

On-Chip Global Signaling by Wave Pipelining

Masanori HASHIMOTO^{1,2}, Akira TSUCHIYA³, Hidetoshi ONODERA³

¹ Dept. ISE, Osaka University, ²PRESTO, JST, ³Dept.CCE, Kyoto University

E-mail: hasimoto@ist.osaka-u.ac.jp, {tsuchiya, onodera}@vlsi.kuee.kyoto-u.ac.jp

1-5 Yamada-oka, Suita, Osaka, 565-0871 Japan. Tel: +81-6-6879-4526, Fax: +81-6-6879-4529

Abstract : This paper discusses the signaling performance of wave pipelining over on-chip transmission lines comparing conventional signaling with CMOS static repeater insertion. We experimentally reveal that the wave pipelining over on-chip transmission lines is about ten times superior in the maximum throughput, latency and dissipates several times less energy per bit compared with the conventional signaling, whereas the required interconnect resource is comparable.

1 Introduction

With advances in LSI fabrication technology, circuit operating frequency is predicted to increase continuously, and local clock frequency will be 15GHz at 45nm technology node [1]. Although interconnect performance will improve due to lower dielectric constant of interlevel metal insulator, global signaling suffers from increase of RC time constant of long wires. High-speed and large capacity signal transmission is a big challenge in the near future. Recently to attack this problem, high-speed signaling and throughput driven interconnection are becoming a hot research topic both in design and EDA communities [2], and flip-flop (latch) pipelining [3, 4] is studied. The problem of flip-flop pipelining is large latency and power dissipation. Another approach to increase signaling throughput is wave pipelining, which is widely used in chip-to-chip and cable serial communication [5].

In this paper, we focus on wave pipelining, and evaluate the performance of on-chip global signaling. References [6, 7] discuss the limitation of interconnect performance, however the signaling performance with driver and receiver is not shown. The interconnect modeling is widely studied both theoretically and experimentally, such as Ref. [8]. Global signaling methods that focus energy efficiency are reported, eg. pulsed signaling [9], transition-aware global signaling [10]. Though the clock distribution over on-chip transmission lines is presented [11, 12], the signaling performance of wave pipelining is not sufficiently discussed so far. We, in this paper, compare conventional repeater-inserted signaling, single-end and differential signaling over on-chip lossy transmission lines, and reveal their performance difference in throughput, latency and energy per bit. The contribution of this work is to demonstrate the potential of wave pipelining over on-chip transmission lines quantitatively supported by a detailed simulation assuming a 45nm technology.

2 Evaluation Setup

We evaluate the performance of wave pipelining over the following three transmission media. The transmission length is 10mm.

Repeater A resistance-dominated global wire into which CMOS static repeaters (4x inverter) are inserted. The interval of repeater insertion is 0.5mm. (Figure 3)

Single-End An on-chip lossy transmission line driven by a CMOS static 16x inverter without termination. The receiver is a CMOS static x1 inverter. (Figure 4)

Differential An on-chip differential transmission line driven by a current mode logic (CML) driver with single-end 100 Ω terminations. The receiver is the same with the driver. (Figure 5)

We assume a 45nm process in Roadmap [1]. Figure 1 is the interconnect structure used for Repeater case, and Figure 2 is the structure for Single-End and Differential. M10 means the tenth metal layer and we assume M11 and M12 are the special thick layers for on-chip transmission lines or power/ground wires. In Figure 1, there are four signal lines ("S") and five ground wires ("G"). A shield ground wire is inserted between signal wires to suppress crosstalk noise. In M9 layer, we assume that there are orthogonal interconnects in all wire tracks. In Figure 2, there eight 4 μ m width signal lines in M12 layer and twenty ground wires in M10 used for power grid. Shielding ground wires are placed for every eight signal wires in M12 layer. When Figure 2 is used for Single-End, S2, S4, S6 and S8 wires are used as shielding wires. As for differential signaling, S1-S2, S3-S4, S5-S6 and S7-8 are differential pairs. The ground wires in the lower layers also affect the characteristics of the signal wires due to return current distribution. Therefore the ground wires in M10 layer are taken into consideration. The interconnect characteristics are modeled by a frequency dependent coupled transmission-line model [13] implemented in a circuit simulator [14].

