Using Clos Switches in Area Efficient Asynchronous SDM Routers

Wei Song and Doug Edwards

School of Computer Science, University of Manchester

Manchester, M13 9PL, United Kingdom

{songw, doug}@cs.man.ac.uk

Abstract—**Most of current asynchronous on-chip routers use the virtual channel flow control method which compromises throughput and latency. Instead of using virtual channels, an asynchronous spatial division multiplexing (SDM) router has been proposed recently. The major design overhead of an SDM router is the enlarged central switch which increases proportionally to the number of virtual circuits. Clos switches provide theoretically the optimal structure for high-radix switches. A novel 2-stage Clos switch is proposed and integrated into an SDM router. It is shown that the 2-stage Clos switch significantly reduce the area consumption of the central switch and the switch allocator. It is also found that SDM and the 2-stage Clos switch reduce power consumption.**

I. INTRODUCTION

Current multiprocessor system-on-chip (MPSoC) systems are capable of integrating tens to hundreds of processing elements into one die. On-chip network or network-on-chip is the state-of-the-art chip architecture for MPSoC systems. Asynchronous on-chip networks are well-known for their low power consumption, unified network interfaces and tolerance to variations [1]. Most asynchronous on-chip networks [2]–[5] utilize the virtual channel (VC) flow control method [6], which leads to frequent reconfiguration in the central crossbar of each router. Since asynchronous routers cannot hide the reconfiguration latency as their synchronous counterparts do, they suffer from degraded throughput [7].

Instead of using virtual channels, a novel spatial division multiplexing (SDM) router has been proposed recently [7]. Rather than sharing links and switches in a time divided manner, SDM divides links and switches into virtual circuits and exclusively allocates virtual circuits to individual frames [8]. Since multiple frames share the same link or channel by occupying multiple virtual circuits, the head-of-line problem is alleviated [9]. Moreover, SDM introduces no extra reconfiguration latency as a virtual circuit is allocated to a frame instead a flit as in a VC router. It is shown that SDM outperforms VC in throughput and area overhead [7]. The major design overhead of SDM routers is the enlarged crossbar. The area overhead of the crossbar is proportional to the number of virtual circuits. Due to the tight area constraints for on-chip routers, this area overhead limits the maximal number of virtual circuits.

Clos switch is theoretically the optimal structure for high-radix switches. It has been used in intra- and inter-chip networks for high throughput performance [10]. Until recently, an asynchronous scheduler for 3-stage unbuffered Clos switch has been proposed [11]. This provides an opportunity to integrate Clos switches into SDM routers to reduce area overhead. This paper will propose a novel 2-stage Clos switch which has the minimal area consumption in all switch structures. Utilizing a simplified asynchronous Clos scheduler, this 2-stage Clos switch replaces the crossbar. It is shown that router area is significantly reduced.

The remainder of this paper is organized as follows: Section II reviews the asynchronous SDM router and its area overhead. Section III introduces the 2-stage Clos switch for SDM routers. Section IV demonstrates the implementation detail of the new SDM router. Section V provides a performance comparison between the new SDM routers using Clos switches and other router structures. The paper is finally concluded in Section VI.
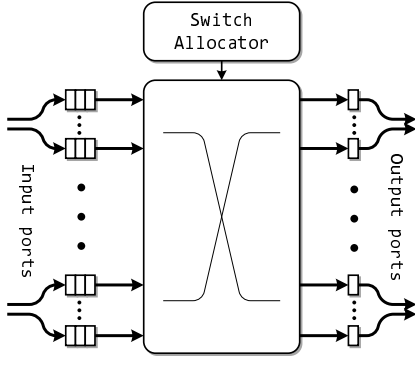
Fig. 1.   SDM router

Fig. 2.   Clos switch

II. AREA ANALYSIS OF THE CROSSBAR IN AN SDM ROUTER

The structure of SDM routers is shown in Fig. 1. The input buffer and the output buffer of each direction are spatially divided into several virtual circuits. The data width of one virtual circuit is a portion of the total data width. Every frame in transmission is exclusively allocated with a virtual circuit using the wormhole flow control method. Since SDM reduces the effective bandwidth blocked by a pausing frame, it alleviates the HOL problem and, therefore, improves the throughput performance for best-effort traffic.

