Optimization of Layered Self-Timed Interfaces

Ghassan Y. Yacoub, Tarun Soni and Walter H. Ku
Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093-0407

Abstract *– This paper presents an approach for designing digital CMOS layered self-timed interfaces. Each layered interface consists of a region containing* m *optimizable transmitter-receiver pairs such that transmission channel* i *is characterized by the 3-vector attribute* $x_i = \{R_o, C_g, l\}$. *A layer is sized by minimizing a hybrid cost function subject to one or more constraint functions selected from a menu. This approach specifically treats 2-Cycle bundled data protocols but can be generalized to other asynchronous signalling schemes as well as synchronous data transmission. The design paradigm is illustrated by solving a nonlinear constrained optimization problem which adequately accounts for the effect of wire inductance on signal delay. Therefore, the proposed unified approach utilizes numerical optimization of communication channels as a precursor to time consuming circuit simulation, thus achieving an appreciable speed up in the overall design task.*

1. Introduction

One of the difficult problems facing self-timed circuit design is preserving the bundling constraint between data wires and their corresponding request wire [1] [2]. The data bus and the associated request signal form a bundle such that their sequence of events are preserved in-spite of communication delays [3] [4]. The transitions on the data wires observed by the receiver must precede the corresponding transition on the request wire to maintain integrity of data transfer.

Figure 1 depicts that the bundled data events, indicated as, ①, precede the corresponding request event, ②. To enforce this bundling constraint, the current inclination of circuit designers is to make the circuit adapt (slow down) by adding buffer or register delays into the request signal's path [2] [3]. This approach is justified for circuits where fast design turnaround is of prime interest and sufficient modeling information of the transmission line effects not available. However, for future high performance sub-micron VLSI circuits with sub-nanosecond switching times, the interface between transmitter and receiver will require more accurate modeling and analysis techniques. In the general case where clustering self-timed modules is not feasible, it becomes necessary to pay special attention to the bundling constraint. This paper addresses the bundling constraint in several steps. Section 2 describes a unified design approach where we propose an optimization taxonomy. Section 3 applies this approach to wide bus 2-Cycle self-timed interfaces. Section 4 discusses our results and future work.

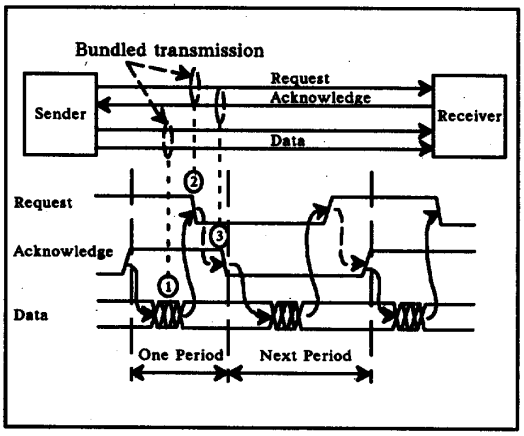

Figure 1: Sequence of events for self-timed 2-Cycle bundled data signalling. The Request and Data events are expected to maintain their sequence of occurrence as they travel from transmitter to receiver regardless of physical placement.

2. A Unified Design Approach

As custom and semicustom CMOS chips are miniaturized to attain faster switching times, issues that were important in high-end designs gain importance in all digital systems, both high and low end. The advent of large-area superchips and multichip package substrates with dense thin film wires are proving that chips are becoming like boards and boards like chips [5]. This implies that on-chip and chip-to-chip signal distribution will need a unified design approach whose underlying models account for RC delay as well as transmission line effects. To this effect we have developed a unified design approach for layered self-timed interfaces based on the telegraph equation with underlying models representing the RC delay domain as well as the transmission line delay domain.

1058-6393/92 $03.00 © 1992 IEEE

2.1 Definition of Layered Self-Timed Interfaces

A simplified view of a layered interface is defined in Figure 2a. The underlying concept implies that the regions containing the interconnect wires can be optimized during the design phase. These layers can be sized as a function of the driver active impedances relative to the interconnect passive impedances. Figure 2b depicts the correspondence between electrical and geometrical parameters.

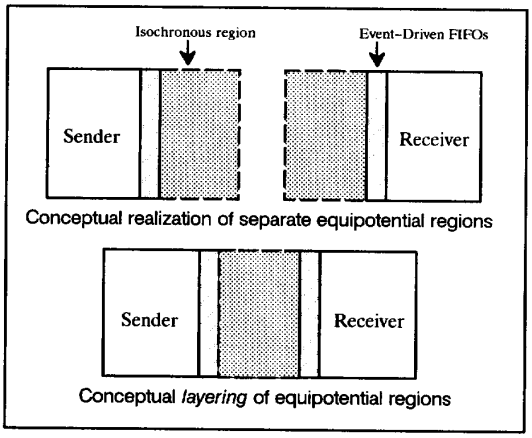

Figure 2a: Conceptual diagram which depicts layered interfaces between processing elements.