In interconnect modeling, we do not consider shunt conductance G , because dielectric loss of insulator is negligible. The attenuation constant of RLGC transmission line is expressed as follows when $R \ll \omega L, G \ll \omega C$ [15].

$$\alpha \simeq \frac{R}{2} \sqrt{\frac{C}{L}} + \frac{G}{2} \sqrt{\frac{L}{C}}, \quad (1)$$

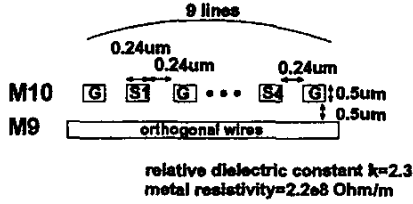


Figure 1: An interconnect structure for repeater-inserted signaling.

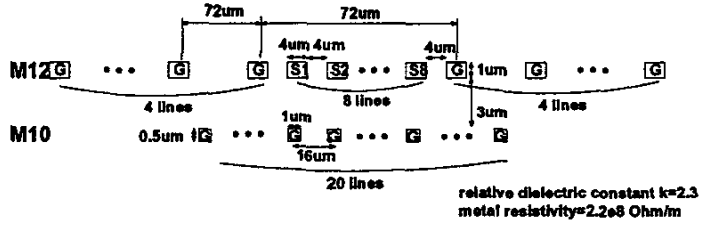


Figure 2: An interconnect structure of on-chip lossy transmission lines.

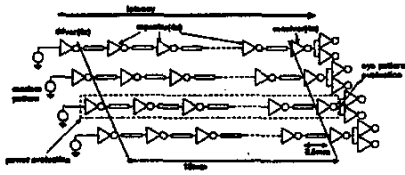


Figure 3: Experimental circuits (Repeater).

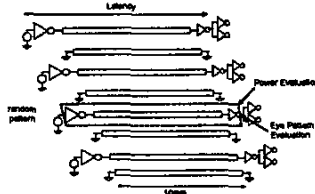


Figure 4: Experimental circuits (Single-End).

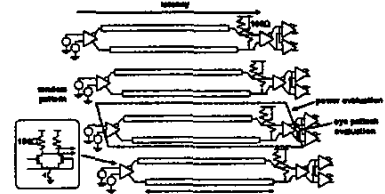


Figure 5: Experimental circuits (Differential).

where R , L and C are resistance, inductance and capacitance per unit length. The first term corresponds to conductor loss and the second term is dielectric loss. Figure 6 shows the conductor loss and the dielectric loss versus frequency. The evaluation situation is also shown in Figure 6. We here assume that resistance R is expressed as follows for simplicity.

$$R \simeq R_{dc} + R_{ac}\sqrt{f}, \quad R_{ac}: \text{fitting coefficient}, \quad (2)$$

$$G = \tan \delta \times \omega C, \quad (3)$$

where R_{dc} is the dc resistance, f is frequency and $\omega = 2\pi f$. $\tan \delta$ is the loss tangent of the insulator, and we assume that it is a constant of 0.0006, though it depends on frequency, because the magnitude of $\tan \delta$ varies at most two or three times. In on-chip situation, we can see that the conductor loss dominates the dielectric loss, and the conductor loss is over one hundred times larger than the dielectric loss even at 1THz. We therefore do not consider the shunt conductance in the experiments shown in the next section.

We use a transistor model of 50nm technology based on the ITRS roadmap [16]. The supply voltage is 0.5V. We evaluate the eye diagram at the receiver output. Each receiver has output loading of fanout 2. The input pulses of signal wires are random non-return-zero patterns that are independent of each other. The pulse shape is trapezoidal pulse with pulse period T and transition time $T/10$.

In this paper, we assume that $0.7T$ eye opening in time is necessary for all signaling, and $0.15V_{dd}$ eye opening in voltage is required for differential signaling. We evaluate the energy per bit dissipated inside the dashed squares in Figures 3-5. We also measure the latency from the driver input to the receiver output.