Assuming the router has P bidirectional ports (directions) and M virtual circuits are implemented for each direction, the central crossbar is an $MP \times MP$ crossbar controlled by a switch allocator. The area of a crossbar is proportional to the number of cross points inside it. The area consumption of the crossbar in an SDM router can be calculated as follow:

$$A_{\mathrm{SDM}} = MP \times MP \times W/M \qquad (1)$$
$$= M \cdot P^2 W \qquad (2)$$
$$= M \cdot A_{\mathrm{Wormhole}} \qquad (3)$$

where W is the overall data width of one direction. It is shown in Equation 3 that the area of the crossbar in an SDM router is proportional to M.

III. 2-STAGE CLOS SWITCH

Traditional crossbars can be recognized as single stage non-blocking switches. Rather than using single stage switches, multi-stage switches may connect the same number of I/O ports with less number of cross points. Clos switches [12] refer
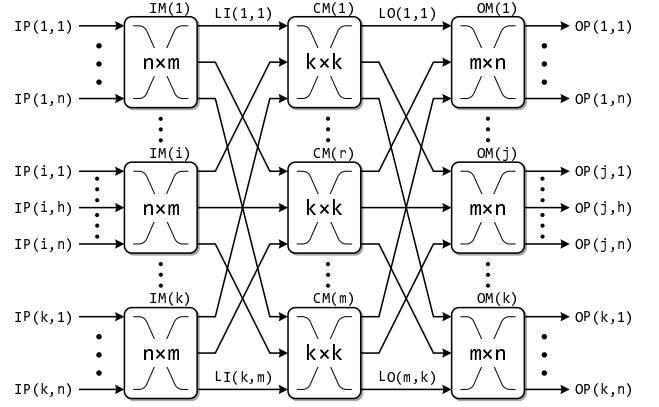
to a family of symmetric 3-stage switches[1] A Clos switch is shown in Fig. 2. The terminologies used in this paper are listed as follows:

IM	Input module at the first stage.
CM	Central module at the second stage.
OM	Output module at the third stage.
n	Number of input ports (IPs)/OPs in each IM/OM.
k	Number of IMs/OMs.
m	Number of CMs.
IM(i)	The (i)th IM.
OM(j)	The (j)th OM.
CM(r)	The (r)th CM.
IP(i,h)	The (h)th IP in IM(i).
OP(j,h)	The (h)th OP in OM(j).
C(n,k,m)	A Clos switch has m CMs and k IMs/OMs with n IPs/OPs.
N	The total number of IPs/OPs ($N = nk$).

A Clos switch consists of three stages of switching modules: IMs, CMs and OMs. The connections between adjacent stages are interleaved. The sizes of an IM, a CM and an OM switch are $n \times m$, $k \times k$ and $m \times n$ respectively. A Clos switch is non-blocking when the number of CMs is equal with n or larger ($m \geq n$). In most cases, m is set to n for the minimal area consumption. The area consumption is related to n. The overall area of a non-blocking switch is the minimum when $k = \sqrt{2N}$. However, the optimal k is not usually an integer leading to suboptimal area consumption.

---

[1]This paper adopts the narrow meaning of Clos. General Clos switches can use more than three stages and be built in an asymmetric way [13].
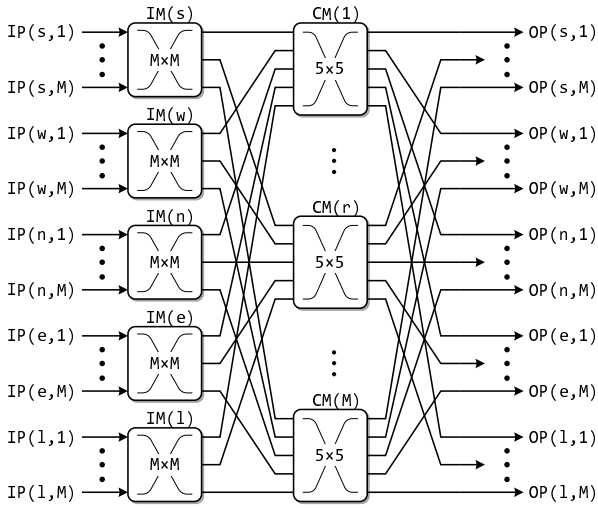
Fig. 3. 2-stage Clos switch

Fig. 4. Area of different switches

Directly utilizing the 3-stage Clos switch in SDM routers has several problems: (1) The area overhead is usually suboptimal due to the non-integer k. (2) The 3-stage Clos switch cannot be optimized by removing disabled turn models, a technique used extensively in on-chip networks. (3) It overlooks the fact that virtual circuits belonging to one direction are equivalent. This feature may be used to build Clos switches which are even smaller than the optimal ones.

