PLA DELAY ESTIMATION

Pravil Gupta and Alice Parker

CENG Technical Report 93-16

Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, California 90089-2562
(213)740-4476

April 1993

PLA Delay Estimation*

Pravil Gupta

Alice C. Parker

Department of Electrical Engineering-Systems

University of Southern California

Los Angeles, CA 90089-2562

April 16, 1993

Abstract

This report describes prediction of delays in controller PLAs prior to state machine synthesis. Prediction of such delays is useful during high-level synthesis, when one wants to include controller delays.

We know that PLAs have a very regular structure, so we can relate the size of the PLA to the number of inputs, number of product terms and the number of outputs; and the size can be used to predict delays in various "planes" of the PLA. Unfortunately, at this stage of synthesis we do not have any information on the actual structure of the PLA and that makes the delay computation very difficult. The model presented here is basically an empirical model derived from a series of SPICE simulations. We attempted to relate the PLA delay in some representative examples to the number of inputs, total number of product terms and the number of outputs. The model gives us the best case, the worst case and some intermediate results, all of which can be used to approximate the actual delay.

*This research described in this study was funded by the Semiconductor Research Corporation under SRC Contract Number 89-DJ-075.

1 Introduction

This task's objective is the prediction of delays in controller PLAs prior to state machine synthesis. Prediction of such delays is useful during high level synthesis, when one also wants to include controller delays.

PLAs have a very regular structure, so we can easily relate the size of the PLA to the number of inputs, number of product terms and the number of outputs. The PLA size can be used to predict delays in various levels of the PLA. However, for precise delay prediction we need to know the actual locations of transistors and the way they are connected. Unfortunately, at early stages of synthesis we do not have any such information and that makes the problem much more difficult.

A variety of design styles can be used to implement PLAs. In our case we preferred to use the pseudo-NMOS design style, which is commonly used. In this design the load is a PMOS transistor and pull down logic consists of NMOS transistors. The AND-OR logic is implemented as a NOR-NOR circuit with inverted inputs and outputs.

2 Approach to PLA Delay Estimation

The model presented here is basically an empirical model derived from a series of SPICE simulations. Our aim is to estimate the delay given the number of inputs, number of product terms and number of outputs. This requires some assumptions about the distribution of transistors if not their exact locations, since there is no way of getting location information before state machine synthesis. Our approach here is to make use of the regular structure of the PLA and estimate the delays at different points (logic levels). To do this we first estimate load capacitance associated with each NOR gate. Since the physical dimensions of the transistors are fixed by the technology, the delays through these levels mainly depend on the associated load capacitance and the number of high inputs to that particular NOR gate. We first developed a model to estimate capacitances at different points in the PLA and then related these capacitances to the delays in signal propagation. This model can be

extended to estimate the delay once the distribution of transistors in the personality matrix is known.

In our model we break a PLA into the following stages :

1. inverters at the input end,
2. the NOR gate which computes the value of the product term,
3. the NOR gate which computes the output (or function) line, and
4. the inverter at the output end.

A number of PLAs of various sizes were generated. If $a_{i,j} = 0$ then we have two inverters at the input end and if $a_{i,j} = 1$ then only one (for NOR-NOR structures we need to invert the input and the output). Using these PLAs we developed empirical models for capacitance in each level of the PLA. To estimate the delay we use a wiring delay model derived from Sakurai's model [2] for each component of the path, with appropriate parameters and transistor sizes. The wiring delay model is discussed in [1].

Total delay can be estimated as follows, depending on whether the load capacitance is being charged or discharged:

$$\begin{aligned} t_{total\_ch} &= t_{inv1\_ch} + t_{inv2\_dis} + t_{pt\_ch} + t_{op\_dis} + t_{inv3\_ch} \\ t_{total\_dis} &= t_{inv1\_dis} + t_{inv2\_ch} + t_{pt\_dis} + t_{op\_ch} + t_{inv3\_dis} \end{aligned}$$

t_{inv1_ch} and t_{inv2_dis} are the input inverter delays, t_{pt_ch} is the product-term stage delay, t_{op_dis} is the output stage delay, and t_{inv3_ch} is the final inverter delay. The two NOR stages are the most significant contributions to delay.

A number of parameters have been determined empirically. As we have seen earlier there could be one or two input inverters in series. The delay for these inverters is determined by the channel resistance of the driver transistors (determined through SPICE simulations), the capacitance due to the poly line, and the gate capacitance of the transistors connected to the product term lines.