3 Experimental Results

Figure 7 shows the maximum throughput of three signaling methods per channel (one signal line in Single-End and Repeater or one differential pair in Differential). The throughput of Differential is the largest and it is 40Gbps. The throughputs of Single-End and Repeater are 20Gbps and 4Gbps respectively. The eye diagrams of each signaling method at the maximum throughput are shown in Figures 8, 9 and 10. The performance of the signaling over repeater-inserted interconnects is poor, and it is not sufficient for 45nm technology, whose predicted local clock frequency is 15GHz [1]. Every repeater injects crosstalk noise through mutual capacitance and inductance, and jitter accumulates though each signal line has shielding wires. The limitation of Differential comes from the attenuation of on-chip transmission lines. Inter-symbol interference (ISI) gets severe, and the eye closes. There are some techniques to reduce ISI, such as pre-emphasis, equalization, which are common in chip-to-chip or cable signal transmission, although we do not evaluate them in this paper. As shown in Eq.(1), the attenuation of transmission lines depends on capacitance C . Therefore the reduction of the dielectric constant helps to improve the signaling throughput in Differential and Single-End. If the relative dielectric constant remains 4.1 of pure SiO₂, we experimentally observe that the throughputs of Differential and Single-End decrease to 20Gbps and 10Gbps respectively.

Figure 11 shows the latency of three methods. We can see that the latency of Repeater is over 700ps and it is about 10 times larger than those of Differential and Single-End. From Figures 7 and 11, global signaling based on repeater insertion

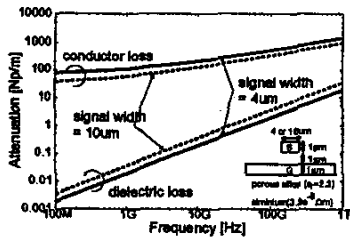


Figure 6: Attenuation.

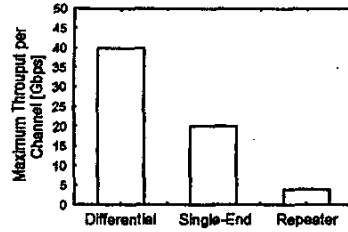


Figure 7: Max. throughput comparison.

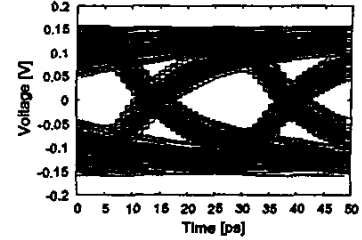


Figure 8: Eye diagram at 40Gbps (Differential).

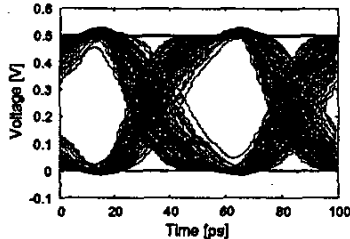


Figure 9: Eye diagram at 20Gbps (Single-End).

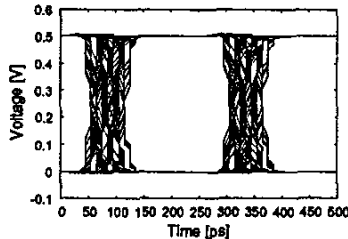


Figure 10: Eye diagram at 4Gbps (Repeater).

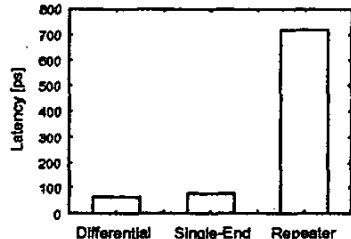


Figure 11: Latency comparison.

degrades the chip performance. On the other hand, wave pipelining over on-chip transmission lines is capable of high speed signaling with 70ps latency. This latency is basically decided by the dielectric constant of insulator. The speed of TEM wave is given by

$$s = \frac{1}{\sqrt{LC}} = \frac{c}{\sqrt{\epsilon_r}}, \quad (4)$$

where c is the light speed in vacuum and ϵ_r is the relative dielectric constant of insulator. In the simulation setup, s becomes 1.98×10^8 m/s. The latency of Differential is 64ps, which is reasonable since it consists of the time of flight (51ps) and the driver and receiver delay times.

