Performance Comparison of Synchronous and Asynchronous VLSI Systems

1G－1

Metehan ÖZCAN Takashi NANYA

Research Center for Advanced Science and Technology, The University of Tokyo

1 Introduction

Scaling integrated circuit (IC) dimensions has significant impact on the performance of VLSI systems. Although switching performance and power consumption can be improved, interconnection delays are becoming more dominant in future technologies. It is necessary to suppress the performance degradation due to the increased interconnection delays in scaled technologies. Scaling the interconnection and the insulator thickness by factors smaller than the scaling factor, using multilayer interconnections, and dividing wires by means of repeaters are the mainly used methods [1].

Due to its average-case behavior and nonexistence of clock skew, asynchronous systems are suggested as an alternative for current synchronous systems. This paper analyses the clock skew in synchronous systems and the average-case behavior of asynchronous systems in detail. Based on this analysis, comparison of synchronous and asynchronous systems is made for different technology generations.

2 Clock Skew

Clock skew due to variances in line parameters and supply voltage can be expressed as

$$\underline{\tau}\ ln\left(1-\frac{V_T}{V_{DD}}\right)-\overline{\tau}\ ln\left(1-\frac{\overline{V}_T}{V_{DD}}\right)$$

where τ is the time constant of the clock line. For 20% variance clock skew is equal to 0.69τ, whereas for 10% variance it is equal to 0.34τ [2].

Furthermore, this time constant is dependent on the path length of the H-tree clock tree, which can be approximated for die size, L_d, as $\frac{L_d}{2}+\frac{L_d}{4}+\frac{L_d}{8}+\frac{L_d}{16}+\cdots\approx L_d$.

If repeaters, usually referred as clock buffers, are used and line thickness and insulator thickness are increased by a factor of 2 for upper k levels, the resulting clock line will have a line constant equivalent to a line of length

$$\frac{L_d}{2^{2k+1}}+\frac{L_d}{2^{2k}}+\cdots+\frac{L_d}{2^{k+2}}+\frac{L_d}{2^k}.$$

For instance, for $k=2$, i.e. two levels of thicker interconnections are introduced, resulting line constant is approximately $0.34L_d$. By similar calculations line constant can be obtained as $0.092L_d$ for $k=4$, and $0.18L_d$ for $k=3$.

In present-day processors, additional 3 or 4 layers of metal is used for interconnections. Since x- and y-dimensions require different layers, this corresponds to having $k=2$.

For 200 MHz DEC ALPHA processor, clock skew is equivalent to a line delay of length $0.34L_d$. This value is $0.33L_d$ for 80 MHz PowerPC^{TM} processor, $0.55L_d$ for 100 MHz Pentium^{TM} processor, $0.22L_d$ for 500 MHz NEC processor, and $0.13L_d$ for 300 MHz Alpha RISC processor. They all agree with the value of $0.34L_d$ for $k=2$.

3 Comparison in Scaled Technology

If the set-up and hold times of storage elements are ignored for simplicity, the possible clock period for a synchronous system can be formulated as

$$T_S = Max[GateDelay + WireDelay + ClockSkew],$$

which can be rewritten as $dg_s + w_{ij}^{max} + s_{ij}$, where d is the basic gate delay, g_s is maximum number of gates between two registers, w_{ij}^{max} is maximum wire delay between two registers, and s_{ij} is maximum value for clock skew.

A similar equation can also be derived for the asynchronous basic data processing time:

$$T_A = Mean[ClosedPathGateDelay + ClosedPathWireDelay]$$

which can be rewritten as $dg_a + w_{ij}^{avg}$, where g_a is average number of gates between two registers and w_{ij}^{avg} is average line delay.

For the synchronous case, maximum line distance between two registers may be approximated as $2L_d$. To decrease the maximum wire delay, repeaters and thick wires may be used for long distance global interconnections. If we use k repeaters, total RC constant of the maximum wire can be diminished to a line of length $2L_d/k$. For $k=4$, RC constant of maximum length wire becomes proportional to $L_d/2$, and for $k=16$ $L_d/8$. Furthermore from Section 2, maximum clock skew due to line parameters is equivalent to a line delay of length between $L_d/3$ and $L_d/6$. Therefore, possible clock period for the synchronous case can be rewritten as $dg_s + (0.50 \sim 0.13L_d) + (0.33 \sim 0.17L_d)$, or

