World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:7, 2011

Speedup of Data Vortex Network Architecture

Qimin Yang

***Abstract*—**In this paper, 3x3 routing nodes are proposed to provide speedup and parallel processing capability in Data Vortex network architectures. The new design not only significantly improves network throughput and latency, but also eliminates the need for distributive traffic control mechanism originally embedded among nodes and the need for nodal buffering. The cost effectiveness is studied by a comparison study with the previously proposed 2-input buffered networks, and considerable performance enhancement can be achieved with similar or lower cost of hardware. Unlike previous implementation, the network leaves small probability of contention, therefore, the packet drop rate must be kept low for such implementation to be feasible and attractive, and it can be achieved with proper choice of operation conditions.

***Keyword*—**Data Vortex, Packet Switch, Interconnection network, deflection, Network-on-chip

I. BACKGROUND

THERE has been tremendous demand and development in using photonics in multi-core processors and computing systems. In particular, recent growths in silicon photonic devices such as modulators, switches and detectors have made Photonic Network-on-chip (NOC) feasible in an integrated form, and such networks become competitive with its electronic counterpart in bandwidth, power and scalability [1-3]. At the same time, different network architectures have been proposed to best utilize these new devices and best combine both electronic and photonic technologies. Among such efforts, Data Vortex network provides a good example because it greatly facilitates optical implementation with minimal routing logic as well as no or minimal optical buffering while it scales to support thousands of processor I/O nodes and each runs at hundreds of Gbit/s [4-5]. Wavelength stacking are utilized for both header encoding and data encoding for best efficiency and simplicity, therefore very high capacity, small latency, high scalability can be achieved at the same time [2][4]. Current prototypes of smaller size networks are based on 1x1 SOA switches because of its nano-second switching speed and good physical cascadability [5], but newly developed micro-resonator switching devices can also be implemented because of faster switching speed and low loss for cascading performance [3]. Proper integration technologies are required to avoid excessive waveguide crossing for such implementations. Therefore, alternative layout of the Data Vortex network should also be studied for better physical implementation that best utilizes different photonic switching devices.

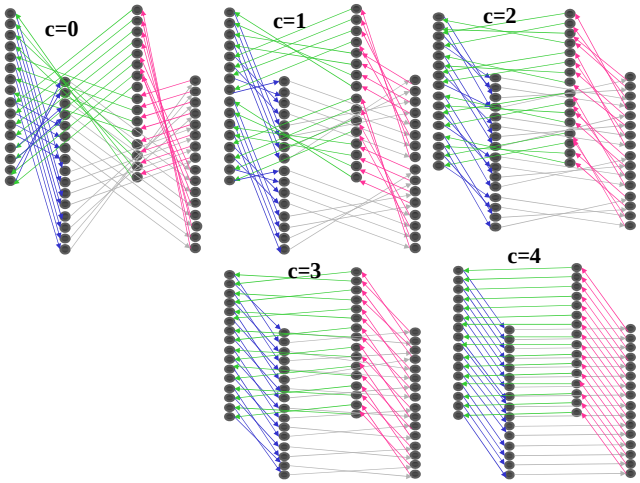

Fig.1 Routing nodes (black dots) and intra-cylinder links (arrowed wires) within Data Vortex of A=4, H=16 and C=5

In Data Vortex networks, the routing nodes are arranged in concentric cylinders in a three dimensional layout with A, H and $C = \log_2 H + 1$, designating the number of nodes along angle, height and cylinder respectively. Fig.1 shows the intra-cylinder link patterns at different cylinders of a network A=4, C=5 and H=16. These links along the same cylinder route a packet back and forth between two height groups which corresponds to a specified binary bit of the height address between "1" and "0". The inter-cylinder links (not shown in Fig.1) are parallel link pattern to simply forward a packet to an inner neighbor cylinder while maintaining its height location, and they are shown as dash lines in Fig.2 from the top view of the same network in Fig.1. As the packet reaches the correct height group, i.e. its specific header bit of the target address matches that of the node's binary height address, the packet is ready to proceed to the next cylinder level until they exit the innermost cylinder.

Q. Yang is with Engineering Department of Harvey Mudd College, Claremont, CA 91711 USA (phone: 909-607-1558; fax: 909-621-8967; Email: qimin_yang@hmc.edu).

