Performance Issues on Micropipelines

Chih–Ming Chang and Shih–Lien Lu

Department of Electrical and Computer Engineering
Oregon State University, Corvallis, Oregon 97331
TEL:(503)737–2980 FAX:(503)737–1300
sllu@ece.orst.edu

Abstract — In this paper, the performance of asynchronous linear micropipelines is investigated. By introducing logic delays for C–elements, linear micropipelines with both fixed and variable physical stage–delays can be fully described by a set of governing equations. Simulation results show that the output loop delay(=1/throughput) could be less than, equal to or greater than the worst stage–delay(maximum stage–delay) within the pipeline. This unexpected phenomenon is explained through the governing equation and the expression of performance upper bound derived in this paper.

1. Introduction

There is a resurgence in the research and development of asynchronous circuits and systems recently due to the continued push for higher performance computation systems at lower power and cost [1–10]. Several works which investigated the performance analysis have been reported. Two of them discussed pipelines with ring architecture, one for 4–phase [11] and the other for 2–phase [12]. In addition, a work that can be applied to more general architecture was shown in [13]. In this paper we investigate the performance of 2–phase, self–timed linear micropipelines[14] with both fixed and variable stage–delays using a different approach. This approach allows us to visualize some interesting phenomena for micropipelines.

It is generally known that for a micropipeline with fixed stage–delays, its output loop delay equals the maximum stage–delay within the pipeline. Is it still true for micropipelines which have variable stage–delays? The answer is no. The governing equation and the expression of performance upper bound derived in this paper explain this unexpected phenomenon. In addition, the performance bound also provides a guideline in designing efficient asynchronous micropipelines.

2. Delay Modelling of C–elements

A C–element coordinates two adjacent stages such that handshaking is appropriately performed between them. In order to model the whole micropipeline mathematically, *physical* delays and *logic* delays, as shown in Figure 1, are introduced for C–element modelling.

Physical delay, denoted as d in Figure 1(a), for a C–element is defined as the time interval between input tokens(data token and space token) consumption and the generation of the output token. It can be further simplified as shown at the right hand side of Figure 1(a). For stage i, the total physical stage–delay $D_i = F_i + B_i$, where F_i is called *forward delay* and B_i is named *backward delay.*

Logic delay is defined as the time period that one token must wait for the other token in order for the C–element to fire. Figure 1(b) shows the logic delay model. It is important to know that logic delay is a function of time(tokens). To maintain the consistency of notation through out the whole pipeline, the *(j+1)th* data token is always matched and operated by the C–element with the *jth* space token(dark color). After firing, the output tokens(gray color) are numbered as *j+1*. According to logic delay definition and firing rule of C–element, $D_i^{24}(j+1,j) * D_i^{42}(j+1,j) = 0$ for all the j, where $D_i^{24}(j+1,j)$ is the logic delay for C_i viewed from stage *i–1* and $D_i^{42}(j+1,j)$ is the logic delay for C_i viewed from stage i. For compact representation, the *2nd* token index of logic delay is omitted. The micropipelines under investigation are assumed initially empty; therefore, $D_i^{24}(1,0) = D_i^{24}(1) = 0$ and $D_i^{42}(1,0) = D_i^{42}(1)$ is undefined for all the i. Note that all the delays, including logic delays, mentioned in this paper are greater than or equal to zero.