A novel 2-stage Clos switch is depicted in Fig. 3. Rather than building an optimum Clos switch, k is fixed to five (**s**outh, **w**est, **n**orth, **e**ast and **l**ocal) in the 2-stage Clos switch. Every direction has an IM sized $M \times M$, where M is the number of virtual circuits implemented in each direction. The number of CMs is set to five to meet the non-blocking requirement. OMs can be implemented but the virtual circuits in each output module are equivalent. The links in OMs can be directly connected without arbitration and thus OMs are removed. The area of the 2-stages Clos switch is:

$$A_{\text{Clos,2-stg}} \le (PM^2 + MP^2) \cdot (W/M) \quad (4)$$

Fig. 4 demonstrates the area consumption of using different switch structures in a 5-port SDM router.

Compared with other switch structures, the 2-stage Clos switch has several advantages:

- As shown in Fig. 4, the 2-stage Clos switch consumes the minimal area as long as $M \le 18$.
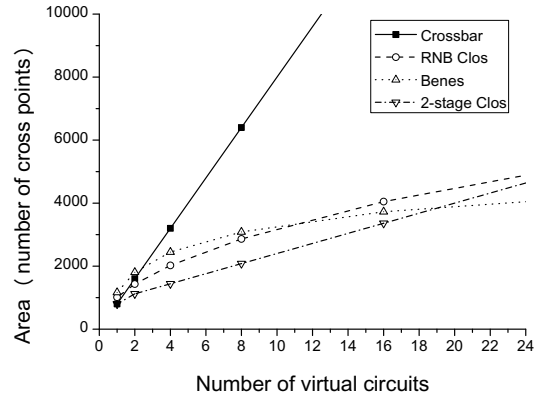- The 2-stage structure simplifies the Clos scheduler and reduces the allocation latency.
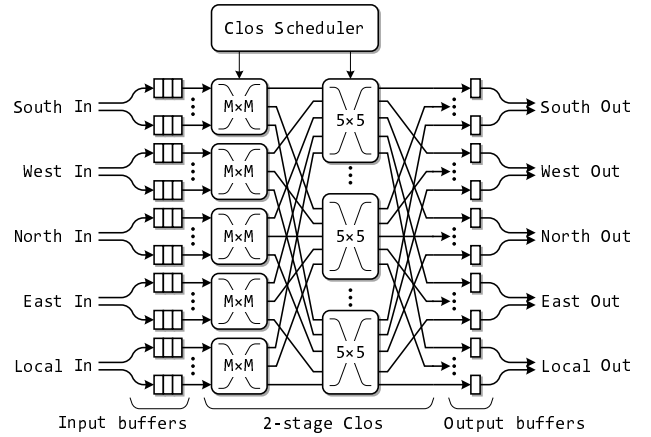

Fig. 5. Using Clos in SDM routers

- Every central module is equivalent to the crossbar in a wormhole router. The disabled turns can be removed in central modules for further area reduction.
- The latency of transmitting data through the switch is also reduced due to the removal of output modules.

IV. Router Structure

The new SDM router replacing the internal crossbar with a 2-stage Clos switch is depicted in Fig. 5. Besides the central switch, the switch allocator in the original design (Fig. 1) is also replaced by a Clos scheduler.

The implementation detail of the input buffer for a virtual circuit is illustrated in Fig. 6. Channel slicing [14] and lookahead pipelines [14], [15] have been used for high throughput and low router latency. The synchronization circuits among low-level pipelines (sub-channels) are removed according to the channel slicing technique. Extra sub-channel
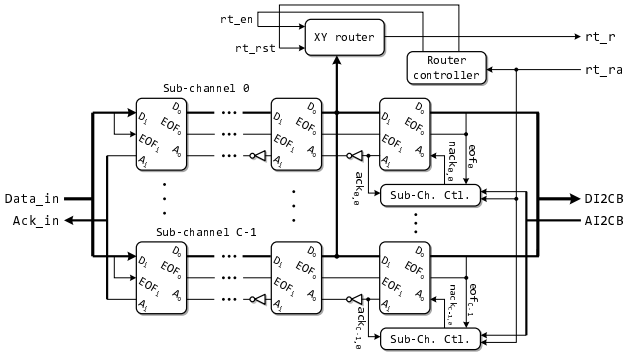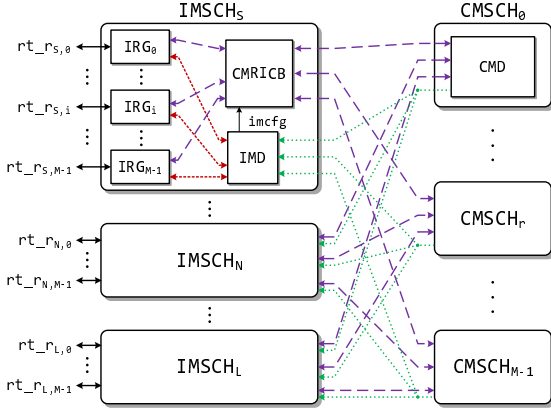
Fig. 6.   Input buffer

