A Robust High Speed Link Design for NoC at 65nm Technology Node

Ravi Kiran B Raghavendra†, Janet M. Wang†, Avinash Karanth Kodi‡, and Ahmed Louri†
† Dept. of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721
‡Dept. of Electrical Engineering and Computer Science, Ohio University, Athens, OH 45701
† {rraghave, wml, louri}@ece.arizona.edu, ‡kodi@ohio.edu

Abstract— Network-on-Chip (NoC) architectures provide a scalable solution to the wire delay constraints in deep submicron Very Large Scale Integrated circuit (VLSI) designs. To reduce power consumption and improve performance, we propose in this paper a technique which provides high speed link design by utilizing a compact model for repeater insertion that takes into consideration the interplay between power and delay simultaneously, as opposed to optimizing either power or delay separately. We scaled our proposed work to the 65nm technology node. We compared the proposed architecture with the other popular schemes. Using clock-data adjustment technique and buffer insertion, this new design achieves 1Tbit/s transmission speed with 0.3697 mw/Gb/S and 0.905e-3 mm^2/Gb/s.

I. INTRODUCTION

The continuous technology scaling will increase the transistor density enabling the integration of tens and hundreds of processing cores and caches on a single die in the future. Network-on-Chips (NoCs) have emerged as a promising solution for interconnecting multi-cores as they overcome the dual problems of wire delay and scalability [2], [3]. Power dissipation, area overhead and performance are the major research challenges for NoCs [4]. With the increasing need for low power and high performance architectures, NoC research has focused on optimizing buffer design [5], [6], [1], minimizing crossbar power [7], [8], and utilizing 3D interconnects [9]. Power is dissipated both for communicating data across links as well as for switching and storage within the routers [4]. The power-performance-area trade-offs are based on the design and optimization of the link. As we move towards nanometer regime, the propagation delay of inter-router link increases asymptotically as the square root of the area. According to ITRS 2007 [10], at 65 nm technology node, the line resistance per unit length is almost twice as much as the one at 180 nm. Even though the length of the inter-router links range from 2mm to 5mm, the total interconnects resistance varies from 80 ohms to 200 ohms which is about 3-4 times the combined driver and receiver resistance. Such high interconnect resistance not only dominates inter-router delay when compared to the total driver and receiver delays, but also deteriorates waveforms at the far end of the link (generally measured by eye opening of the waveform), and thus hurts the NoC system performance due to transient errors. One popular strategy used in most existing multi-core systems such as RAW, TRIPS and ASAP is to trade-off gate latency for total delay (gate delay + interconnect delay) through pipelining to achieve a scalable delay. This strategy works well when gate delay plays a dominant role for a local connection. However, the strategy is not applicable in the NoC where the interconnect delay plays a dominant role in inter-router delay. Therefore, it is important that NoC link design be optimized for power, delay and performance.

In this paper, we propose a technique which provides a high-speed link design by utilizing a compact model for repeater insertion that takes into consideration the interplay between power and delay simultaneously, as opposed to optimizing either power or delay separately. We propose clock-data adjustment to achieve high data transmission speed. By using compact device model, we formulate sizing factors for buffers inserted in inter-router link to achieve optimal delay and power. Synthesized designs using Synposys Power Compiler in the 65nm technology at 2 Ghz and 1.0 V show that the proposed design requires four repeaters with 79.6 sizing factor and provides a bandwidth in excess of 1.0Gb/s at power/bit rate of 0.3697 mW/Gb/s and area/bit rate of 0.905e-3 mm2 /Gb/s. This is a significant improvement over many proposed schemes including, RAW [11], TRIPS [12], ASAP [13] and INTEL [14] designs. Moreover, cycle accurate simulation on an 8 × 8 mesh network topology show a power reduction of 52% without significant loss in throughput for various network traffics.

In Section 2, we discuss the design of the proposed high-speed links. In Section 3, we estimate and evaluate the power, area and performance of the approach, and in Section 4, we conclude.