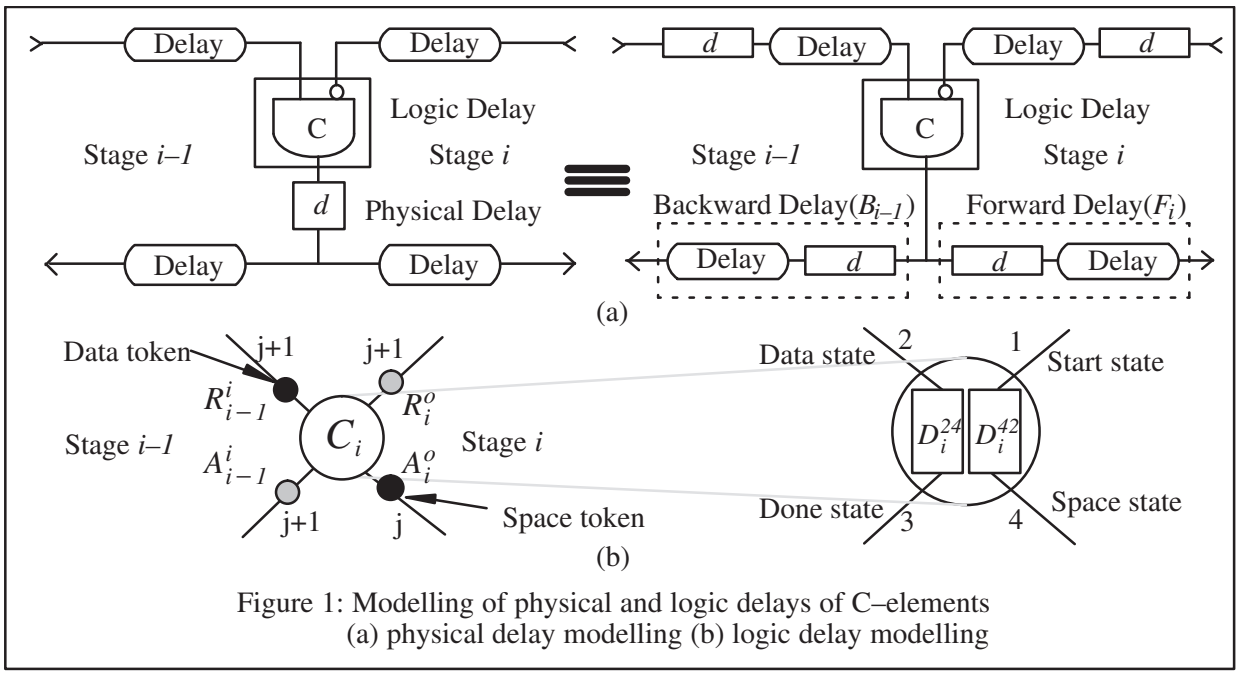


Figure 1: Modelling of physical and logic delays of C–elements
(a) physical delay modelling (b) logic delay modelling

3. Micropipelines with Fixed Delays

The performance of fixed–delay micropipelines is studied by introducing some definitions and theoretical development based upon C–element delay model mentioned in section 2. To conserve space, most of the proofs are omitted. A more detailed treatment can be found in [15,16].

Definition: A loop i is defined as a circle that token will travel around within stage i. The *ith* loop delay is the time required for a token to travel loop i. There are two different loop delays defined in Figure 2. They are

$$T^1_i(j+1) = F_i + D^{24}_{i+1}(j+1) + B_i + D^{42}_i(j+2)$$
$$= D_i + D^{24}_{i+1}(j+1) + D^{42}_i(j+2)$$
$$T^3_i(j+2) = B_i + D^{42}_i(j+2) + F_i + D^{24}_{i+1}(j+2)$$
$$= D_i + D^{42}_i(j+2) + D^{24}_{i+1}(j+2),\ j \geq 0.$$

Definition: The output loop delay is defined as $T^1_i(j+1)$ for $i=m+1$. That is, as shown in Figure 3, $T^1_{m+1}(j+1) = D_{out} + D^{42}_{m+1}(j+2)$. The pipeline performance(throughput) $T(j+1)$ is then defined as the reciprocal of output loop delay. That is, $T(j+1) = 1/(D_{out} + D^{42}_{m+1}(j+2))$.

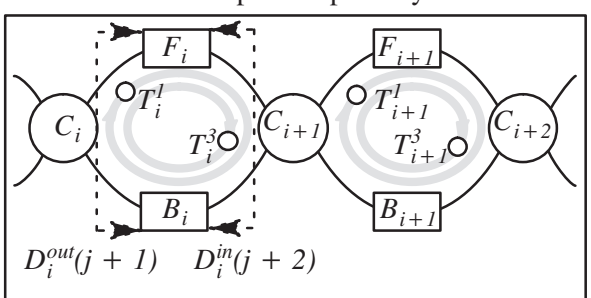


Figure 2: Definitions of loop delays

Theorem 1: *(Equal loop delay theorem.)* For C–element C_{i+1}, $T^3_i(j+2)=T^1_{i+1}(j+1)$ for all the i and j. This is the governing equation for each C–element.