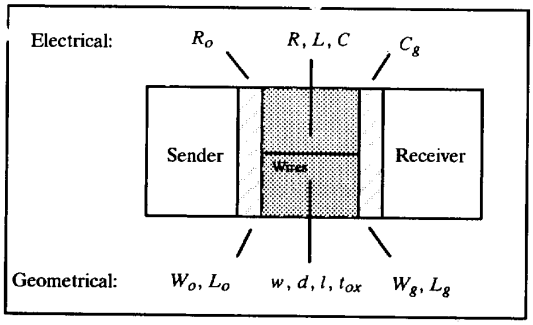

Figure 2b: Electrical and geometrical parameter correspondences for the layered elements.

Figure 3 depicts a more detailed view of the layered interface environment along with the designable parameters R_o, C_g, and l for each data channel. The event-driven FIFOs and the C elements are similar to those described in [2]

2.2 Analytical Approach

The design approach consists of modeling the interconnect wires as uniform RLC sections and solving the telegraph equation to obtain the following empirical solution in two distinct time domains [6]

RC Delay Domain ($\tau_{rc} \geq \pi\zeta\gamma\tau_p$)

$$\tau_d \approx \frac{9.2\tau_{rc}}{\pi^2\gamma^2} = \frac{9.2}{\pi^2\gamma^2}\left(2R_oCl + \zeta^2RCl^2\right) \quad (1)$$

Transmission-Line Delay Domain ($\tau_{rc} < \pi\zeta\gamma\tau_p$)

$$\tau_d \approx \frac{4.6\zeta^2\tau_p^2}{\tau_{rc}} = \frac{4.6\left(\gamma^2 + \frac{2C_g}{Cl}\right)LCl^2}{2R_oCl + \zeta^2RCl^2} \quad (2)$$

τ_p is the time-of-flight of the signal from transmitter to receiver if it were to travel distance l at the speed of light in the medium.

According to [6], $|\gamma| \leq 1.1$ for the following parameter ranges; wire length l in the range [500μm, 50,000μm], wire width W in the range [0.1μm, 10μm], driver's output resistance R_o in the range [200Ω, 4KΩ], and wire resistivity ρ in the range [2x10$^{-8}\Omega$-m, 5x10$^{-5}\Omega$-m]. For the above parameter ranges, it has been verified in [6] that the approximation of τ_d can be used to predict delays within an error of 15% compared to the numerical results based on the analytical solutions of the telegraph equation.

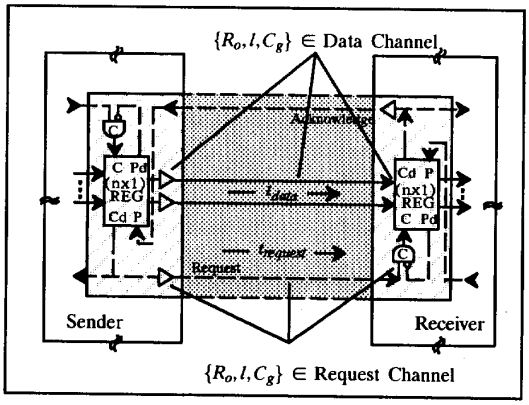

Figure 3: An illustration of a layered 2-Cycle self-timed interface. This interface is representative of simple adjacent sender-receiver blocks. The constraint $t_{request} - t_{data} \geq t_{setup}$ must be satisfied at all times for safe data transfer.

2.3 Circuit Model and Electrical Performance of Interconnection

Figure 4 illustrates a microstrip cross section along with the circuit and interconnection models used in our analysis.

The driver's output impedance R_o specified at the maximum drain conductance point [8] is given by

$$R_o = (W_o\mu C_{ox}(V_{DD} - V_{th}))^{-1}L_o \quad (3)$$

where W_o, μ, C_{ox}, and L_o are the driver's effective channel width, carrier mobility, gate oxide capacitance, and effective channel length respectively.

