

# Optimal Repeaters for Sub-50nm Interconnect Networks

Deepak C Sekar, Raguraman Venkatesan\*, Keith A Bowman\*, Ajay Joshi, Jeffrey A Davis and James D Meindl  
Georgia Institute of Technology, \*Intel

## Abstract

Power consumed by interconnect repeaters is a serious concern for future ICs. Ways to tackle this issue such as unique optimization of repeater and logic transistor technologies, improved repeater insertion methods and 3D integration are discussed. These techniques reduce total power of a 22 nm 1.4 GHz low power combinational logic block by 55% with negligible performance and area overheads.

## Introduction

About twenty years ago, Bakoglu and Meindl discussed the introduction of interconnect repeaters to speed up across-chip wires [1]. Recent studies indicate that repeater count increases exponentially with scaling [2][3], and 70% of the cells in a microprocessor's logic block at the 32 nm node could be repeaters, as shown in Fig. 1 [2]. This raises concerns about repeater power dissipation.

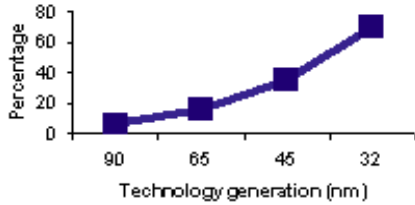


Figure 1: Number of repeaters as a percentage of total cell count of a high performance microprocessor's logic block [2]

A simulation tool called MINDS is used to find the trends for repeater power dissipation with scaling. The n-tier methodology used in MINDS is described thoroughly in [3] and [4]. To summarize, MINDS arranges wires in metal levels based on a stochastic wiring distribution and available wire area. The pitch of every orthogonal pair of metal levels is calculated by equating a specified fraction of the clock period to the delay of the longest wire in that pair of metal levels. Logic gates are modeled as two-input NAND gates and are sized based on average wire length estimates. Simulations using MINDS have been shown to match data from industrial designs in previous work [3]. Leakage power models from [5] are used.

Fig. 2 shows results from MINDS that indicate that while repeaters take up 12% of a low-power combinational logic block's power at 65 nm, they could consume a staggering 53% of the power at 22 nm. Repeater leakage power, in particular, can be seen from Fig. 2 to be a serious concern. Leakage power of logic gates is not so alarming, since logic gates scale better with technology than repeaters. Note that the data in Fig. 2 is generated considering: (1) Low Operating Power (LOP) ITRS transistor parameters (2) Sub-optimal repeater insertion [4] with 10% delay penalty. Also, Rent's constants  $k$  and  $p$  are 4 and 0.6 respectively.

In this paper, we extend previous work and discuss techniques that reduce repeater power of a future 1.4 GHz 22 nm low-power combinational logic block by 80% with negligible performance and area overheads.

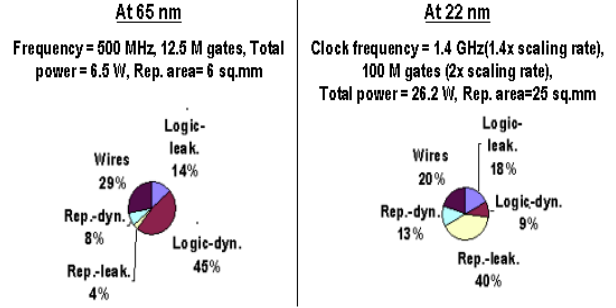


Figure 2: Scaling trend for repeater power in a 60 sq.mm low power combinational logic block

## Derivation of Interconnect Repeater Insertion Model

In this scenario where repeater power is a significant fraction of system power, a compact model that minimizes Energy-Delay Product (EDP) of a repeated wire is important. Fig. 3 shows the derivation of such a model.