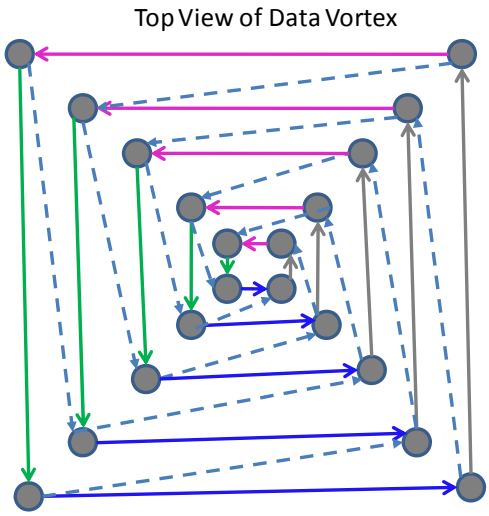

Fig.2 Top view of Data Vortex routing links

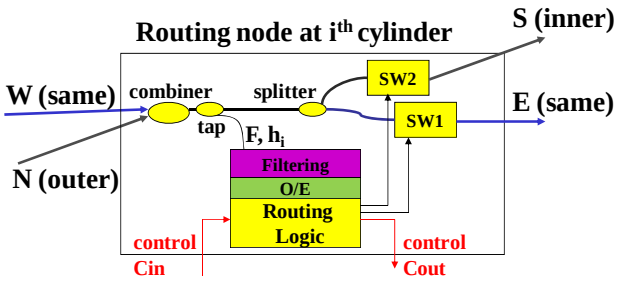

Fig.3 Original node design at the i^{th} cylinder

The Data Vortex network routing operates in synchronous and slotted manner. A distributive traffic control mechanism is used in combination with deflection routing to maintain a single packet routing principle at each node. In Fig.3 a routing node at the i^{th} cylinder is shown, where input ports are known as the West (from the same cylinder) and North (from the outer cylinder) ports, and the output ports are known as the East (to the same cylinder) and South (to the inner cylinder) ports, and the routing decision is based on the packet frame bit, which tells the presence of the packet, the i^{th} binary header bit both extracted from optical packet path as well as an electrical control bit sent from its inner competing node. Due to the orthogonal wavelength encoding in header bits and low packet rate, passive filtering and low rate O/E detector are required for header bit extraction. The control bit C_{in} or C_{out} properly permits or blocks the outer-cylinder traffic so that the single-packet-routing rule is satisfied to greatly simplify the routing node implementation. When a packet receives the blocking control, it can be deflected by staying on its current cylinder, which acts as a virtual buffer with a two hop delay penalty before it recovers to the correct height group. As shown in Fig.1, the last cylinder maintains the same height to provide additional optical buffering in addition to the electrical buffering present in the output ports. It should also be noted that, all of the inter-cylinder routing paths illustrated in Fig. 2 are of the same physical length, and so are the intra-cylinder paths. The inter-cylinder paths must be made slightly shorter than the intra-cylinder paths to allow for the establishment of the control signal, and such time difference depends on the both the generation as well as the transmission time of the control signal, which can be limited to a very small portion of the packet length, but it does require strict path alignment [4]. More details can be found in references on the Data Vortex architecture design [5-6].

II.NEW NODE BASED ON 3x3 CONFIGURATION

While the single packet routing rule greatly simplifies the routing logic at nodes, it also limits the processing capability of the node, which builds up traffic backpressure and degrade throughput and latency at heavier load conditions. There have been several approaches to enhance the network performance, and most of them use additional switching hardware in the hope that future technological development can drastically reduce the cost of switching elements within the network [6-7]. For this study, we will specifically compare our new node to previous implementation based on nodal buffering, referred to as 2- input node in [7]. Fig.4 shows such node implementation based on 6 SOA switches. With the nodal buffering capability, each node can process two packets simultaneously. The previous study has shown benefit of the 2-input node implementation, and networks of same cost (same number of switches per I/O port supported) have been compared to show improvement in network latency. For proper scheduling, the traffic control mechanism is similar to that in the original Data Vortex network, and the control is necessary to limit the inputs to two active packets including the packet within the buffer of the node.