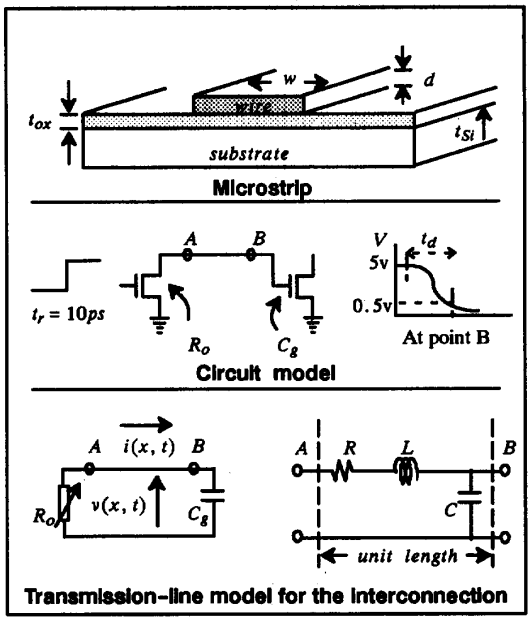

Figure 4: A microstrip on a silicon substrate along with the circuit and interconnection models used during analysis. The wire impedances, *R*, *L*, and *C* are given in units per micron.

Furthermore, C_g is related to the receiver's geometrical dimensions by

$$C_g = C_{ox} W_g L_g \tag{4}$$

where C_g, W_g, and L_g are the receiver's input gate capacitance, channel width, and channel length respectively.

For a microstrip on a semiconductor such as silicon, the relationships we used (assuming that the skin effect is negligible) are [5] [6]

$$\text{For } h = t_{ox} + t_{Si}$$

$$C = \frac{2\pi\epsilon_{eff}\epsilon_o}{\ln\left(\frac{8t_{ox}}{W} + \frac{W}{4h}\right)} \tag{5}$$

$$L = \frac{1}{Cv_p^2} \quad ; \quad R = \frac{\rho}{Wd} \quad ; \quad Z_o = \sqrt{L/C} \quad ;$$

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2}\left(1 + \frac{10h}{W}\right)^{-1/2}$$

where ϵ_{ox} is the dielectric constant of silicon dioxide, v_p is the speed of light in silicon dioxide, ρ is the resistivity of the wire material, and Z_o is the wire's characteristic impedance.

We modeled R_o as a function of W_o in equation (3) for values of W_o up to 200um using the MOSIS 2 um CMOS process. Then we simulated the device behavior in response to a fast input pulse as illustrated in Figure 4, where the wire is modeled as a distributed *RLC* transmission line. In the simulation, the wire length varied between 100um and 5cm while the values of R_o and C_g were fixed at 200Ω and 100ff respectively. The results are plotted in Figure 5 with the empirical relationships listed in the previous section from [6]

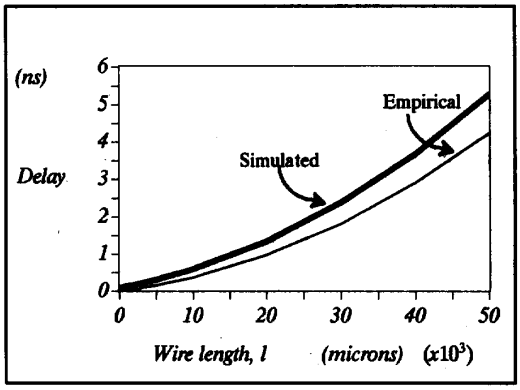

Figure 5: Comparison of the empirical result found in [6] with a SPICE simulation of the circuit model shown in Figure 4 for R_o = 200 Ω and C_g = 100 ff.

As can be seen from Figure 5, the RC delay domain is dominant for most of the wire lengths. However, for the particular microstrip we considered ($W = d = t_{ox} = 1$um, $t_{Si} = 200$um) and the above driver and receiver impedance values, one observes a transition from TL to RC domains at a wire length of 2875um. This transition between domains will be further analyzed in the next section.

2.4 Optimization Taxonomy

We propose an optimization taxonomy for layered self-timed interfaces. The taxonomy stems from the classical matching condition between the impedances of the transmitter, channel, and receiver. This is specified as

$$\pi\gamma\zeta Z_o = 2R_o + \zeta^2 Rl \tag{6}$$

This proposed taxonomy is depicted in Figure 6 for the condition where C_g = 100 ff. The taxonomy encourages layer size (i.e., *l*) selections that satisfy the RC delay domain. Of prime interest is finding appropriate values of *l* which facilitate the simplest lumped interconnection model selection for ease of delay analysis.The optimization taxonomy differentiates between

RC and TL delay domains by guiding the designer towards the RC delay domain and finding an optimum $x_i = \{R_o, C_g, l\}$.