Definition: Referring to the right side of Figure 1(b), the time needed for a token to travel from the done state back to the data state within stage i is called *equivalent input delay* $D^{in}_i(j+2)$. The time for a token to travel from the start state to the space state within stage i is called *equivalent output delay* $D^{out}_i(j+1)$. From Figure 2, we have

$D^{in}_i(j+2) = D^{42}_i(j+2) + D_i$,and

$D^{out}_i(j+1) = D^{24}_{i+1}(j+1) + D_i$

Theorem 2: Given a m–stage micropipeline, equivalent input delay for each stage is bounded by $D_i \leq D^{in}_i(j+2) \leq Max(D_{in}, D_1.., D_i)$ for $1 \leq i \leq m$ and logic delay $D^{42}_k(j+2)$ is bounded by $0 \leq D^{42}_k(j+2) \leq Max(0, D^{in}_{k-1}(j+2) - D_k)$ for $1 \leq k \leq m+1$, where $D^{in}_0(j+2) = D_{in}$ is the (environmental) input delay and $D_{m+1}=D_{out}$ is the (environmental) output delay. This is true for all the $j \geq 0$.

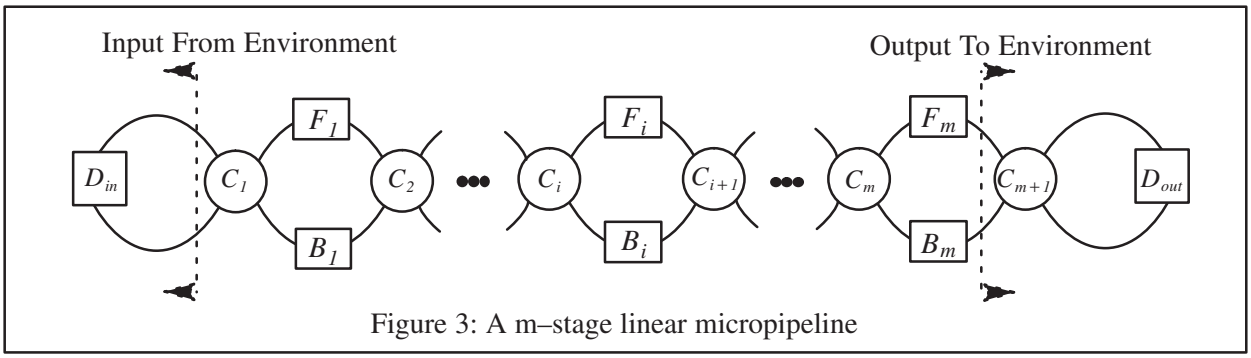


Figure 3: A m–stage linear micropipeline

<Proof> Consider C_1 in Figure 3. According to theorem 1,

$D_{in} + D_1^{24}(j+2) = D_1 + D_2^{24}(j+1) + D_1^{42}(j+2)$

If $D_{in} \geq D_1 + D_2^{24}(j+1)$ then $D_1^{24}(j+2) = 0$ and

$$\begin{aligned} D_1^{42}(j+2) &= D_{in} - D_1 - D_2^{24}(j+1) \\ &\leq D_{in} - D_1 \end{aligned}$$

$$\begin{aligned} \Rightarrow D_1^{in}(j+2) &= D_1 + D_1^{42}(j+2) \\ &= D_{in} - D_2^{24}(j+1) \leq D_{in} \end{aligned} \quad (1)$$