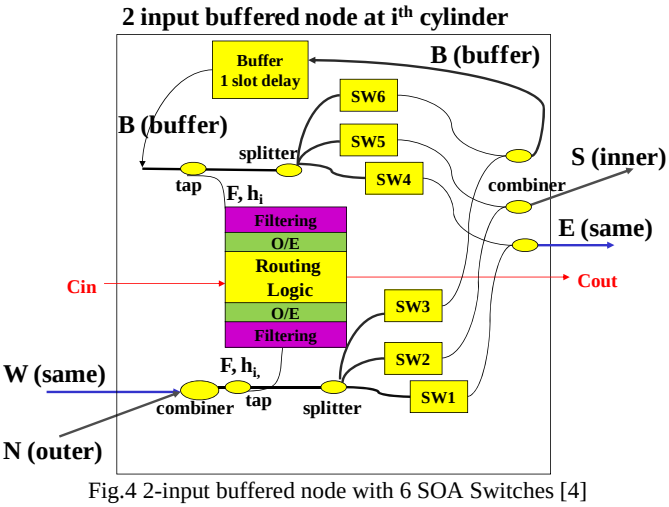

Fig.4 2-input buffered node with 6 SOA Switches [4]

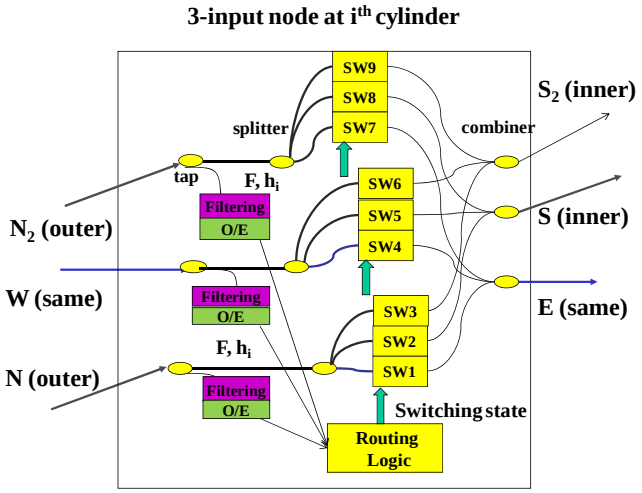

Fig.5 3-input routing node implementation

In new proposal, the capability of routing node is further extended to process three input packets simultaneously based on a 3x3 switching configuration. Instead of a buffer path, additional inter-cylinder paths S_2 or N_2 are provided. These are paths allowing for a packet to switch to a different height group based on its current binary header bit. Therefore it should provide the same function as the East path in the original network, except it allows for the packet to proceed to the next cylinder level instead of staying in the current cylinder level. The additional path S_2 should reduce the overall latency of the packets. If implemented with 1x1 SOA switch, it requires 8 or 9 switches instead of 6 switches in the 2-input buffered node networks. A detailed diagram of the routing node is shown in Fig.5, and a complete truth table is provided in Table 1.

TABLE I
ROUTING TABLE FOR 3 SIMULTANEOUS INPUTS

Cases	W	N	N_2	Routing Table
1	0	0	0	W→S_2, N→E, drop N_2
2	0	0	1	W→S_2, N→E, N_2→S
3	0	1	0	W→S_2, N_2→E, N→S
4	0	1	1	W→S_2, N_2→E, N→S,
5	1	0	0	W→S, N_2→E, N→S_2
6	1	0	1	W→S, N_2→E, N→S_2
7	1	1	0	W→S, N→E, drop N_2
8	1	1	1	W→S, N→E, drop N_2

1: To maintain its height group
0: To switch its height group

While all three input packets are accepted, the node becomes blocking in the case all three inputs desire to go to the same height group which is not possible with the combination of S, S_2 and E output paths. In our study, to further save hardware, we used 8 switch implementation (SW_9 in Fig.5 is reduced) so that N_2 incoming packet is dropped for cases 1, 7 and 8 in Table 1 when traffic contention happens. In other cases, a desired output path will be selected for the incoming packet. For cases where less than three packets simultaneously arrive, there is always a guarantee for the desired routing path for the input packets.