- (1)  $Delay = k[0.7 \frac{R_o}{h} (\frac{C_{int}}{k} + hC_o) + \frac{R_{int}}{k} (0.4 \frac{C_{int}}{k} + 0.7hC_o)]$
- (2)  $Power = (a \frac{1}{2} C_o V_{dd}^2 f + b V_{dd} I_{leak}) h k + a \frac{1}{2} C_{int} V_{dd}^2 f$
- (3)  $Energy-delay product (EDP) = Delay^2 \cdot Power$   
Set  $\frac{d(EDP)}{dh} = 0$ ,  $\frac{d(EDP)}{dk} = 0$ , simplify & approximate to get,

$$(4) \text{ Optimal } k = (0.73 + 0.07 \ln \phi_{gate})^2 \sqrt{\frac{R_{int} C_{int}}{R_o C_o}}$$

$$\text{Optimal } h = (0.88 + 0.07 \ln \phi_{gate})^2 \sqrt{\frac{C_{int} R_o}{R_{int} C_o}}$$

$$\text{where } \phi_{gate} = \frac{\frac{1}{2} a C_o V_{dd}^2 f}{\frac{1}{2} a C_o V_{dd}^2 f + b V_{dd} I_{leak}}$$

$k$  = Number of repeaters,  $h$  = Size of repeaters,  $R_{int}$  = Wire resistance,  $C_{int}$  = Wire capacitance,  $b$  = Percentage of time circuit is not sleep gated,  $R_o, C_o$  &  $I_{leak}$  = Resistance, capacitance & leakage of minimum sized repeater respectively,  $V_{dd}$  = Supply voltage,  $a$  = Activity,  $f$  = Frequency

Figure 3: Derivation of new repeater insertion model

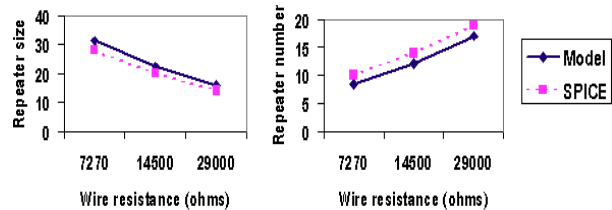


Figure 4: Validation of new repeater insertion model for a 1mm length wire with 100nm BSIM transistor models [6]

This model has <15% difference compared with SPICE simulations and its validation is shown in Fig. 4. The differences of this model from previous work are explained later in this paper. Note that all the data in this paper are generated using LOP transistor parameters and 1mm intermediate length wires for a 22 nm ITRS technology unless specified otherwise.

## Techniques to Reduce Interconnect Repeater Power in Future Technology Generations

The following technology directions are suggested for future low power microchips so that this repeater power problem can be substantially minimized.

### (A) Unique optimization of logic and repeater transistor technologies

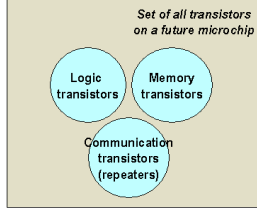


Figure 5: Three main types of transistors in future microchips

Fig. 5 illustrates that transistor area of future microchips would consist *mainly* of logic transistors, memory transistors and communication transistors (repeaters). Logic and memory transistors perform inherently different functions; they used to have the same device parameters, but this is not the case anymore. Similarly, we propose that logic and communication transistors also need to have their own uniquely optimized device parameters in the future. For example, communication transistors could have different values of threshold voltage ( $V_t$ ) from logic transistors.

$$EDP = R_{int} C_{int}^2 [R_o (a \frac{1}{2} C_o V_{dd}^2 f + b V_{dd} I_{leak}) (\frac{0.7}{\delta} + 0.7\gamma + \frac{0.4}{\gamma} + 0.7\delta)^2 (\gamma\delta + \phi_{gate})]$$

$$\text{where } \gamma = (0.73 + 0.07 \ln \phi_{gate})^2 \text{ and}$$

$$\delta = (0.88 + 0.07 \ln \phi_{gate})^2$$

Figure 6: Expression for minimum EDP of an interconnect repeater chain

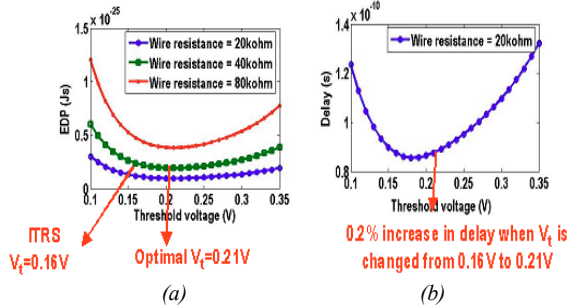


Figure 7: (a) An optimal  $V_t$  exists that minimizes the EDP of a repeater chain (b) Delay of a repeater chain using the new model

