

Optimal Repeater Insertion for N-Tier Multilevel Interconnect Architectures

Raguraman Venkatesan, Jeffrey A. Davis, Keith A. Bowman and James D. Meindl
Georgia Institute of Technology, Atlanta, USA

Abstract— A new methodology to optimally insert repeaters for n-tier multilevel interconnect architectures is demonstrated. For a 0.1 μ m ASIC macrocell case study, repeater insertion either decreases the macrocell area 4-fold (and lowers power dissipation by 50%), increases clock frequency by 22% or reduces number of metal levels by 25%.

I. INTRODUCTION

The increasingly restrictive limits posed by interconnects on processor performance make their optimal design a primal concern for future technology generations. Repeaters improve the dependency of time delay on interconnect length from a square-law to a linear relationship [1]. A novel optimal repeater insertion methodology is presented here to either minimize the macrocell size (and power dissipation) or number of metal levels or maximize the clock frequency for an n-tier multilevel interconnect architecture. An ASIC macrocell with $N = 12.4M$ logic gates, copper interconnects, a low-k dielectric ($\epsilon_r = 2.0$) and 8 metal levels is optimized for the 0.1 μ m technology International Technology Roadmap for Semiconductors (ITRS) [2] projections.

II. N-TIER MULTILEVEL ARCHITECTURE

The n-tier multilevel architecture is designed by scaling the pitch on each tier so that the longest interconnect on a tier meets the timing delay constraint exactly [3]. A stochastic wiring distribution, validated previously with measured data [4], is used to estimate the longest interconnect on each tier. For the n^{th} tier, the range of interconnect lengths is calculated by equating the area available for wiring (A_{av}) to the area that is required for wiring (A_{req}),

$$A_{av} = n_l e_w A_m = \chi p_n \sqrt{\frac{A_m}{N}} \int_{L_{n-1}}^{L_n} li(l) dl = A_{req} \quad (1)$$

where n_l is the number of metal levels in the n^{th} tier, e_w ($=0.4$) is the wiring efficiency factor [5], A_m is the macrocell area, χ ($=0.667$) is a factor that converts the point-to-point interconnect length to wiring net length, p_n and L_n are the wire pitch and longest interconnect length of the n^{th} tier in microns and gate pitches, respectively and $i(l)$ is the interconnect density function [4].

The wire pitch in (1) (without repeaters) for all non-local tiers (i.e. $p_n > 2F$) is calculated using RC models for time delay as

$$p_n = 2 \sqrt{\frac{1.1 \rho \epsilon_r \epsilon_o 6.2 f_c}{\beta}} \sqrt{\frac{A_m}{N}} L_n \quad (2)$$

where ρ is resistivity of metal, f_c is the clock frequency and β is the interconnect time delay expressed as a fraction of the clock period [5]. The maximum time delay is assumed to be 25% of the clock period ($\beta = 0.25$) on the first tier and 90% ($\beta = 0.9$) on all other tiers.

III. REPEATER INSERTION MODELS AND METHODOLOGY

Bakoglu [6] derives an expression for the pitch of an interconnect when an “optimal” number of equi-spaced repeaters are inserted so that the delay of each repeater equals the delay of the interconnect segment between repeaters. When the number of repeaters in an interconnect is less than this optimal number, it is considered a “sub-optimal” design. If the number of repeaters is some factor ζ times the “optimal” number of repeaters, $0 < \zeta \leq 1$, then the pitch is given by

$$p_n = \left(1.4 + 0.53\zeta + \frac{0.53}{\zeta}\right) \frac{2f_c}{\beta} \sqrt{6.2\rho\epsilon_r\epsilon_o R_o C_o} \sqrt{\frac{A_m}{N}} L_n \quad (3)$$

where R_o and C_o are the output resistance and input capacitance of a minimum feature size inverter respectively.

Since the area of the chip is wire-limited, there is some unutilized silicon real estate which is used to accommodate the repeaters. It is assumed that only 60% of this free area is used for repeater insertion to account for practical routing and placement constraints.