If $D_{in} \leq D_1 + D_2^{24}(j+1)$ then $D_1^{42}(j+2) = 0$ and

$$\Rightarrow D_1^{in}(j+2) = D_1 + D_1^{42}(j+2) = D_1 \quad (2)$$

Combining Eq. (1) and (2),

$$\Rightarrow D_1^{in}(j+2) \leq Max(D_{in}, D_1) \quad (3)$$

From above, we also notice that

$$\Rightarrow D_1^{42}(j+2) \leq Max(0, D_{in} - D_1) \quad (4)$$

Consider C_2 in Figure 3. According to theorem 1,

$D_1^{in}(j+2) + D_2^{24}(j+2)$

$= D_2 + D_3^{24}(j+1) + D_2^{42}(j+2)$

If $D_1^{in}(j+2) \geq D_2 + D_3^{24}(j+1)$ then $D_2^{24}(j+2)=0$ and

$$\begin{aligned} D_2^{42}(j+2) &= D_1^{in}(j+2) - D_2 - D_3^{24}(j+1) \\ &\leq D_1^{in}(j+2) - D_2 \end{aligned}$$

$$\begin{aligned} \Rightarrow D_2^{in}(j+2) &= D_2 + D_2^{42}(j+2) \\ &= D_1^{in}(j+2) - D_3^{24}(j+1) \\ &\leq D_1^{in}(j+2) \end{aligned} \quad (5)$$

If $D_1^{in}(j+2) \leq D_2 + D_3^{24}(j+1)$ then $D_2^{42}(j+2)=0$

$$\Rightarrow D_2^{in}(j+2) = D_2 + D_2^{42}(j+2) = D_2 \quad (6)$$

Combining Eq. (5) and (6),

$$\begin{aligned} \Rightarrow D_2^{in}(j+2) &\leq Max(D_1^{in}(j+2), D_2) \\ &\leq Max(Max(D_{in}, D_1), D_2) \\ &\leq Max(D_{in}, D_1, D_2) \end{aligned} \quad (7)$$

From above, we also notice that

$$\Rightarrow D_2^{42}(j+2) \leq Max(0, D_1^{in}(j+2) - D_2) \quad (8)$$

The conclusion can be readily reached if similar analysis applies to all the stages. □

Corollary 2.1: If $Max(D_{in}, D_1 \ldots D_i) = D_n$, where $1 \leq i \leq m$ and $n \leq i$, then $D_n^{42}(j+2) = 0$ for all the j.

Theorem 3: Given a m–stage linear micropipeline, equivalent output delay for each stage is bounded by $D_i \leq D_i^{out}(j+1) \leq Max(D_{out}, D_m, ..., D_i)$ for $1 \leq i \leq m$, where D_{out} is the (environmental) output delay. This is true for all the $j \geq 0$.

Corollary 3.1: If $Max(D_i \ldots D_m, D_{out}) = D_n$, where $1 \leq i \leq m$ and $n \geq i$, then $D_{n+1}^{24}(j+1) = 0$ for all the j.

Theorem 4: If $D_{i-1}^{in}(j+2) \geq D_i$ and $D_{i-1}^{in}(j+2) \geq D_{i+1}^{out}(j+1)$ for all the j, then $D_i^{24}(j+1) = 0$, $D_{i+1}^{24}(j+1)=0$ and $D_i^{42}(j+2)=D_{i-1}^{in}(j+2) - D_i$.

Corollary 4.1:If $Max(D_{in},D_1,...D_m,D_{out})=D_n$, then $D_k^{42}(j+2) = D_n - D_k$ and $D_{k+1}^{24}(j+1) = 0$ for $n+1 \leq k \leq m+1$ and for all the $j \geq 0$.

Corollary 4.2: If $Max(D_{in},D_1,...D_m,D_{out})=D_n$, then $D_n^{in}(j+2) = D_{n+1}^{in}(j+2) = ... = D_m^{in}(j+2) = D_n$ for all the $j \geq 0$.

According to the definition and from Corollary 4.1, 4.2, if $Max(D_{in},D_1,...D_m,D_{out})=D_n$, then the throughput

$$\begin{aligned} T(j+1) &= 1/(D_{out} + D_{m+1}^{42}(j+2)) \\ &= 1/(D_{out} + D_n - D_{out}) = 1/D_n. \end{aligned}$$

Note that throughput is not a function of token index j. It is a constant and equal to the reciprocal of maximum stage–delay within the whole linear micropipeline. Figure 4 shows the logic and equivalent input/output delays of stages following the max delay stage n.