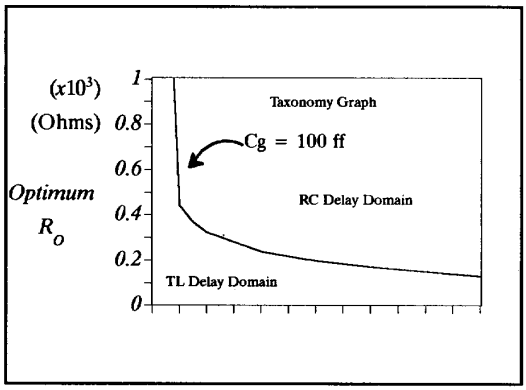

Figure 6: The proposed optimization taxonomy for optimizing layered self-timed interfaces. This taxonomy encourages RC analysis whenever possible because it is much simpler and less time consuming than that of transmission lines.

This intermediate numerical optimization step also represents a formal technique for optimal interconnect model selections from those accurately tabulated in [7] [8] [9]. This approach can be typically applied after a thorough architectural design study [10] and before launching the circuit design phase.

3. Numerical Optimization

A hybrid cost function, F(x), was chosen to illustrate the unified design approach. The constraint functions, $c_j(x)$, were selected from a menu listed in Table I. $c_j(x)$ are circuit-specific design constraints dependent on the real n-vector designable parameters, $x = \{x_1, x_2, ..., x_n\}$. We can therefore, compactly express our nonlinear constrained design optimization problem as

$$\begin{aligned} &\text{minimize} \quad F(x) \\ &x \in \Re^n \\ &\text{subject to} \quad c_j(x) = 0, \quad j = 1, 2, \ldots, m'; \\ &\qquad\qquad\quad c_j(x) \geq 0, \quad j = m'+1, 2, \ldots, m; \end{aligned}$$

In particular,

$$F(x) = \tau_d + M(\pi\gamma\zeta\tau_p - \tau_{rc}) \qquad (7)$$

The first component, τ_d, in the above cost function is the delay in the bundle. The second component, $(\pi\gamma\zeta\tau_p - \tau_{rc})$, refers to the distance of the chosen interconnect parameters from the RC delay domain which was defined in section 2.2. The parameter, M, defines the relative importance the designer places between circuit speed and the attractiveness to staying in the RC delay domain. The τ_d component may cause a parameter choice which requires transmission line analysis, thus slowing down design time. On the other hand, the latter $(\pi\gamma\zeta\tau_p - \tau_{rc})$ component will tend to keep all parameters in the RC delay region, thus slowing down the circuit speed.

With this philosophy in mind, optimization was done with a number of values of M. To keep the optimization conditions identical to the spice simulations, R_o and C_g were kept constant and optimization was done over the single parameter, l, effectively converting this into a one dimensional optimization problem. For these conditions, with the constraints shown in Table 1, the values of "optimal" wire length obtained are shown in Figure 7.

Table I

Constraint function menu for the optimization of self-timed layered interfaces	
$0.2 \leq \frac{C_g}{Cl} \leq 2$	(receiver-wire constraint)
$0.2 \leq \frac{R_o}{Rl} \leq 2$	(transmitter-wire constraint)
$0.2 \leq \frac{Rl}{Z_o} \leq 2$	(wire-Zo constraint)
$0.2 \leq \frac{R_o}{Z_o} \leq 2$	(transmitter-Zo constraint)
$\frac{(C_g + Cl)V_{DD}^2}{\tau_d} \leq 0.25$	(dynamic power constraint)

As can be seen, when the importance of "being in the RC delay domain" is increased (by increasing M in the cost function), the wire length returned as the optimal for that cost function is seen to increase. This happens for increasing M, till the value of l is constrained by one of the other bounds in Table 1. Conversely, when M is reduced this corresponds to more weightage being given by the designer to minimizing the delay, than to modeling the interface in the RC domain. In such conditions, the "optimal" wire length returned is lower and keeps reducing with reducing M until it is constrained by one of the other bound of Table 1.

4. Conclusions and Discussion

It is apparent that for more than one parameter being under the control of the designer, the search for an optimal set of circuit and layout values is a very complex procedure. Optimization techniques aided by good models, will greatly speed up the design time and reduce the number of circuit simulations a designer has to perform before reaching the optimum design.

The long term intentions of our research is to investigate optimization strategies for layered (i.e., overlapping equipotential regions) self-timed interfaces within IC systems. The findings of this paper complements the research reported in [10] by using analytical tools

whose underlying physical models adequately represent interconnection signal propagation delay.

