency of packets that reach the destination for a long period of simulation time after the initial injection transient period. The network throughput is measured as the successful injection rate at the input port as previously reported. Once the packet reaches the correct target height, it exits the network immediately, therefore no angular resolution is considered in the simulation study. The networks that incorporate the two enhancement methods are compared where A=5, C=9 and H=256 as an example. Fig. 7 and Fig. 8 shows the results of the delay and throughput performance where two redundant conditions are considered. Because both methods are for performance enhancement purpose when the Data Vortex network is heavily loaded or under less redundant operation, we choose A_{in}=3 and A_{in}=5 for the study. Keep in mind, for the buffer implementation, each buffer stay requires a two packet slots delay even though the number of node hop is one.

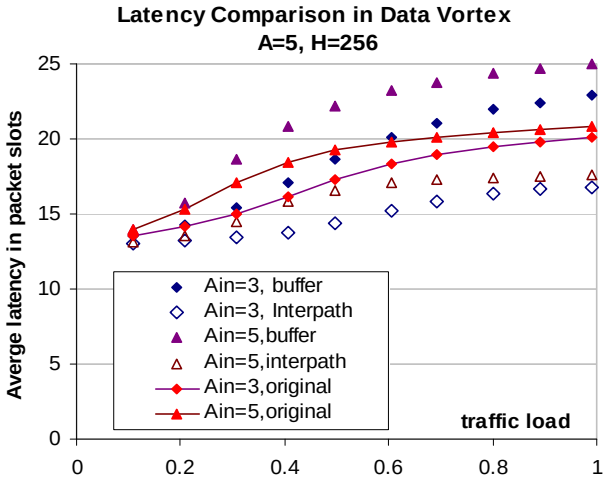

Figure 7. Latency comparison under various traffic load and redundant conditions

For comparison purpose, the original network performances are shown with the solid lines. From these results, we can see that the inter-cylinder paths provide a smaller latency in general compared to that with an additional buffer within the routing node. In fact, the latency is worse for the case of node buffering compared to the original network especially at higher load conditions and less redundant network conditions. This is mainly because of the two hop delay requirement on the buffer path for timing issue, which does not provide efficient reduction of latency even though the deflection events are reduced by keeping the packet at the open path to inner cylinder. The traffic backpressure remains significant because as the buffer packet re-enters the node for routing, there is no acceptance of additional traffic from neighbouring nodes. On the other hand, the inter-cylinder paths provide a better shared configuration of the redundant resource because when such resource is available, the additional routing paths always push more traffic through towards the inner cylinders. As a result, the traffic backpressure has been effectively reduced.

A similar performance edge in inter-cylinder path implementation is also reflected in the data throughput comparison as shown in Fig. 8. In this rather busy network conditions, the buffer implementation has little improvement compared with the original networks, while the inter-cylinder path approach provides much more visible improvement. The results follows very similar trend for the two different redundant conditions. In reference [6], more detailed cost performance study is provided on this buffer implementation in comparison to the original network and a two input buffer scheme which uses much more hardware. Similar conclusion is provided that the overall the

 ISBN: 978-1-61208-174-8

improvement in throughput and latency in this buffer scheme is rather limited and this implementation is only attractive for certain network conditions. In our comparison for more heavily loaded network conditions, the results have proved that the buffered implementation could even degrade the overall network performance once the system reaches saturation in load. On the other hand, the inter-path approach maintains the performance enhancement in both throughput and latency, and it provides a much more attractive implementation for the same amount of hardware cost. Such performance enhancement also scales to very demanding network conditions.