4. Micropipelines with Variable Delays

The performance(throughput) of micropipelines becomes interesting when all the delays, including forward, backward, input and output delays,

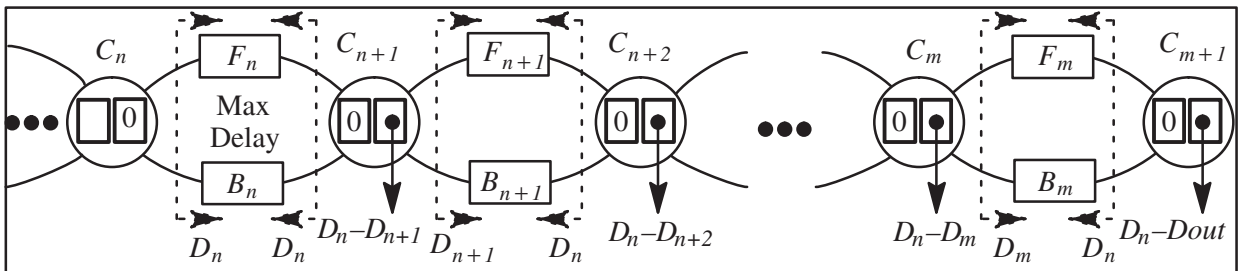


Figure 4: Logic and equivalent input/output delay analysis for the stages following max delay stage.

are not fixed any more. That is, they are variable and are functions of token index j. More precisely, the *ith* forward delay, *ith* backward delay, input delay and output delay are represented as $F_i(j+1)$, $B_i(j+1)$, $D_{in}(j+2)$ and $D_{out}(j+1)$, respectively.

Most of theorems and corollaries stated in section 3 are still applicable to the micropipelines with variable delays. Before we proceed further, it is worth restating the two different loop delays.

$$T_i^1(j+1) = F_i(j+1) + D_{i+1}^{24}(j+1) + B_i(j+1) + D_i^{42}(j+2)$$

$$T_i^3(j+2) = B_i(j+1) + D_i^{42}(j+2) + F_i(j+2) + D_{i+1}^{24}(j+2)$$

Since it is difficult to get an analytical solution in closed form for the throughput(this involves solving a set of non–linear difference equations), we will give lower and upper bounds. As mentioned in section 3, given a m–stage linear micropipeline, its throughput is

$$T(j+1) = 1/T_{m+1}^1(j+1) = 1/(D_{out}(j+1) + D_{m+1}^{42}(j+2))$$

Using similar approach as in theorem 2, we can conclude that the output loop delay has lower bound

$$T_{m+1}^1(j+1) \geq DF_{m+1}(j+1) \quad (9)$$

,and upper bound

$$T_{m+1}^1(j+1) \leq Max(DF_{m+1}(j+1), DB_m(j+1), DB_{m-1}(j+1) + \sum_{l=m}^{m}(F_l(j+2) - F_l(j+1)), DB_{m-2}(j+1) + \sum_{l=m-1}^{m}(F_l(j+2) - F_l(j+1)), \ldots\ldots DB_0(j+1) + \sum_{l=1}^{m}(F_l(j+2) - F_l(j+1))) \quad (10)$$

where $DF_i(j+1) = F_i(j+1) + B_i(j+1)$,
$DB_i(j+1) = F_i(j+2) + B_i(j+1)$,
$DB_0(j+1) = D_{in}(j+2)$,
$DF_{m+1}(j+1) = D_{out}(j+1)$.

4.a: Average output loop delay > worst stage–delay

The term $F_l(j+2) - F_l(j+1)$, defined as *delay slope*, plays an important role in the performance analysis of the inequality stated by (10). In the case of fixed–delay micropipelines as mentioned in section 3, the delay slope is equal to zero. In addition, $DB_i(j+1) = DF_i(j+1) = D_i$. Therefore, the inequality (10) reduces to

$$T_{m+1}^1(j+1) \leq Max(D_{in}, D_1, D_2, \ldots, D_m, D_{out}) = D_n,$$

which is identical to the previous conclusion described in section 3. The equal sign always holds for the fixed–delay case since $D_k^{24}(j+1) = 0$ for $n+1 \leq k \leq m+1$ as shown in corollary 3.1 and 4.1.