Figure 12 shows the relation between energy per bit and throughput. Differential dissipates static power, which is decided by the current sources of CML driver and receiver, and hence energy per bit becomes large when the throughput is low. However as the throughput increases, energy per bit gets small, and at 40Gbps signaling it becomes 0.073 pJ/bit, which is less than Repeater case of 0.13 pJ/bit at 4Gbps, since energy per bit is constant in Repeater case. The energy efficiency of Repeater is worse as well as the maximum throughput. In power dissipation, signaling of Single-End is the most efficient. The energy per bit is 0.031 pJ/bit. The end of the transmission line is open-end, and hence static power is not consumed. Also as throughput increases, the energy per bit decreases, because the wire is not charged fully at every transition [9, 11].

We next discuss the required interconnect resource. Suppose we need to design 160Gbps interconnection between memory and a processor. Figure 14 shows the comparison in interconnect resource. In the case of Differential, four differential pairs are necessary to achieve 160Gbps signaling, and hence the total width of $72\mu\text{m}$, which includes metal width and spacing, is necessary. Single-End requires eight channels, and the total width becomes $144\mu\text{m}$. It is twice of Differential signaling. As for Repeater case, it becomes $153.5\mu\text{m}$. Although it is difficult to compare the required interconnect resource due to the

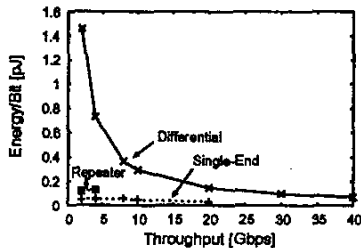


Figure 12: Relation between throughput and energy per bit.

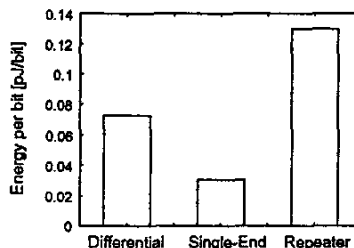


Figure 13: Comparison in energy per bit.

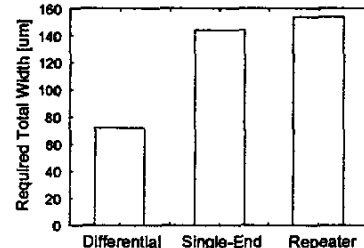


Figure 14: Comparison in wire resource required for 160Gbps signaling.

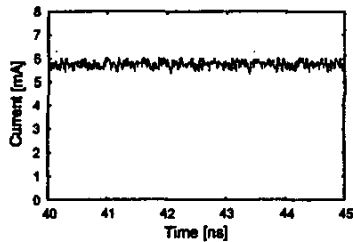


Figure 15: Power current (Differential, 40Gbps).

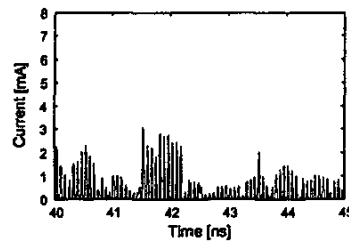


Figure 16: Power current (Repeater, 4Gbps).

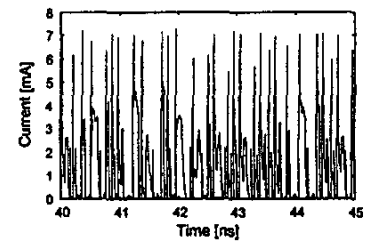


Figure 17: Power current (Single-End, 20Gbps).

difference in the metal thickness, the signaling over on-chip transmission lines is not so expensive even in wire resource.

Figures 15, 16 and 17 show the current supplied by the power supply. The current in Differential signaling is almost constant. On the other hand, the current in Repeater and Single-End varies drastically, which causes severe di/dt noise. We can see that Differential signaling is much friendly to power supply network, and it may mitigate a potential problem of on-chip simultaneous switching noise.