II. HIGH-SPEED LINK DESIGN

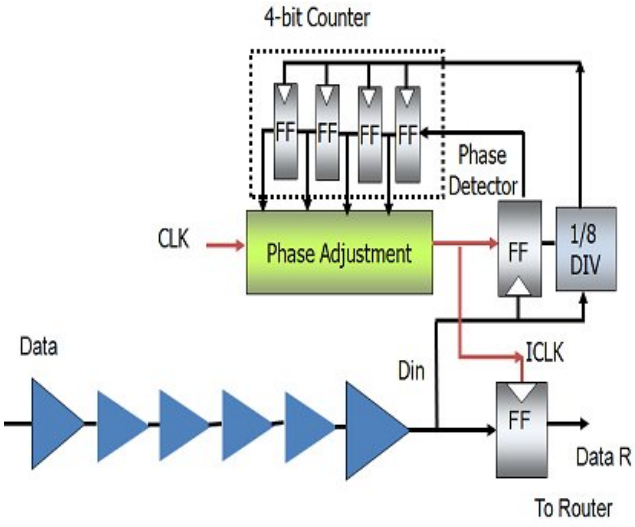
Inter-router link in most multi-core systems

consists of three parts: one driver, one receiver and one interconnect connecting the driver and the receiver. The driver and receiver are designed as buffers. Complicated link design such as [15], [16] are only used for global interconnects at inter-chip level where power and area consumptions are not constrained. Here, we achieve high performance by applying multiplexed transceiver and receiver, using precise phase generation by ring-based voltage controlled oscillators to derive multiple phases, and utilizing phase interpolators to obtain higher resolution and phase tuning. Figure 1 displays the proposed high speed link design. Figure 1(a) shows the adjustment block. Data enters the transceiver, travels though four buffers, then exist from the receiver (”Din”). Before it enters the router (”Data R”), the output of the receiver (”Din”) is adjusted together with clock through the use of a two flip-flops controlled by phase adjustment circuits. The adjustment block consists of the phase adjustment, phase detector, 1/8 divider and 4-bit delay counter.

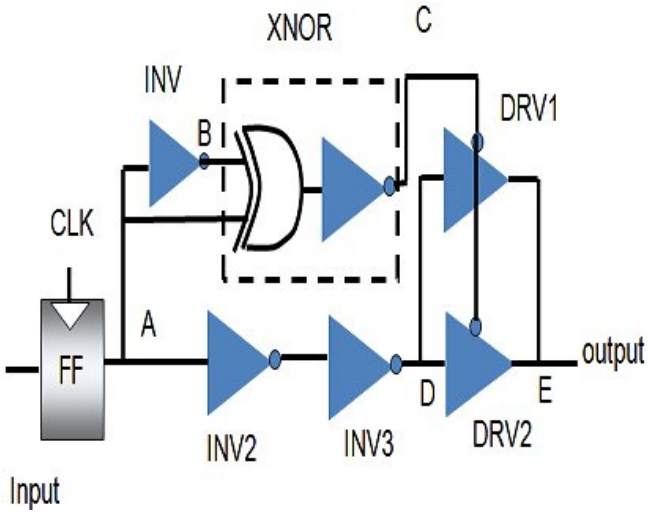
A. Transceiver and Receiver Design

The transceiver circuit contains one D Flip flops (FF), three inverters (INV1, INV2 and INV3), one XNOR gate (XNOR) and two tri-state drivers (DRV1 and DRV2). Input signals (i.e. from router) first enter one Flip-flop (FF) controlled by clock signal and then propagate through two paths (signal A and B in the Figure) converging at the output of the transceiver circuit (signal E in the Figure). INV1 delays the signal A by a small inverter delay to get signal B. INV2 and INV3 are designed to have similar delay as XNOR gate. We compare signal A and B. If A equals B, the DRV1 and DRV2 are enabled so that we have two parallel driver (DRV1 and DRV2) with double driving ability. When A and B are different, no signal is propagated as both drivers are at high impedance state. The proposed transceiver has a much simpler design compared to multiplex based transceivers used in intra-chip links while it still includes error detection feature by using XNOR. This new design doubles the driver strength by using parallel drivers while traditional on chip multi-core link designs only use on driver.