For ease of discussion and interpretation, the longest stage–delay(or worst stage–delay) at the instant of $j=k$ in a linear micropipeline is defined as $Max(D_{in}(k+2), DB_i(k+1), D_{out}(k+1))$ for $1 \leq i \leq m$. Now the question becomes: is the output loop delay bounded by the longest stage–delay for each j? Apparently, by looking at (10), the answer is no. This is because, depending on the delay pattern, the delay slope could be either greater than, equal to or less than zero. This implies that the output loop delay could be greater than, equal to or less than the longest stage–delay in a micropipeline. For better understanding, only two–stage linear micropipeline is investigated. To further simplify the complicated situation, let the delay $D_{in}(j+2)$, $DF_2(j+1)$ and $D_{out}(j+1)$ be constant. Figure 5 shows the simulation result with D_{in}=30.8, D_2=4 and D_{out}=15. From Figure 5(c), it is obvious to see that the output loop delay(maximum=39.13 at j=8) can be larger than the worst stage–delay(=31). Moreover, its average(arithmetic)output loop delay(=31.3632) is larger than the worst stage–delay(=31) also, which is contrary to common belief.

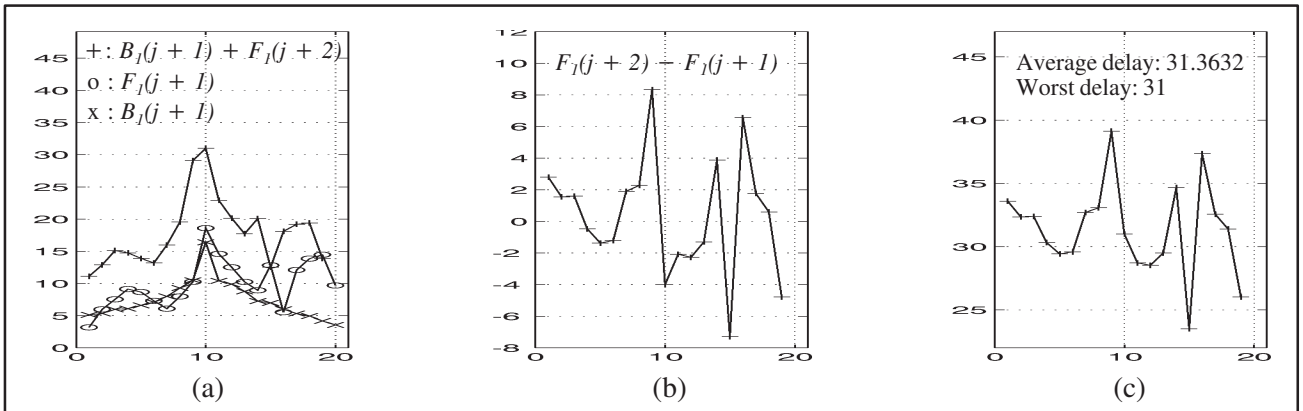


Figure 5: Simulation result with D_{in}=30.8, D_2=4 and D_{out}=15 (a)Delays of $F_1(j+1)$,$B_1(j+1)$ and $B_1(j+1)+F_1(j+2)$ (b) Delay slope (c) Output loop delay.

4.b: Output loop delay < worst stage–delay

In this section, a more interesting case which output loop delay is less than worst stage–delay for **all** the *j,* is addressed. This is exciting since this is what we have been hoping for that even a larger delay element exists in a micropipeline, the overall delay(output loop delay) is less. In other words, the micropipeline performance is not limited by the longest stage–delay anymore if delay pattern could be arbitrarily determined. This gives an added incentive to employ asynchronous approach over synchronous approach of which performance is limited by the longest stage–delay in a pipeline.