The area of a gate, A_g is calculated as [7]

$$A_g = k_I \left(1 + \frac{4\sqrt{G_{ar}}(f_i - 1)}{\sqrt{k_I}}\right) \left(1 + \frac{(1 + \beta_g)(w_k - 1)}{\sqrt{k_I G_{ar}}}\right) \quad (4)$$

where k_I is the area of a minimum sized inverter, G_{ar} is the gate aspect ratio, f_i is the number of inputs, β_g is the ratio of PF T to NF T width and w_k is the NF T width to feature size ratio. The PF T width is constrained to satisfy equal worst case rise and fall times and w_k is calculated by equating the critical path delay to the clock period ($=\frac{1}{f_c}$),

$$\frac{1}{f_c} = \frac{n_{cp} T_{PDn} f_{ineff}}{b} \quad (5)$$

where T_{PDn} is the NF T propagation delay including the transition (rise or fall) time effect that is derived from the physical alpha-power law model [8], f_{ineff} is the effective fan-in factor for series connected MOSF Ts [9], n_{cp} is number of gates in a critical path ($=10$) and b is the clock skew factor ($=0.9$). The supply voltage is chosen as 1V which is in the range of the ITRS projections [2].

Repeater insertion begins from the topmost tier and continues downward to lower tiers in a top-down design style. On each tier either the “optimal” number ($\zeta = 1$) or a “sub-optimal” number ($\zeta < 1$) of repeaters are inserted, depending on the number of repeaters available.

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, 791 Atlantic Drive NW, Atlanta, GA 30332-0269. email: vrugu@ee.gatech.edu
The support of the Semiconductor Research Corporation (SJ-374.002) and the Defense Advanced Research and Projects Agency (BAA 9415-A-007) is gratefully acknowledged.

IV. OPTIMIZATION RESULTS

Four repeater insertion optimizations are defined in this section that either minimize the macrocell area (and the power dissipation) or the number of metal levels or maximize the clock frequency. These optimizations are compared against the same designs without repeaters. Table I provides all the interconnect parameters for the various design points.

A. Minimum macrocell area

The area of the macrocell with 8 metal levels and $f_c = 2\text{GHz}$ is minimized using repeaters. Fig. 1 plots the macrocell area versus the number of repeaters for this design. Repeaters make interconnects narrower and this decreases the area needed for wiring. The reduction in average wiring capacitance, due to smaller wire limited macrocell area, reduces the logic gate size and decreases the transistor limited area. As number of repeaters increases, the wire limited macrocell area decreases until it equals the transistor limited area. If more repeaters are inserted, then the macrocell area becomes transistor limited and starts increasing (due to increases in the number of repeaters and the size of logic gates). Therefore, for a given number of metal levels, the macrocell area is minimized when the wire and transistor limited areas become equal. As seen from Fig. 1, optimal repeater insertion decreases the macrocell area from 1.79cm^2 to 0.43cm^2 , more than a 4-fold reduction in the cell size.

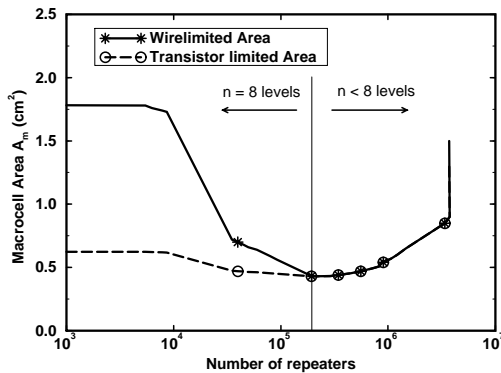


Fig. 1. Minimum macrocell area optimization for baseline design of $f_c=2\text{GHz}$ and at most $n=8$ metal levels

B. Maximum clock frequency

The clock frequency of the macrocell with $A_m = 1.79\text{cm}^2$ and 8 metal levels is maximized using repeaters. Fig. 2 plots the clock frequency versus number of repeaters for this design. The wire limited clock frequency is the maximum frequency for which all the transistors can be wired within 8 metal levels for the specified macrocell area. The wire limited clock frequency increases with increase in number of repeaters because repeaters speed up lines for a constant pitch. The transistor limited clock frequency is the maximum frequency at which the logic critical path transistors can operate if they are enlarged to occupy the remaining macrocell area (after accounting for repeaters). The transistor limited clock frequency decreases with an