The outputs of the inverters are inputs to the NOR gates which evaluate the product terms. The delays in this stage are determined by the channel resistance of the driver transistors, the line capacitance and the load capacitance. In case of discharge, the delay is also going to depend on the number of ON transistors (how many input lines are high). The line capacitance and the load capacitance for these NOR gates are given by empirical equations.

For the NOR gates which evaluate the outputs, the delay is determined by the line capacitance of the output line and the load capacitance of the output NOR gate, which are given by empirical equations, as well as the channel resistance of the driver transistors.

For the output inverter (inv3), the delay is almost fixed and is approximated by a constant.

2.1 Ramp inputs

Since the input to the PLA is not going to be a step input because of the finite rise or fall time from the previous stage, the effect of ramp inputs was also investigated. It was found that the delay from from 50% of the input value to 50% of the output value differs by not more than 5% compared to the delay when we have a step input. This supports our use of step inputs for analysis even when we have some distortion in the actual input.

2.2 Verification of the model

To verify the model we generated PLAs of different sizes and performed SPICE simulations on them. Table 1 gives some representative SPICE and estimated results. Despite the simplicity and the speed of the estimator, fairly good predictions of delay can be obtained.

2.3 The PLA Delay Software

The delay software does a complete analysis of various components of the total delay in the PLA. Since the capacitance of the product term depends on the number of inputs connected

to it and the capacitance of the output (or function) line depends on the number of product term lines connected to it, we consider three different cases for our load capacitance analysis:

i. Best case : We assume that only one input forms the product term in ($\tilde{\text{ii}}$) and only one product term forms the function in (iii)

ii. Average case: Half the inputs form the product term (in the delay path) in (ii) and half of the product terms form each output line in (iii).

iii. Worst case: All the inputs form the product term in (ii) and all the product terms forms the output line in (iii).

For the ON transistor analysis, we assume the following subcases:

i. Best case: all the inputs to the NOR gate are high.

ii. Average case : half of the inputs connected to the NOR gate are high.

iii. Worst case : only one input is high.

We consider these 5 cases for our combined delay analysis:

i. Best-Best: best case with respect to both 1 and 2. (load capacitance and ON transistor analysis).

ii. Average-Average: average with respect to both 1 and 2.

iii. Average-Worst : average with respect to 1 and worst with respect to 2.

iv. Worst-Average : worst with respect to 1 and average with respect to 2.

v. Worst-Worst : worst case with respect to both 1 and 2.

Table 2 gives charging and discharging delays for three different size PLAs for all the above-mentioned 5 cases.

No.	No. of Inputs	No. of Outputs	No. of Product Terms		Delay	
					Estimated ns	SPICE ns
1	4	10	16	charging	19.25	20.75
2				discharging	18.31	19.50
3	5	20	30	charging	37.19	36.75
4				discharging	34.07	30.50
5	5	1	25	charging	19.75	20.75
6				discharging	21.09	24.75
7	5	20	20	charging	34.24	33.50
8				discharging	28.81	24.50

Table 1: Estimated results compared with SPICE.

No.			Delay	
			Charging(ns)	Discharging(ns)
1	i=5	Best-Best	9.10	7.94
	o=5	Average-Average	13.00	11.89
	pt=10	Average-Worst	13.24	11.89
		Worst-Average	19.93	17.78
		Worst-Worst	20.92	17.96
2	i=8	Best-Best	8.86	7.91
	o=15	Average-Average	20.91	19.04
	pt=20	Average-Worst	22.29	19.29
		Worst-Average	33.07	30.33
		Worst-Worst	36.40	31.14
3	i=15	Best-Best	8.30	7.82
	o=20	Average-Average	27.85	24.32
	pt=30	Average-Worst	29.97	24.80
		Worst-Average	47.48	41.13
		Worst-Worst	52.09	42.36

Table 2: Delay Analysis of PLAs

3 Conclusions

It is difficult to model PLA delays prior to the actual synthesis. In this study, we derived an empirical model which is heavily dependent on the technology used. However, it can be parameterized for a given technology once the corresponding constants are determined. The model gives us the best case, the worst case and some intermediate results, all of which can be used to approximate the actual delay.

References

[1] P. Gupta and A. C. Parker. Wire Delay Estimation. Technical report, University of Southern California, 1990.

[2] T. Sakurai. Approximation of Wiring Delay in MOSFET LSI. *IEEE J. Solid-State Circuits*, SC-18(4):418–426, August 1983.