By applying the Equal loop delay theorem (Theorem 1 of section 3) to C–element C_2 and C_3 in a two–stage micropipeline and assume the maximum stage–delay for this pipeline occurs at first stage at j=k, we have

$$B_1(k+1) + D_1^{42}(k+2) + F_1(k+2) + D_2^{24}(k+2) = F_2(k+1) + D_3^{24}(k+1) + B_2(k+1) + D_2^{42}(k+2) \quad (11)$$

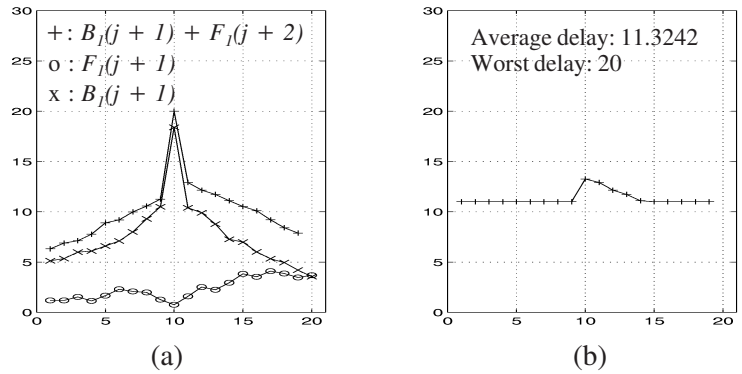
$$B_2(k+1) + D_2^{42}(k+2) + F_2(k+2) + D_3^{24}(k+2) = D_{out}(k+1) + D_3^{42}(k+2) \quad (12)$$

Eq.(12) is the expression of output loop delay. To simplify the discussion, let $D_1^{42}(k+2) = 0$, $D_2^{24}(k+2) = 0$, $D_3^{24}(k+1) > 0$ and maximum stage–delay $B_1(k+1) + F_1(k+2) = v$. From Eq. (11), this implies $F_2(k+1)+B_2(k+1)+D_2^{42}(k+2) < v$. Again, assume $D_{in}(j+2), DF_2(j+1)$ and $D_{out}(j+1)$ are constant. Then in Eq. (12), $B_2(k+1) + D_2^{42}(k+2) + F_2(k+2) < v$. Also, by previous assumption,$D_{out}(k+1)<v$,so if $D_{out}(k+1) \geq B_2(k+1) + D_2^{42}(k+2) + F_2(k+2)$ then logic delay $D_3^{42}(k+2)=0$, and output loop delay becomes $D_{out}(k+1)$, which is less than v; if $D_{out}(k+1) \leq B_2(k+1) + D_2^{42}(k+2) + F_2(k+2)$ then $D_3^{42}(k+2) = B_2(k+1) + D_2^{42}(k+2) + F_2(k+2) - D_{out}(k+1)$ and output loop delay becomes $B_2(k+1) + D_2^{42}(k+2) + F_2(k+2)$, which is less than v. From the previous discussion, we found that it is possible to have output loop delay less than the longest stage–delay in a micropipeline. Figure 6 shows the simulation result for the case in which the output loop delay is less than the maximum stage–delay for all the j with D_{in}=9, D_2=4 and D_{out}=11. Table 1 indicates that the output loop delay becomes 13.26, which is less than worst stage–delay 20, for j=9 due to $D_3^{24}(10)$=6.74>0. We can consider that only part of the longest stage–delay is passed(=13.26) to the output(or environment) and the rest of it(=6.74) is "absorbed" by C–element when passing. The amount of delay(=6.74) absorbed by C–element is not detectable by the environment, leading to a better performance(=13.26).

5. Conclusion

This paper has developed a different approach in viewing linear micropipelines by introducing delay model for C–elements. Based upon this model, the behavior of linear micropipelines can be described by logic delays and physical delays(forward and backward) according to the equal loop delay theorem.

The performance of two types of linear micropipelines — fixed–delay and variable delay — is investigated in this paper. The result shows



(a) (b)

Figure 6: Simulation result with D_{in}=9, D_2=4 and D_{out}=11
(a)Delays of $F_1(j+1)$, $B_1(j+1)$ and $B_1(j+1)+F_1(j+2)$
(b) Output loop delay .