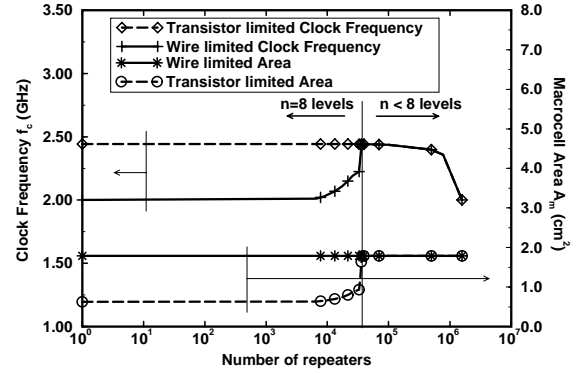


Fig. 2. Maximum clock frequency optimization for baseline design of $A_m = 1.79\text{cm}^2$, $n=8$ metal levels

increase in the number of repeaters due to a decrease in the area available for logic gates. For a small number of repeaters (wire limited macrocell area), the clock frequency is also wire limited. As the number of repeaters increases, the wire and transistor limited frequencies eventually become equal when the macrocell area becomes transistor limited. More repeaters can be added only by shrinking the logic transistors which reduces the transistor limited frequency. Hence, the clock frequency peaks at the point where the wire and transistor limited frequency curves meet. Therefore, repeater insertion increases the maximum clock frequency from 2GHz to 2.44GHz , which is a 22% improvement.

C. Minimum number of metal levels

The number of metal levels required for the macrocell with $A_m = 1.79\text{cm}^2$ and $f_c = 2\text{GHz}$ is minimized using repeaters. Fig. 3 plots the number of metal levels versus the number of repeaters for this design. As the number of repeaters increases, the number of levels decreases while the transistor limited area increases. Once the wire limited and transistor limited areas become equal, no more repeaters can be inserted without increasing the macrocell area. However, in Fig 3, all the tiers become satiated with repeaters (either they have optimum number of repeaters or have reached minimum width) and hence further repeater

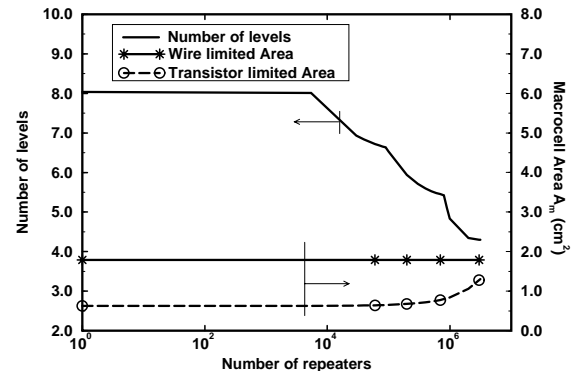


Fig. 3. Minimum number of levels optimization for baseline design of $f_c=2\text{GHz}$, $A_m = 1.79\text{cm}^2$

Tier # (n)	number of levels	L_n		p_n (μm)	number of repeaters	NFET W/L ratios	Dynamic Power
		gate pitches	(cms)				
without repeaters : $A_m=1.79cm^2$, $f_c=2GHz$ and $n=8$ metal levels.							
Tier 4	2	7042.7	2.6758	1.79	0	$W/L_{logic}=22$	$P_{total}=47.93W$
Tier 3	2	1845.2	0.7011	0.94	0		$P_{logic}=71\%$
Tier 2	2	880.9	0.3347	0.45	0		$P_{int}=29\%$
Tier 1	2	208.2	0.0791	0.2	0		$P_{rep}=0\%$
minimum area and power: $A_m=0.43cm^2$, $f_c=2GHz$ and $n=8$ metal levels.							
Tier 3	2	7042.7	1.3115	0.29	1.94E5	$W/L_{logic}=11$	$P_{total}=24.34W$
Tier 2	2	865.1	0.1611	0.22	0	$W/L_{rep}=57$	$P_{logic}=68\%$
Tier 1	4	198.5	0.0370	0.2	0		$P_{int}=28\%$, $P_{rep}=4\%$