For the receiver circuit, we adopt the design from [17], which consists of three important components, the phase detection module, the control module and the phase adjustment module. The phase detection module consists of a strobe signal generation circuit and a D Flip flop. The strobe signal generator creates a strobe signal on both edges of the input data signal. The D Flip flop uses the strobe signal and the delayed clock to generate the ’down’ signal which is further used in the phase control module. The phase control module has an 8 bit shift register and has the ’down’ signal as enable. The input data signal is divided by 8 and is used as the clock for the 8 bit shift register. The phase adjustment module delays and adjusts the phase of the clock signal according to the control signal. This module has buffers that are controlled by the output bits of the shift register. These buffers adjust the clock signal to generate a new clock signal after adjustment. This process is continuous until the new clock signal iclk has its positive edge aligned with the middle of the data signal.



(a)

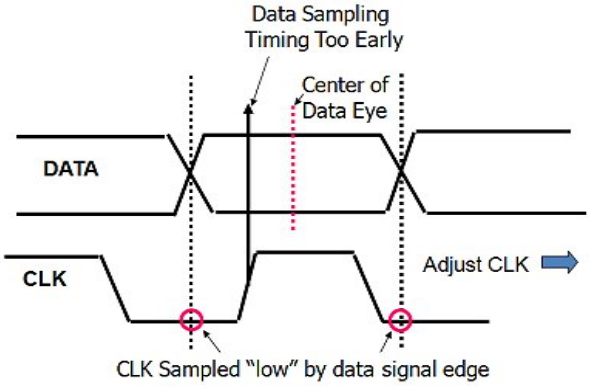


(b)

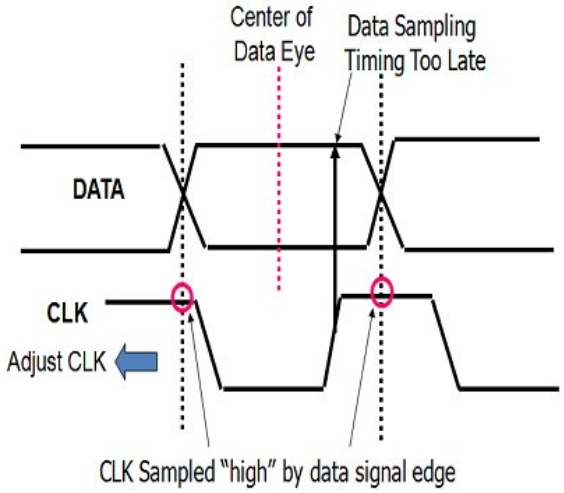
Fig.1. The adaptive, CLK-Data adjustment block design (a) Adjustment block detailed design and (b) Proposed transceiver design.

B. Clock-Data Adjustment Design

A salient feature of the proposed adjustment block design is the mutual sampling where the CLK can sample the data and the data can sample the CLK, depending on which signal arrives first. Consequently, the signaling speed is increased as the sampling frequency is achieved at the rate of the fastest signal. Figure 2 illustrates the concept of the proposed clock-data adjustment scheme. When the CLK rising edge is "earlier" than the center of the data eye, CLK is sampled as "low" by data. The proposed design thus adjusts the CLK to "right" as displayed in Figure 2(a). When the CLK rising edge is "later" than the center of the data eye, CLK is sampled "high" by data, and is adjusted to "left" as displayed in Figure 2(b). The core part of the design (in Figure 2(b)) is the feedback loop which consists of a 1/8 divider and a digital delay line. This feedback loop generates the data-sampling signal by delaying the CLK input. As such the latency of the digital delay line adaptively increases when the CLK is ahead of the data eye. Similar consideration takes place when the data eye is ahead of the CLK. The net sampling frequency of the signal is adjusted to reflect the highest sampling rate of the two signals, CLK and data. The divider further optimizes the feedback loop response by counting the phase detector output to optimize the loop response. Note that the proposed design samples the CLK at every data edge and feeds back the phase error for compensation. Therefore, this adjustment is continuous and allows the CLK to be adjusted reliably at high data rate transmission. We implemented the above circuitry using Verilog and then synthesized and simulated in Xilinx. We then generate SPICE netlist using Cadence tool. Figure 3 presents the simulation results from Xilinx showing the clock and data adjustment. Figure 4(a) and (b) details the result. Here, after one clock cycle, the clock and data are adjusted with clock positive edge at the center of data eye.