Since the routing node can accept all three packets simultaneously, there is no need for electronic control paths between the neighboring cylinders. As a result, there is also no need for intra-cylinder fibers to be longer than inter-cylinder fibers, and thus the synchronization of the overall network can be greatly simplified. Furthermore, synchronization at different cylinders allows for the potential global traffic control and support for quality of service (QoS) within the network, which will be studied in future researches for such benefit. A 3-input configuration also allows for the injection ports to accept two packets simultaneously and allows for two exiting packets at each output port, which leads to great speedup of the routing and thus the throughput of the network.

III. PERFORMANCE STUDY

To evaluate the routing performance of networks with the 3-input node, an event simulator in C/C++ is used with a focus on the following performance criteria: *throughput*, *mean latency*, *99.9th percentile latency* and *packet drop rate* [7]. Only random and uniform traffic is considered and no angular resolution is considered, i.e. when a packet reaches the correct height, it reaches its destination and exits the optical network. *Throughput* is defined as the steady-state number of packets per port measured at all exiting ports. To fairly compare networks of different angle (i.e. different number of exit ports), we also use the concept of *normalized throughput* by choosing one network as a reference, and the total arrival packets will be normalized to the network height (which should be the same for all networks under comparison) and the reference network's angle. *Mean latency* is the number of nodes traversed by a packet averaged over all exit packets, including the injection and exit hop. In addition, *99.9th percentile latency* is an important measurement because packets that encounter long delay, i.e. at the latency distribution tail, accumulate more noise from the SOA switches, and an unacceptably low signal-to-noise ratio in physical layer performance may lead to the discard of the packet [9]. The network performance is measured after initial period of transient when the network first gets populated, and data is collected over a sufficiently long period after such steady state has been reached for an accurate performance measurement. *Packet drop rate* is measured by the number of dropped packets normalized to the total arrival packets.

TABLE II
NETOWKRS IN COMPARISON STUDY

Type	Network Name	A	H	A_{in}	Number of Nodes $A \times C \times H$	Number of SOA switches
3-input	A_1	5	512	5	25600	204800
	B_1	5	512	3	25600	204800
	C_1	5	512	1	25600	204800
2-input	A_2	7	512	5	35840	215040
	B_2	7	512	3	35840	215040
	C_2	7	512	1	35840	215040

To make a fair comparison of the performance, the additional hardware cost must be included in consideration. Even though the elimination of the control signal in the 3-input network greatly simplifies its system implementation and cost, we still use the number of switches as our main metric of cost comparison, so the result provides a conservative estimate of the performance per cost in 3-input networks in comparison to 2-input networks. A set of networks are listed in Table 2 with similar cost factor in two different implementations. The number of angles A_{in} indicates the number of I/Os supported and condition of the network redundancy. For example, network A_1 and A_2 support same number of I/Os and have similar level of cost in terms of switching element count. A reasonably large size network is studied for high performance computing applications, in this case H=512 and network scalability has been shown at various network sizes. Because different angles are used for comparable hardware cost, it is not fair to include angular resolution for networks with larger angles; therefore, no angular resolution is included. We also measure *normalized throughput* using reference angle of the 3-input networks for a fair comparison. The I/O ports for 3-input networks are updated to reflect additional processing capability, so we can study the effect on the resulted throughput.