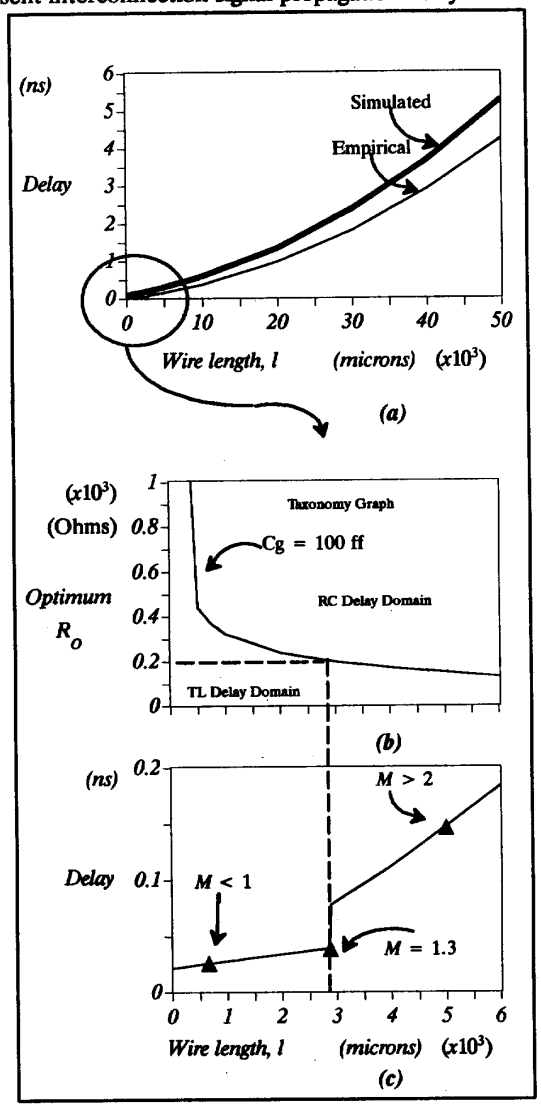

Figure 7: .An illustration of the RC and TL delay domains in the design. (a) shows the behaviour of the delay as a function of the wire length. (b) shows the two domains, RC and TL for the smaller range of wire length values. (c) shows the optimal wire length returned by the optimization module for various values of M. As the value of M is increased the optimized wire length is seen to increase and the design goes from the TL domain to the RC domain.

The optimization taxonomy we are proposing can be useful for VLSI performance driven place and routing. This will allow for rapid prototyping of IC systems by utilizing the less time consuming RC domain delay estimation.

5. References

[1] G. M. Jacobs, "Self-Timed Integrated Circuits for Digital Signal Processing," *Ph. D. Thesis, University of California, Berkeley*, UCB/ERL Memorandum No. M89/128, November 30, 1989.

[2] I. E. Sutherland, "Micropipelines", *Communications of the ACM*, Vol. 32, No. 6, pp. 720–738, June 1989.

[3] I. E. Sutherland, R. F. Sproull and I. Jones, Standard Asynchronous Modules, *Technical Memo 4662*, Sutherland, Sproull and Associates, 1986.

[4] T. H.-Y. Meng, R. W. Brodersen and D. G. Messerschmitt, "A Clock-free Chip Set for High-Sampling Rate Adaptive Filters," *Journal of VLSI Signal Processing, Vol. 1, pp. 345–365, 1990.*

[5] H. B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley, Reading, Massachusetts, 1990.

[6] D. Zhou, F. P. Preparata and S. M. Kang, "Interconnection Delay in Very High-Speed VLSI," *IEEE Transactions on Circuits and Systems, Vol. 38*, No. 7, pp. 779–790. July 1991.

[7] T. Sakurai, "Approximation of Wiring Delay in MOSFET IC LSI," *IEEE Journal of Solid-State Circuits, Vol. SC-18*, pp. 418–426. August 1983.

[8] G. Y. Yacoub and E. G. Friedman, "An Environment Sensitive Circuit Design Technique for Modeling VLSI Interconnect Impedances", *Proceedings of 1987 Government Microcircuit Applications Conference*, pp. 391-394, October 1987.

[9] S. Long, S. Butner, *Gallium Arsenide Digital Integrated Circuit Design*, McGraw-Hill, New York, 1990.

[10] G. Y. Yacoub and W. H. Ku, "Self-Timed Simultaneous Bidirectional Signalling for IC Systems," *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 2957–2960, May 1992.

[11] T. Sakurai, "Simple Formulas for two- and Three-dimensional Capacitances," *IEEE Transactions on Electron Devices*, Vol. ED-30, pp. 183–185, February 1983.

[12] P. E. Gill, W. Murray, M. A. Saunders and M. A. Wright, User's guide for NPSOL (Version 4.0), Report SOL 86-2, Department of Operations Research, Stanford University California, January 1986.