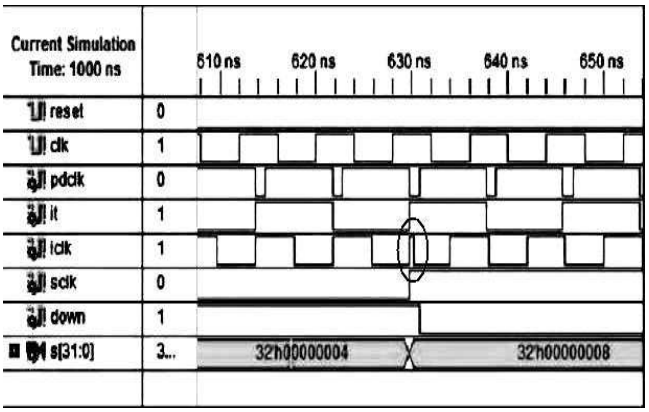


(a)

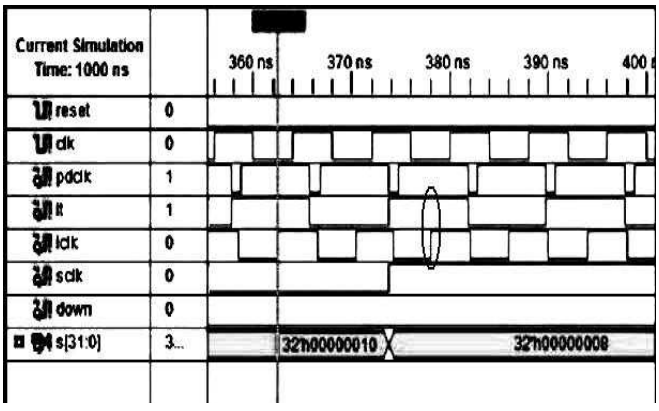


(b)

Fig. 2. The clock adjustment method (a) Clock advancement (b) Clock retreat.

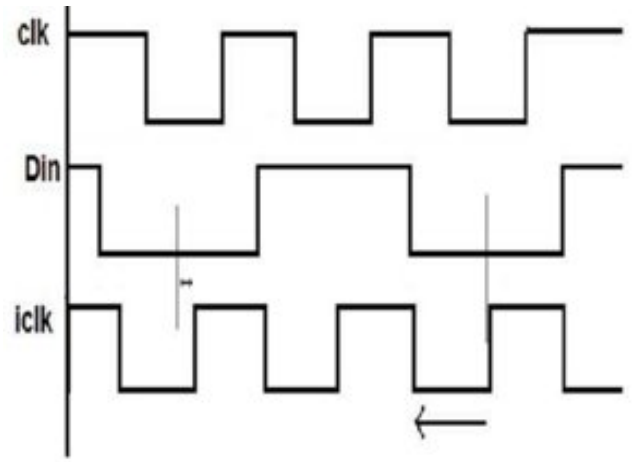


(a)

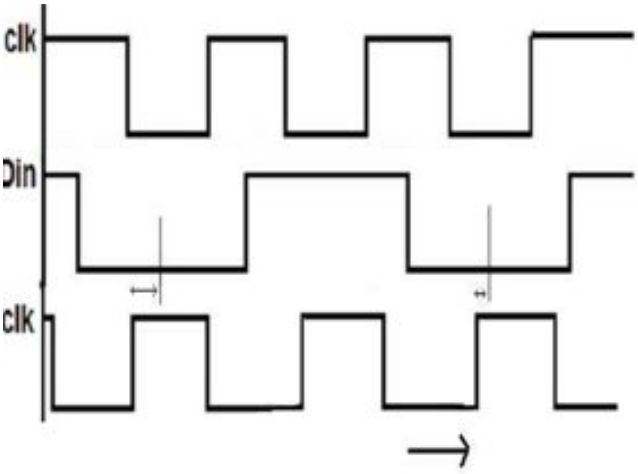


(b)

Fig. 3. Clock and data adjustment from Verilog simulated in Xilinx.



(a)



(b)

Fig. 4. Zoomed in from Figure 3 (above).