$$T_S = dg_s + (0.56 \pm 0.27)L_d \quad Note\ that\ T_A = dg_a + w_{ij}^{avg}$$

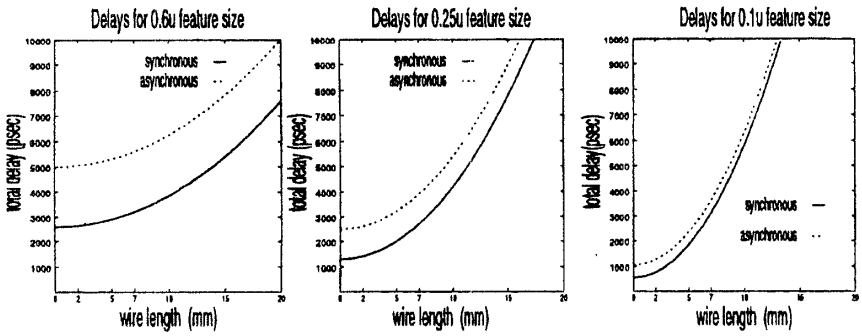

Figure 1: Data rates in 0.6μ, 0.25μ, and 0.1μ tech.

3.1 Evaluations for TITAC I chip

TITAC I is an asynchronous version of an 8-bit von Neuman microprocessor based on the delay-insensitive model with the isochronic-forks assumption [3]. TITAC has been fabricated as a CMOS gate array using 1.0μ-rule silicon gate technology.

To find the effective average wire length and logic depth of TITAC I chip, instruction set measurements are used for evaluation. For this purpose, the hypothetical DLX machine [4] is taken as the base machine. To calculate average values, first, wire lengths and logic depths for each instruction of TITAC I is calculated. After mapping the instructions of TITAC I with instructions of DLX, weighted average values are calculated. Average logic depth is 31 against the worst-case value of 65, and average wire length is 74mm against the worst-case value of 138mm. Therefore, it can be concluded that average gate delay and wire delay values are half of the corresponding worst case values.

3.2 Comparison for Different Generations

Figure 1 give the data processing times of synchronous and asynchronous processors for three technologies. The size of a reasonable chip varies from 10 to 15 mm. The difference between two curves for the same wire length is due to extra gates needed by asynchronous logic. In the these figures, a logic depth of 13 is assumed for the synchronous processor, whereas 25 for the asynchronous counterpart. Therefore, asynchronous processors are to compensate their extra logic delay by having shorter average path delays.

When the Figure 1 is observed in detail for wire lengths of 3-12 mm, it can be concluded that asynchronous processors cannot reach the performance of synchronous processors in 0.6μ technology. Because gate delays are much larger than wire delays in that technology sizes. As the technology size is scaled down, the lag between curves for synchronous and asynchronous processors becomes smaller.

This is due to smaller gate delays in scaled technologies. Therefore, in scaled technologies asynchronous processors will be comparatively more feasible. For instance, for a typical maximum wire length of 6 mm, synchronous processors will have a data rate of 400 MHz in 0.10μ. Having an average wire length smaller than 5 mm in 0.10μ technology will be enough for an asynchronous processor to perform better than synchronous counterpart.

4 Conclusion

With the estimated ultimate feature size of · 200 Angstroms, there is still a tenfold miniaturization potential from todays one quarter micron technologies. However, interconnections are becoming the bottleneck, thus the main challenge, in ultra high-speed VLSI systems.

If the average wire delay of asynchronous systems could be reduced to the orders of a line delay of length one fifth or one tenth of the chip size, they can compete and even surpass their synchronous counterparts in advanced technologies. For this purpose, methods applied for reducing *RC* constants of clock lines in synchronous systems should also be employed in asynchronous systems. Therefore, multilayer interconnections with very thick layers and repeaters should be applied and migrated to asynchronous design methodology, which constitutes the primary candidate for future continuation of this work.

References

[1] H. B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI,* Addison-Wesley, MA, 1990.

[2] K. Saraswat and F. Mohammadi, "Effect of scaling of interconnections on the time delay of VLSI circuits," *IEEE J. of Solid-State Circuits,* Vol. SC–17, No. 2, April 1982.

[3] T. Nanya, Y. Ueno, M. Kuwako, and A. Takamura, "TITAC: Design of a quasi-delay-insensitive microprocessor," *IEEE Design and Test,* vol.11, No.2, pp.50–63, Summer 1994.

[4] J. L. Hennessy and D. A. Patterson, *Computer Architecture: a Quantitative Approach,* M. Kaufman, CA, 1990.