Fig. 7.   Clos scheduler

controllers are added in the last pipeline stages of all sub-channels in order to occasionally re-synchronize the buffer. This re-synchronization occurs when the old frame is delivered and the header flit of a new frame is going to be analysed by the XY router. The process is controlled by the router controller. The circuit implementations of sub-channel controllers and the router controller can be found in [7].

The output buffer for a virtual circuit is merely a single stage of lookahead pipelines. The pipeline stage is also divided into independent sub-channels according to the channel slicing technique. An AND gate is added on the ack line generating the early evaluated acknowledge signal [14], [15].

The internal structure of the Clos scheduler is revealed in Fig. 7. It is a simplified version from the Clos scheduler for 3-stage Clos switches [11].

Every IM or CM is dynamically reconfigured by an IM scheduler (IMSCH) or a CM scheduler (CMSCH). The CM scheduler has the same internal structure as the arbiter in a wormhole router. An IM scheduler contains M input request generators (IRGs), an IM dispatcher (IMD) running the asyn-

chronous dispatching algorithm [11] and a forwarding crossbar (CMRICB).

The routing request from an input virtual circuit (rt_r) is sent to an IRG. The request is translated into two requests: imr and cmr for IMD and CM-SCH respectively. Besides request translation, IRG also enforces the right request and acknowledge procedure to avoid false allocation.

The IM dispatcher receives all requests from the IRGs in one IM. It tries to allocate a CM to each active request using the asynchronous dispatching algorithm. A state feedback scheme is utilized in IM dispatchers: it avoids allocating occupied CMs to active requests using the state feedback from CM schedulers. It is shown that the state feedback scheme alleviates the HOL blocking in Clos switches [11]. The allocation results of IM dispatchers are sent to the forwarding crossbars (CMRICBs). The cmr signals from IRGs are forwarded to CM schedulers accordingly.

Simplified forward acting $m \times n$ arbiters [16] have been utilized in IM dispatchers to replace the area consumption multi-resource arbiters in the original design [11].

The new SDM router has been implemented into layout using the Faraday 0.13 μm standard cell library. The data width of each direction is 32 bits. Each direction has four virtual circuits, each of which is eight bits wide. The buffer depths of input and output buffers are set to one stage which is the minimal size. The router is placed in the centre of a $2.5 \times 2 : 5\text{mm}^2$ tile which provides enough space for an embedded processing element in 0.13 μm technologies [17]. The long-range links between routers incurs long latency. Two pipeline stages are inserted symmetrically on every long range link to ensure that they are not the throughput bottleneck. The pin order and position of all top-level I/O ports are manually specified; therefore, adjacent tiles can be aligned seamlessly into a mesh network.

V. PERFORMANCE ANALYSES

The new SDM router using the 2-stage Clos switch (SDM-Clos) along with an SDM router using crossbars (SDM) and a wormhole router (WH) has been implemented into layout. The post-layout area breakdown of different routers is revealed in Table I. It is shown that the area of the crossbar in an SDM router is around 3.9 times of that in a wormhole

TABLE I
AREA CONSUMPTION

	WH	SDM	SDM-Clos
IP Buf.	17,147	23,356	19,995
OP Buf.	11,034	9,016	8,893
Switch	17,754	69,701	33,417
Alloc.	909	83,485	19,884
Overall	48,170	187,872	84,054

(unit: μm^2)

TABLE II
SPEED PERFORMANCE

	WH	SDM	SDM-Clos
Cycle period	2.24	2.86	2.67
Router latency	0.92	1.18	1.29
XY router	0.64	0.68	0.63
Allocation	1.24	2.01	2.13

(unit: ns)

Fig. 8. Detailed bandwidth requirement of MPEG-4

router. The area overhead of the switch allocator is even larger than the switch area overhead. The area of the allocator built from multi-way MUTEX arbiters increases quadratically with the number of virtual circuits. Using the 2-stage Clos switch in SDM routers reduces the switch area by 52.1%, which is consistent with Fig. 4. The area of the switch allocator is also reduced by 76%.