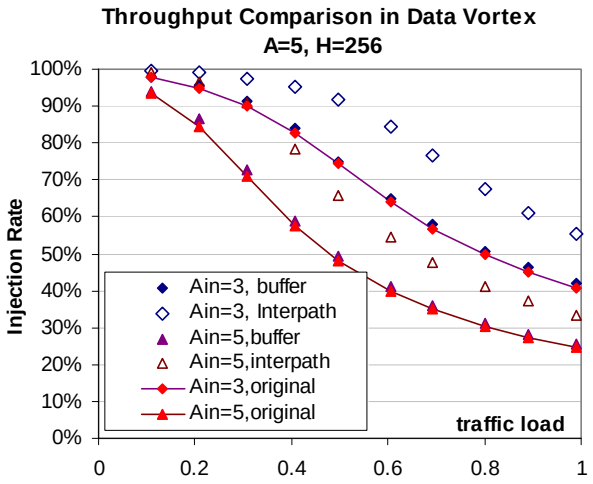

Figure 8. Throughput comparison under various load and redundant conditions

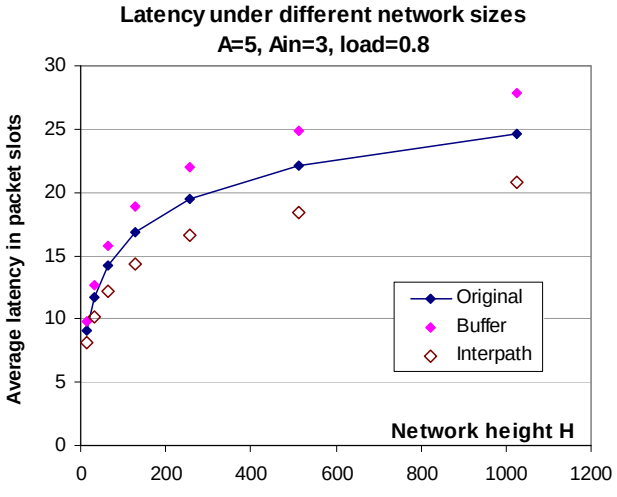

Figure 9. Latency performance comparison at different network sizes

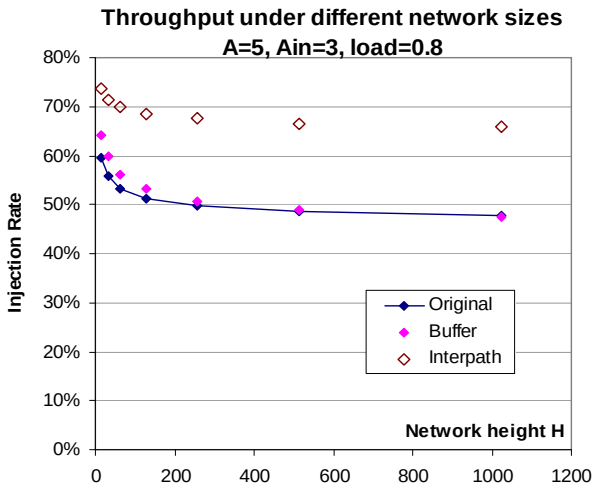

Figure 10. Throughput performance comparison at different network sizes

In order to study the scalability of such performance comparison, networks of different heights are also compared in the study. In Fig. 9 and Fig. 10, networks with A=5 and injection angles of A_{in}=3 with buffer and with inter-cylinder paths are compared and the original Data Vortex network performances are also shown as references. All cases shown are with a medium to high traffic load of 0.8. It is seen that for all network sizes, the inter-path cylinder approach provides better performance over the buffer implementation, and there is especially significant difference for larger networks.

IV. CONCLUSIONS AND FUTURE WORKS

This study focuses on two different modification schemes for Data Vortex networks improvement. With similar hardware cost and complexity, the inter-cylinder paths provide better configuration of shared redundant routing resource. Such arrangement effectively reduces the traffic backpressure present in the original network at high load network conditions, and it provides much better performance in latency and data throughput compared with the modified network with buffering implementation. Future developments in switching device integration are important and relevant for this investigation, and allow us to further quantify the benefits of different modification schemes. For future development in novel enhancement methods, researchers should consider not only the hardware cost but also the routing performance in both delay and throughput especially for less ideal network operation conditions so that a fair and effective evaluation of the proposal can be achieved.

 ISBN: 978-1-61208-174-8

REFERENCES

[1] Keren Bergman, *Optical Fiber Telecommunications, B: Systems and Networks (Editor Ivan P. Kaminow, Tingye Li, Alan E. Willner), Chapter 19*, "Optical interconnection networks in advanced computing systems", Academic Press.

[2] Ronald Luijten, Cyriel Minkenberg, Roe Hemenway, Michael Sauer, and Richard Grzybowski, "Viable opto-electronic HPC interconnect fabric", *Proceedings of the 2005 ACM/IEEE SuperComputing*, Seattle, pp. 18-18, November 2005.

[3] Roberto Gaudino, Guido A. Gavilanes Castilo, Fabio Neri, and Jorge M. Finochietto, "Can Simple Optical Switching Fabrics Scale to Terabit per Second Switch Capacities?", *Journal of Optical Communication Networks*, Vol.1, No.3, pp. B56-B68, August 2009.

[4] Odile Liboiron-Ladouceur, Assaf Shcham, Benjamin A. Small, Benjamin G.Lee, Howard Wang, Caroline P. Lai, Aleksandr Biberman, and Keren Bergman, " The Data Vortex Optical Packet Switched Interconnection Network", *Journal of Lightwave Technology*, Vol. 26, No. 13, pp. 1777-1789, July 2008.

[5] Cory Hawkins, Benjamin A. Small, D.Scott Wills, and Keren Bergman, "The Data Vortex, an All Optical Path Multicomputer Interconnection Network", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 18, Issue 3, pp. 409-420, March 2007.

[6] Assaf Shacham and Keren Bergman, "On contention resolution in the data vortex optical interconnection networks", *Journal of Optical Networking*, Vol.6, pp. 777-788, 2007.

[7] Qimin Yang, "Enhanced control and routing paths in data vortex interconnection networks", *Journal of Optical Networking*, Vol. 6, No.12, pp. 1314-1322, December 2007.

[8] Neha Sharma, D. Chadha, and Vinod Chandra, "The augmented data vortex switch fabric: An all-optical packet switched interconnection network with enhanced fault tolerance", *Optical Switching and Networking*, Elsevier, Vol. 4**,** pp. 92-105, 2007.

[9] Qimin Yang, "Performance Evaluation of *k*-ary Data Vortex Networks with Bufferless and Buffered Routing Nodes", *Asia Photonics and Communication Conference (ACP) 2009*, pp. 1-2, Shanghai, November 2009.

 ISBN: 978-1-61208-174-8