Table 1
Several delay value for j=8,9 and 10

j	8	9	10
$B_1(j+1) + F_1(j+2)$	11.26	20	12.92
$D_1^{42}(j+2)$	0	0	0
$D_2^{24}(j+2)$	0	0	0
$D_3^{24}(j+1)$	7	6.74	0
Output Loop Delay	11	13.26	12.92

that the output loop delay for fixed–delay micropipeline is equal to the worst stage–delay in the pipeline. This is, however, not the case for the micropipelines with variable delay. Both simulation and explanation show that maximum output loop delay for this type of micropipelines can be either larger than, equal to or less than the worst stage–delay in a pipeline. Indeed, a shorter output loop delay, which is less than the worst stage–delay, is always desired to achieve higher performance. A possible guideline may be to design a negative delay slope for all the stages to reduce the magnitude of upper bound.

Acknowledgement

This research is funded by NSF under grant #MIP–9211510.

References

[1] G. M. Jacobs and R. W. Brodersen, "Self–Timed Integrated Circuits for Digital Signal Processing Applications," *VLSI Signal Processing III*, IEEE Press, 1988.

[2] C. H. Lau, D. Renshaw, J. Mavor, "Data Flow Approach to Self–Timed Logic in VLSI," *Proc. of ISCAS*, 1988, pp. 479–482.

[3] A. J. Martin, S. M. Burns, T. K. Lee, D. Borkovic and P. J. Hazewindus, "The Design of an Asynchronous Microprocessor," *Proc. of the Caltech Conf. on VLSI*, pp. 351–373, Mar. 1989.

[4] T. H. Meng, *Synchronization Design for Digital Systems,* Kluwer Academic Publishers, 1991, ch. 3, pp. 45–63.

[5] S. L. Lu, "The Design of Hardware Efficient Self–Timed Circuits," *IEEE Electronics Letters*, January 1993.

[6] C.M. Chang and S.L.Lu, "Micro Data Flow Processors", *Proceedings of IEEE Pacific Rim Conference on Comm., Computers and Signal Processing*, pp.149–153, 1993

[7] P. Kudva and V. Akella, "A Technique for Estimating Power in Asynchronous Circuits," *International Symposium on Advanced Research in Asynchronous Circuits and Systems,* Salt Lake City, Utah, USA, November, 1994.

[8] J. A. Tierno and A. J. Martin, "Low–Energy Asynchronous Memory Design,"*International Symposium on Advanced Research in Asynchronous Circuits and Systems,* Salt Lake City, Utah, USA, November, 1994.

[9] C. Farnsworth, D.A. Edwards and S.S. Sikand, "Utilising Dynamic Logic for Low Power Consumption in Asynchronous Circuits,"*International Symposium on Advanced Research in Asynchronous Circuits and Systems,* Salt Lake City, Utah, USA, November, 1994.

[10] C.M Chang and S.L. Lu, "Design of a Static MIMD Data Flow Processor Using Micropipelines," *IEEE Trans. on VLSI Systems*, Vol. 3, No. 3, September, 1995.

[11] T. E. Williams, "Performance of Iterative Computation in Self–Timed Rings," *Journal of VLSI Signal Processing*, 7, pp17–31, 1994

[12] M. R. Greenstreet and K. Steiglitz, "Bubbles Can Make Self–Timed Pipelines Fast," *Journal of VLSI Signal Processing*, 2, pp139–148, 1990

[13] S. Burns, "Performance Analysis and Optimization of Asynchronous Circuits," *Ph.D Thesis*, Computer Science Department, Caltech, 1991.

[14] I. E. Sutherland, "Micropipelines", *CACM*, Vol. 32, No. 6. June 1989.

[15] C.M. Chang and S.L. Lu, "Transient Delay and Steady State Analysis of Micropipelines," technical report, *TRECE 94.005 September 26, 1994*, Department of Electrical and Computer Engineering, Oregon State University,

[16] C.M. Chang and S.L. Lu, "Delay Analysis of Micropipelines," *International Symposium On VLSI Technology, Systems and Applications,* Taipei, Taiwan, ROC, pp.247–251, May 1995.