The speed performance of different routers is demonstrated in Table II. SDM routers prolong cycle period due to the enlarged switches. The allocation latency is increased for the same reason. Compared with the SDM router using crossbars, the SDM router using the 2-stage Clos switch leads to longer router latency and allocation latency because the multi-stage switch structure is more difficult to control and deeper than the single stage crossbar. Unexpected, the cycle period is reduced. It is believed that the reduced overall area has reduced the average cell load, which produces the short cycle period.

A 4×3 mesh network has been built to test the network performance with a traffic pattern extracted from an MPEG-4 system [18]. Assuming the bandwidth is equally divided into forward and backward directions [18], the detailed bandwidth requirement of all links and routers using the XY routing algorithm is depicted in Fig. 8. The values between nodes are the bidirectional bandwidth requirement in unit of MByte/s. The total bandwidth of a router
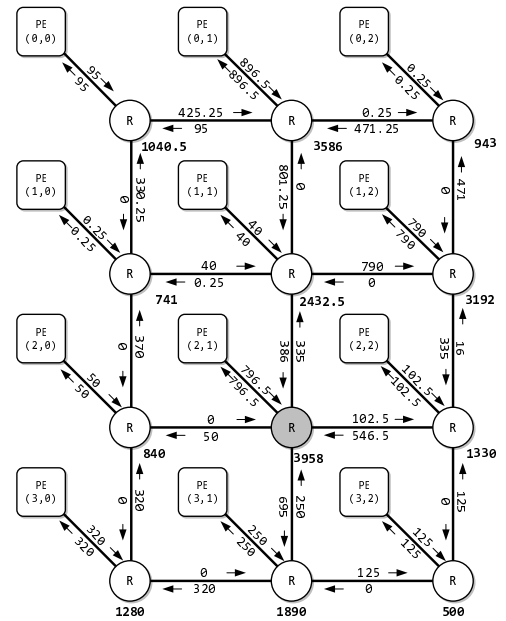
is labelled beneath it in bold. The router connected with PE(2,1) is the busiest one. 57.1% of the overall 3.446 GByte/s data goes through it.

In simulation, the expectation of the data generated by each processing element complies with the bandwidth requirement shown in Fig. 8. The payload size of each frame varies from eight to 64 bytes. The communication with a small bandwidth requirement uses short frames and the communications with more than 64 Byte/s are delivered by multiple frames with the largest 64 bytes payload.

Table III reveals the performance of different router structures. The power consumption is the power consumed by the router connected with PE(2,1), which is the busiest router. The overall throughput requirement of the MPEG-4 system is 3,466 GByte/s. It is shown that all routers achieve this throughput requirement. The average frame latencies of SDM routers are longer than the wormhole router because frames are serialized in transmission. Note that the data width for a single virtual circuit is eight bits while it is 32 bits for a link in the wormhole router. The average frame latency is not prolonged four times; therefore, the wormhole router is near its saturation point while SDM routers can support much larger traffic load. An important advantage of using SDM and the 2-stage Clos switches is their capability of reducing power consumption. As shown in the table, Using SDM routers reduce the overall power by 2.6% and

TABLE III
NETWORK PERFORMANCE FOR MPEG-4

	WH	SDM	SDM-Clos
Throughput (MByte/s)	3414.6	3458.1	3458.1
Avg. latency (ns)	108.6	239.4	248.5
Power (mW)	15.6	15.2	13.6

using the 2-stage Clos switch reduces further 10.2%.

VI. CONCLUSION

In this paper, a novel 2-stage Clos switch has been proposed for asynchronous SDM routers. This switch has been integrated into a 5-port SDM router using the Faraday 0.13 μm standard cell library. The post-layout area breakdown shows that the 2-stage Clos switch reduces the area consumption of the central switch and the switch allocator by 52.1% and 76.2% respectively. It is also shown in the speed performance that the cycle period is slightly reduced although the router latency and allocation latency are increased.