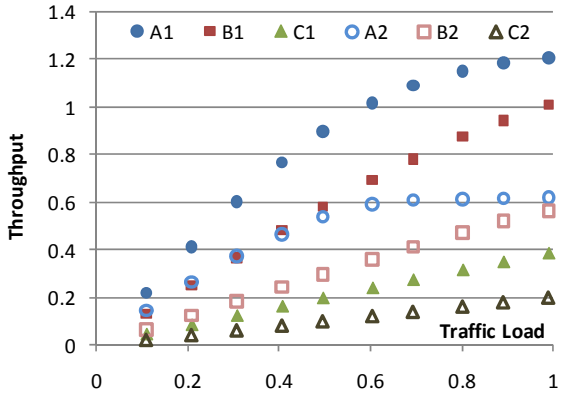

Fig.6 Throughput performance comparison of networks

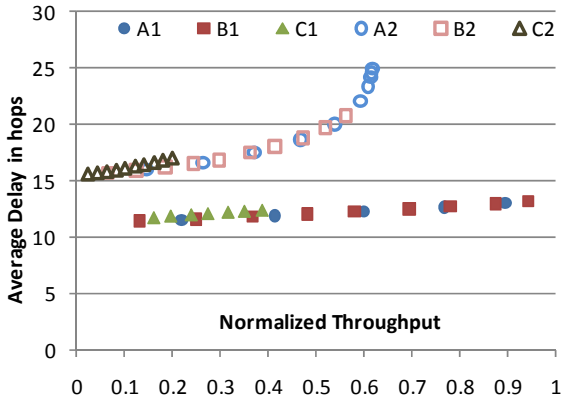

Fig.7 Latency performance comparison of networks

Fig.6 and Fig.7 show the *throughput* and *mean latency* performance of networks in Table 1 with various injection conditions. Fig.6 shows that even with the slightly less hardware (less number of nodes and no control paths), the 3-input networks provide dramatic improvement in throughput in comparison to that in 2-input buffered networks especially in A_{in}=5 and A_{in}=3 cases. For example, the 2-ipnput network's *throughput* is saturated at 0.6 in network A_2, while 3-input network can achieve as high as 1.0 in network B_1or 1.2 in network A_1. In the case of A_{in}=5, the saturation in throughput happens at a slightly higher load in A_1 than that in A_2. There is no saturation at more redundant networks with A_{in}=3 and A_{in}=1. Fig.7 shows the much smaller *mean latency* in comparison to that in the 2-input networks, and it is in a much narrower range even at high throughput operations. The deflection is greatly reduced because even when packet's header bit not matching with the node height position, it proceeds to the next cylinder through S_2 path. In addition to this speedup, additional packet processing capability also leads to larger throughput and lower mean latency.

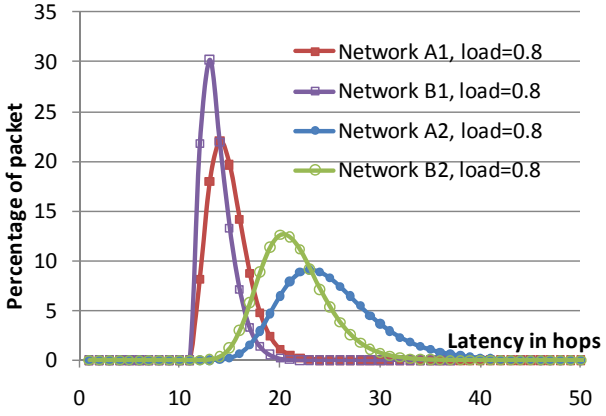

Fig.8 Latency distribution comparison under load=0.8

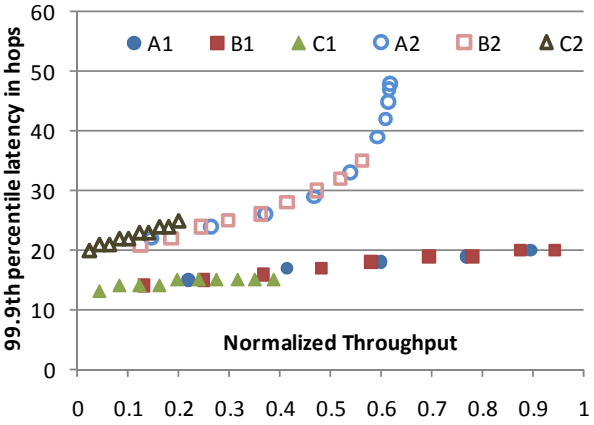

Fig.9 99.9th percentile latency comparison

In addition to the average latency performance, we also studied the latency variation within the network. This is either presented by the *latency distribution* for various load and network conditions or evaluated using the *99.9th percentile latency*. These measurements further confirm the excellent performance in packet latency. Fig. 8 shows the distribution curves of network A_1, A_2, B_1 and B_2 respectively under the

same load of 0.8. Both network A_1 and B_1 dramatically outperforms A_2 and B_2 with narrower distribution and small distribution tails. We can clearly see that 3-input networks push packets through the network much faster and more efficiently. The corresponding comparison in *99.9th percentile latency* for the same networks is shown in Fig.9 for various load conditions. It is very consistent with the average latency performance and the latency distribution curve shown earlier. In comparison to 2-input network where the 99.9% latency can increase dramatically at high throughput operations, 99.9% of packets in 3-input networks can limit their latency to under 20 hops, and this allows for great relaxation in physical performance constraint, and it is highly desirable.