III. PERFORMANCE EVALUATION OF THE PROPOSED ARCHITECTURE

A. Power and Area Estimation for the high-speed links

To estimate the power of the high-speed links, accurate models of the inverter chain are required. Certain models either only optimize one performance metric such as delay [18], or using traditional gate model developed for 0.25 $\propto\mu$ to optimize the power and delay of the links at 65nm [19]. Another issue is that delay and power optimization are performed as two steps, first delay optimization and then power optimization [19]. In most cases, the optimization results would be very different from optimizing power first and then delay [20]. In this work, we applied finite point gate model [20] in sizing the inverter chain to determine the most-efficient design for different combination of the sizing and inverter chain as shown in Table 1. This gate model is developed for 90nm and below. In addition, it includes both delay and power into the model itself. Thus, once applied, it takes into consideration both delay and power at the same time during the optimization. Also, it should be noted that the various number of repeater stages/inverter chain directly relates to the amount of storage available on the link. That is, when there is congestion, the tri-state buffers on the link become storage units. Whenever there is no congestion, the tri-state buffers are traditional buffers/inverters and propagate data. The number of repeater stages thus represents the number of tri-state buffers (or adaptive link buffers) needed to propagate data at optimal delay and power consumption.

Table 1 shows the propagation delay (t p nsec) and average power consumption for three different cases - 4, 6 and 8 repeater (inverter) stages along the inter-router link. t(L − > H) and t (H − > L) represent the inverter output switching delay from Low to High and High to Low respectively. In each of the three cases, we vary the repeater sizing from 60 to 90, in order to determine the optimum 'design point' considering both the propagation delay and the power consumption. As seen in Table 1, for the case with 4 stages, a sizing of 80 provides the least delay and power. Similarly, in each case we determine the optimum sizing and use the corresponding average power to calculate the total link power per flit traversal. In the design case with 4 repeater stages along the link, the total link power per flit traversal for the 128-bit link is found to be 47.32 mW , and in the case with 8 stages the link consumes 57.6 mW .

TABLE I
INVERTER CHAIN (REPEATER STAGES) SIZING AND POWER VALUES

Inverter Chain	Sizing	*t(L -> H) (ns)*	*t(H -> L) (ns)*	*tp (ns)*	*Avgpwr per bit (mw)*
4	60	0.21	0.21	0.21	0.3768
4	70	0.18	0.18	0.18	0.4193
4	**80**	**0.15**	**0.16**	**0.16**	**0.3697**
4	90	0.16	0.16	0.16	0.4535
6	60	0.54	0.54	0.54	0.4253
6	70	0.31	0.31	0.31	0.3943
6	80	0.24	0.24	0.24	0.4585
6	90	0.20	0.20	0.20	0.5068
8	60	0.36	0.35	0.36	0.5928
8	70	0.25	0.28	0.27	0.6048
8	80	0.17	0.18	0.18	0.4501
8	90	0.26	0.25	0.26	0.5019

TABLE II
LINK IMPLEMENTATION COMPARISON TABLE

	ASAP	RAW	TRIPS	INTEL	This work
Technology Node (nm)	180	150	130	65	65
Speed (Ghz)	0.5	0.5	0.5	4	2
Power (mW/Gb/s)	0.5323	0.7549	0.8605	0.7089	0.3697
Area ($X10^{-3}mm^2$ /Gb/s)	6.64	5.60	4.82	0.451	0.905

B. Comparison of the Proposed High-Speed Design

We compare this work with link designs in ASAP, RAW, TRIPS and INTEL. That is, we only consider link or interconnect network related circuits such as transceiver, receiver, interconnect, and circuits used to increase data transmission speed on interconnect. While this work and INTEL design use 65 nm technology node, the rest of the designs use 180nm to 130nm technology node. To provide accurate assessment of each link design even implemented with different technology node and clock frequency, we use power/bit rate/s and area/bit rate/s to estimate the performance of each link. Table 2 lists the key parameters of each design and its performance. Figure 5 shows the performance comparison of the proposed approach with ASAP, RAW, TRIPS, and INTEL. Obviously, the lower, left corner of the figure has the least power and area usage per bit rate per second and thus the best performance, while the upper right corner has the most power and area usage per bit rate per second, which implies worst performance.