Three router implementations, a wormhole router, an SDM router using crossbars and an SDM router using the 2-stage Clos switch, have been used in a 4×3 mesh network running an MPEG-4 system. All router structures achieve the throughput requirement of the MPEG-4 system. SDM routers introduce longer frame latencies than the wormhole router due to the serialized frame transmission but they are capable of supporting larger throughput requirement. It is shown that SDM and the 2-stage Clos switch achieve lower power consumption than the wormhole router. The SDM router using the 2-stage Clos switch reduces 12.8% overall power.

REFERENCES

[1] M. Krstić, E. Grass, F. K. Gürkaynak, and P. Vivet, "Globally asynchronous, locally synchronous circuits: overview and outlook," *IEEE Design and Test of Computers*, vol. 24, no. 5, pp. 430–441, 2007.

[2] T. Felicijan and S. B. Furber, "An asynchronous on-chip network router with quality-of-service (QoS) support," in *Proc. of IEEE International SOC Conference*, September 2004, pp. 274–277.

[3] E. Beigné, F. Clermidy, P. Vivet, A. Clouard, and M. Renaudin, "An asynchronous NOC architecture providing low latency service and its multi-level design framework," in *ASYNC'05 Proceedings. 11th IEEE International Symposium on Asynchronous Circuits and Systems, 2005.*, March 2005, pp. 54–63.

[4] T. Bjerregaard and J. Sparsø, "A router architecture for connection-oriented service guarantees in the mango clockless network-on-chip," in *DATE '05 2005. Proceedings of Design, Automation and Test in Europe*, 2005, pp. 1226–1231.

[5] R. R. Dobkin, R. Ginosar, and A. Kolodny, "QNoC asynchronous router," *Integration, the VLSI Journal*, vol. 42, no. 2, pp. 103–115, March 2009.

[6] W. J. Dally, "Virtual-channel flow control," *IEEE Transactions on Parallel and Distributed Systems*, vol. 3, no. 2, pp. 194–205, March 1992.

[7] W. Song and D. Edwards, "Asynchronous spatial division multiplexing router," *Microprocessors and Microsystems*, vol. 35, no. 2, pp. 85–97, 2011.

[8] A. Leroy, D. Milojevic, D. Verkest, F. Robert, and F. Catthoor, "Concepts and implementation of spatial division multiplexing for guaranteed throughput in networks-on-chip," *IEEE Transactions on Computers*, vol. 57, no. 9, pp. 1182–1195, September 2008.

[9] C. Gómez, M. E. Gómez, P. López, and J. Duato, "Exploiting wiring resources on interconnection network: increasing path diversity," in *Proc. of Euromicro Conference on Parallel, Distributed and Network-Based Processing*, February 2008, pp. 20–29.

[10] S. Scott, D. Abts, J. Kim, and W. J. Dally, "The blackwidow high-radix clos network," in *International Symposium on Computer Architecture*, 2006, pp. 16–28.

[11] W. Song and D. Edwards, "An asynchronous routing algorithm for clos networks," in *Proc. of International Conference on Application of Concurrency to System Design*, June 2010, pp. 67–76.

[12] C. Clos, "A study of non-blocking switching networks," *Bell System Technical Journal*, vol. 32, no. 5, pp. 406–424, March 1953.

[13] *Broadband Packet Switching Technologies: A Practical Guide to ATM Switches and IP Routers.* John Wiley & Sons, Inc., 2001.

[14] W. Song and D. Edwards, "A low latency wormhole router for asynchronous on-chip networks," in *Proc. of Asia and South Pacific Design Automation Conference*, 2010, pp. 437–443.

[15] M. Singh and S. M. Nowick, "The design of high-performance dynamic asynchronous pipelines: lookahead style," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 15, no. 11, pp. 1256–1269, November 2007.

[16] S. S. Patil, "Forward acting n x m arbiter," Massachusetts Institute of Technology, Tech. Rep. 67, 1972.

[17] S. R. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, A. Singh, T. Jacob, S. Jain, V. Erraguntla, C. Roberts, Y. Hoskote, N. Borkar, and S. Borkar, "An 80-tile sub-100-W TeraFLOPS processor in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 1, pp. 29–41, january 2008.

[18] M. Kim, D. Kim, and G. E. Sobelman, "MPEG-4 performance analysis for a CDMA network-on-chip," in *Proc. of International Conference on Communications, Circuits and Systems*, May 2005, pp. 493–496.