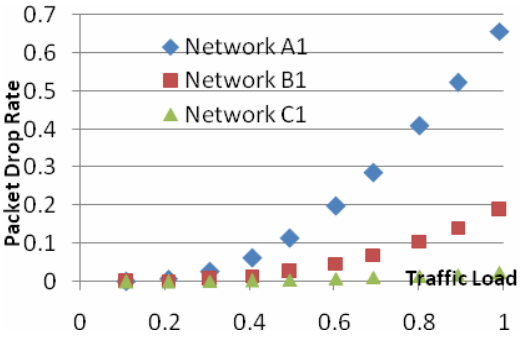

Fig.10 Packet Drop Rate in 3-input networks

The only downside of the new implementation is that it leads to a blocking network. As shown in Fig.10, while A_{in}=5 in network A_1 shows very good throughput and latency performance, there is an unacceptably high packet drop rate at high load conditions. On the other hand, at a slightly redundant network such as A_{in}=3 in network B_1, the packet drop rate becomes much smaller and degrade much slower as traffic load increases. For example, at a load of 0.8, only 10% packets are dropped while its throughput is still as high as 0.875 and its latency performance maintains very low. While $A_{in}=1$ in network C_1 provides a much smaller packet drop rate (less than 2%), its enhancement in throughput shown in Fig.6 is not significant. Therefore, a medium redundant network such as A_{in}=3would be optimum choice for implementation. Typically Packet drop requires additional signaling at the network layer for proper retransmission, which introduces additional complexity in comparison to have such signaling directly at I/O ports in the original Data Vortex network where the injection rate at I/O port is limited at high load conditions, but the network maintains non-blocking.

IV. Conclusions

This study shows significant performance benefit with the 3-input node design in Data Vortex network in comparison to 2-input nodes. The traffic control can be eliminated and no nodal buffering is required. The network can achieve global synchronization because of the elimination of control signals between cylinders. The optimum operation condition is at a reasonable level of network redundancy, under which it is able to achieve superior throughput and low latency, while maintaining a low blocking rate less than ~10% even for very large network sizes. The feasibility depends on development of high speed switching elements of 3x3, and newly developed devices such as micro-resonator switches can potentially provide easier fan out ports than SOA based switches, and it should be researched in more details with the layout of the Data Vortex network to achieve more cost effective implementations.

References

[1] Andrew Poon, Fang Xu and Xianshu Luo, "Cascaded active silicon microresonator array cross-connect circuits for WDM networks-on-chip", Photonics West, San Jose, January 2008.
[2] Madeleine Glick, Michael Dales, Derek McAuley, Tao Lin, Kevin Williams, Richard Penty, Ian White, "SWIFT: A testbed with optically switched data paths for computing applications", Proceedings of 7th International Conference on Transparent Optical Networks (ICTON), July 2005.
[3] Huimin Zhang, Yaojun Qiao, Yuefeng Ji, " A novel asymmetric optical interconnection network architecture for network-on-chip", Proceedings of International Conference on Network Infrastructure and Digital Content, November, 2009.
[4] Benjamin A Small and Karen Bergman, "Slot Timing Considerations in Optical Packet Switching Networks", IEEE Photonic Technology Letters, VOL. 17, NO. 11, pp. 2478-2481, NOVEMBER 2005.
[5] A. Shacham, B.A. Small, O. Liboiron-Ladouceur and K. Bergman, "A Fully Implemented 12x12 Data Vortex Optical Packet Switching Interconnection Network," Journal of Lightwave Technology, vol. 23, No.10, pp. 3066-3075, October, 2005.
[6] O.Liboiron-Ladouceur, B.A.Small and K.Bergman, "Physical Layer Scalability of WDM Optical Packet Interconnection Networks", J. Lightwave Technol. 24, 262-270, (2006)
[7] A. Shacham and K. Bergman, "On contention resolution in the data vortex optical interconnection network", Journal of Optical Networking, vol. 6, No.6, pp.777–788, June 2007.
[8] Neha Sharma, D.Chadha, Vinod chandra, "The augmented data vortex switch fabric: An all-optical packet switched interconnection network with enhanced fault tolerance", Optical Switching and Networking, Elsevier, 4, 92-105 (2007).
[9] Qimin Yang, "High throughput exploration of Data Vortex network", accepted for opto-electronics communication conference (OECC), July 2011.