From the above experimental results, we made several conclusions. First, we incorporated some high speed design feature into on-chip link design by doubling the driver strength and adjusting the clock-data. This is the reason why the current design has 1 Tb/s. On the other hand, due to the power and area constraints in on-chip communication, we cannot afford to use complicated control and synchronization circuitry found in chip-to-chip communication, which is more constrained by the transmission speed (10 to 30T b/s) requirement rather than power or area considerations. Second, the proposed link design is for 65nm technology node where the line resistance per unit length is almost twice as much as the one at 180 nm, and the interconnect delay plays a dominant role in inter-router delay. This is very different from link designs in RAW, TRIPS and ASAP where interconnect resistance was neglected and delay scalability was achieved by trading off gate latency for total delay (gate delay + interconnect delay) through pipelining. Third we believe that the router and the inter-router links should be tackled simultaneously and not in isolation. A case in point here is the recently announced INTEL design where the router can achieve high speed in the range of 1 Tb/s but the inter-router link consumes excessive amount of power. In contrast, our proposed approach optimizes the designs of both the router and the link as shown in Table 1 or Fig 7 where the link consumes 50% less power per bit rate than INTELs.

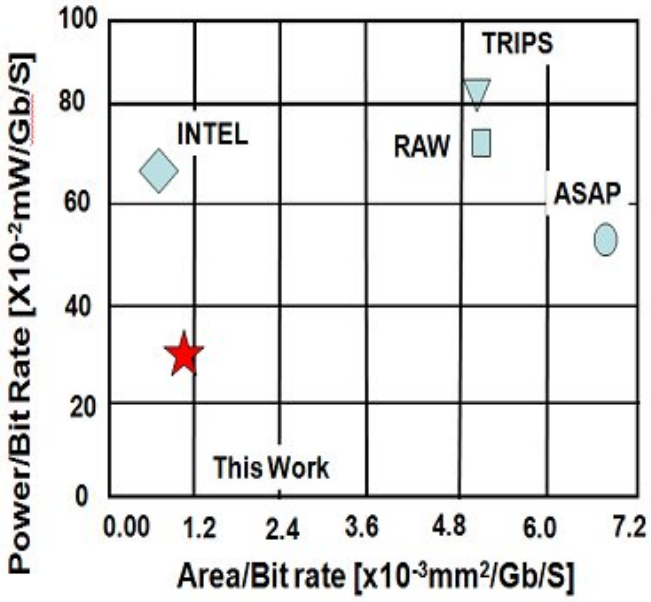


Fig. 5. Power/Bitrate and Area/Bitrate comparison.

IV. CONCLUSION

Two of the major power consuming units in an NoC architecture are router buffers and inter-router links. Recent research in NoC has been focusing on optimizing the use of buffers and improving the link power. Of notable approaches to tackle this problem, is an architecture called iDEAL. In this paper, we propose two circuit optimization techniques to improve on iDEAL. The first technique optimizes the

design of the control circuit to reduce area and power and the second technique provides high-speed inter-router link design by utilizing a compact model for repeater insertion that takes into consideration the interplay between power and delay si- multaneously, as opposed to optimizing either power or delay separately. We utilize dynamic buffer management within the router architecture to prevent degradation of performance. Our simulation results have shown that our proposed high- speed design can achieve a bandwidth in excess of 1.0Gb/s at power/bit rate of 0.3697 mW /Gb/s and area/bit rate of 0.905 mm^2 /Gb/s. Moreover, this design also results in an overall power reduction of 52% over the baseline without significant loss in performance and improve performance (throughput and latency) by more than 10%. We further compared the proposed approach with some popular NoC designs, namely, RAW, TRIPS, ASAP and INTEL designs and we found that the proposed methodology provides a significant improvement over these designs.

REFERENCES

[1] A.K. Kodi, A. Sarathy, and A. Louri, "ideal: Inter-router dual-function energy and area-efficient links for network-on-chip (noc) architectures," in Proceedings of the International Symposium on Computer Architecture (ISCA), June 2008, pp.241–250.

[2] W. J. Dally and B. Towles, "Route packets, not wires," in Proceedings of the Design Automation Conference (DAC), Las Vegas, NV, USA, June 18-22 2001.