4 Conclusion

In this paper, we experimentally evaluate and compare the performance of three on-chip global signaling methods. Signaling over on-chip transmission lines by wave pipelining is about ten times superior in the maximum throughput and latency to the conventional signaling with CMOS static repeater insertion. Also in the required energy per bit, signaling over on-chip transmission lines is several times better, and the required interconnect resource is comparable. The results reveal that wave pipelining using on-chip transmission lines should replace the conventional global signaling method based on repeater insertion. In the comparison of single-end signaling with CMOS static driver and receiver without termination to the differential signaling with CML driver and the single-end termination, the former is superior in energy per bit where as the latter has a good characteristics of flat current consumption.

Acknowledgment

This work is supported in part by the 21st Century COE Program (Grant No. 14213201). The authors thank Akinori Shinmyo and Kenji Furusawa for their contribution in CML design and technical discussions.

References

- [1] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors", 2003 ed., 2003.
- [2] T. Lin and L. T. Pileggi, "Throughput-Driven IC Communication Fabric Synthesis," *Proc. ICCAD*, pp.274–279, 2002.
- [3] L. Zhang, Y. Hu and C. Chen, "Statistical Timing Analysis in Sequential Circuit for On-Chip Global Interconnect Pipelining," *Proc. DAC*, pp.904–907, 2004.
- [4] V. Nookala and S. Sapatnekar, "A Method for Correcting the Functionality of a Wire-Pipelined Circuit," *Proc. DAC*, pp.570–575, 2004.
- [5] William J. Dally and John W. Poulton, "Digital Systems Engineering," Cambridge University Press, 1998.
- [6] A. Tsuchiya, M. Hashimoto and H. Onodera, "Performance Limitation of On-chip Global Interconnects for High-speed Signaling," *Proc. CICC*, to appear.
- [7] D. A. B. Miller and M. H. Özaktaş, "Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture," *Journal of Parallel Distributed Computing*, vol.41, no.1, pp.42–52, 1997.
- [8] A. Deutsch, P. W. Coteus, G. V. Kopsay, H. H. Smith, C. W. Surovic, B. L. Krauter, D. C. Edelstein and P. J. Restle, "On-Chip Wiring Design Challenges for Gigahertz Operation," *Proc. IEEE*, Vol. 89, No. 4, pp. 529–555, April 2001.
- [9] P. Wang, G. Pei and E. C.-C. Kan, "Pulsed Wave Interconnect," *IEEE Trans. on VLSI Systems*, Vol. 12, No. 5, pp.453–463, May 2004.
- [10] H. Kaul, D. Sylvester, "Low-Power On-Chip Communication Based on Transition-Aware Global Signaling (TAGS)," *IEEE Trans. on VLSI Systems*, Vol 12, No. 5, pp.464–476, 2004.
- [11] M. Mizuno, K. Anjo, Y. Sumi, H. Wakabayashi, T. Mogami, T. Horiuchi, and M. Yamashina, "On-chip multi-GHz clocking with transmission lines," *Proc. ISSCC*, pp.366–367, 2000.
- [12] S. Tam, R. D. Limaye and U. N. Desai, "Clock Generation and Distribution for the 130-nm Itanium 2 Processor With 6-MB On-Die L3 Cache," *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 4, pp.636–642, 2004.
- [13] Dmitri Borisovich Kuznetsov and José E. Schutt-Ainé, "Optimal Transient Simulation of Transmission Lines," *IEEE Trans. Circuits and Systems*, vol.43, no.2, pp.110–121, Feb 1996.
- [14] Avant! Corporation and Avant! subsidiary, "Star-Hspice Manual", 2003.
- [15] C.-K. Cheng, J. Lillis, S. Lin, and N. H. Chang, "Interconnect Analysis and Synthesis," A Wiley-Interscience Publication., 2000.
- [16] K. Inagaki and T. Sakurai, "Standard SPICE model based on ITRS'99," http://lowpower.iis.u-tokyo.ac.jp/~ina/index_e.html.