[3] L. Benini and G. D. Micheli, "Networks on chips: A new soc paradigm," IEEE Computer, vol. 35, pp. 70–78, 2002.

[4] J. D. Owens, W. J. Dally, R. Ho, D. N. Jayasimha, S. W. Keckler, and L. S. Peh, "Research challenges for on-chip interconnection networks," IEEE Micro, vol. 27, no. 5, pp. 96–108, September-October 2007.

[5] J. Hu and R. Marculescu, "Application-specific buffer space allocation for network-on-chip router design," in Proceedings of the IEEE/ACM International Conference on Computer Aided Design (ICCAD), San Jose, CA, USA, November 7-11 2004, pp. 354–361.

[6] C. A. Nicopoulos, D. Park, J. Kim, N. Vijaykrishnan, M. S. Yousif, and C. R. Das, "Vichar: A dynamic virtual channel regulator for network- on-chip routers," in Proceedings of the 39th Annual International Symposium on Microarchitecture (MICRO), Orlando, FL, USA, December 9-13 2006, pp. 333–344.

[7] Y. Hoskote, S. Vangal, A. Singh, N. Borkar, and S. Borkar, "A 5-ghz mesh interconnect for a teraflops processor,"IEEEMicro, pp. 51–61, Sept /Oct 2007.

[8] H. S. Wang, L. S. Peh, and S. Malik, "Power-driven design of router microarchitectures in on-chip networks," in Proceedings of the 36th Annual ACM/IEEE International Symposium on Micro architecture, Washington DC, USA, December 03-05 2003, pp. 105–116.

[9] S. E. Dongkook Park, R. Das, A. K. Mishra, Y. Xie, N. Vijaykrishnan, and C. R. Das, "Mira: A multi-layered on-chip interconnect for router architecture," in Proceedings of the International Symposium on Computer Architecture (ISCA), June 2008, pp. 251–261.

[10] "International technology roadmap for semiconductors, http://www.itrs.net/links/2007itrs/home2007.htm," 2007.

[11] M. B. Taylor and et. al., "The raw microprocessor: A computational fabric for software circuits and general purpose programs," IEEE Micro, vol. 22, pp. 25 – 35, 2002.

[12] P. Gratz, C. Kim, K. Sankaralingam, H. Hanson, P. Shivakumar, S. W. Keckler, and D. Burger, "On-chip interconnection networks of the trips chip," IEEE Micro, vol. 27, pp. 41 – 50, 2007.

[13] Z. Yu and et. al., "An asynchronous array of simple processors for dsp," in IEEE ISSCC Dig. Tech. Papers, February 2006, pp. 428–429.

[14] S. R. Vangal and et. al., "An 80-tile sub-100-w teraflops processor in 65-nm cmos," IEEE Journal of Solid-State Circuits, vol. 43, pp. 29 – 41, 2008.

[15] J. Yamada, H. Adachi, Y. Mori, A. Harada, S. Okada, and H. Ando, "High-speed interconnect for a multiprocessor server using over 1tb/s crossbar," in IEEE ISSCC Digest Technical Papers, February 2006, pp. 343–352.

[16] J. Yamada and Y. Mori, "Driver driving method, driver circuit, transmission method using driver, and control circuit," U.S. Patent, no. 7,149,271, December 12 2006.

[17] J. Yamada, "Clock adjustment apparatus and method thereof," U.S. Patent, no. 2005/0281365 A1, 2005.

[18] R. Venkatesan, J. A. Davis, and J. D. Meindl, "Compact distributed rlc interconnect models part iv: unified models for time delay, crosstalk, and repeater insertion," IEEE Transactions on Electron Devices, vol. 50, no. 4, pp. 1094–1102, 2003.

[19] K. Banerjee and A. Mehrotra, "A power-optimal repeater insertion methodology for global interconnects in nanometer designs," IEEE Transactions on Electron Devices, vol. 49, no. 11, pp. 2001–2007, 2003.

[20] D. Ganesan, A. Mitev, J. M. Wang, and Y. Cao, "Finite-point gate model for fast timing and power analysis," in IEEE International Symposium on Quality Electronic Design, March 2008, pp. 657–662.