An expression for minimum Energy-Delay Product (EDP) of a repeater chain is obtained in Fig. 6 by using equations (3) and (4) of Fig. 3. The repeated wire EDP vs.  $V_t$  plot of Fig. 7(a) indicates that an optimal  $V_t=0.21V$  exists for all repeated wires on a microchip that minimizes their EDP. This is because the *wire-independent* term inside square brackets in Fig. 6 must be minimized to minimize EDP.

Fig. 7(b) shows that delay of a repeated wire with the new model is fairly insensitive to increase in  $V_t$  near the optimal point. Any delay increases can be compensated using increased wire sizes. Thus, repeated wires could give the same performance using the new model with a  $V_t=0.21V$  as they give with a ITRS specified  $V_t=0.16V$  by increasing wire sizes.

The delay of the generic logic path shown in Fig. 8 is more sensitive to  $V_t$  (Fig. 9(a)). This is because output resistance of the inverter, which depends on  $V_t$ , is much larger than wire resistance. The generic logic path cannot have its gates sized bigger to maintain performance with higher  $V_t$  values, because of large die area increases and the need to space a new  $V_t$  value far enough from the existing  $V_t$  of 0.16V for manufacturability, as Fig. 9(b) shows.

$$\text{TOTAL WIRE } C_{int} = 1.06fF$$

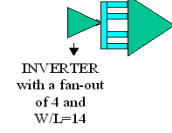


Figure 8: A generic logic path where an inverter drives a fan out of 4 through average length wires

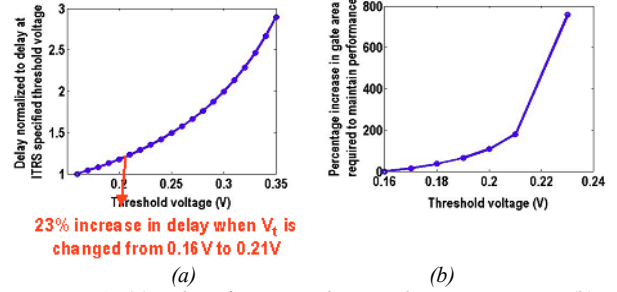


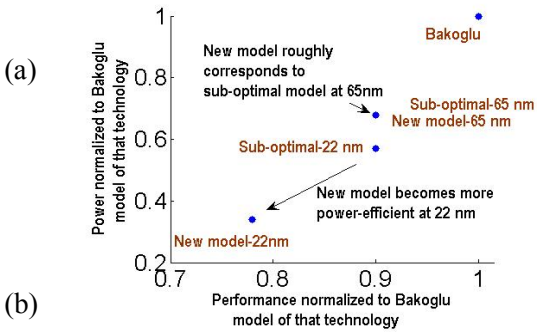
Figure 9: (a) Delay of a generic logic path is sensitive to  $V_t$  (b) Increasing gate sizes within practical values does not compensate performance losses associated with higher  $V_t$  values

Communication transistors can thus have higher  $V_t$  values than logic transistors, because repeated wires can achieve their target performance by more power-efficient techniques than use of lower  $V_t$  values, such as wire sizing and optimized repeater insertion. It is also useful to have a different channel length, gate dielectric thickness and/or supply voltage for logic and communication transistors. To the best of the authors' knowledge, this is the first time this idea has been proposed.

### (B) Optimized repeater insertion using new EDP model

Fig. 10(a) shows that the new model reveals designs that are more power-efficient than other models for leaky future technologies, since it allows increased performance loss with scaling. Fig. 10(b) indicates that increasing wire sizes compensates this performance loss but retains power efficiency. Fig. 10(b) also shows that use of the new model with separate  $V_t$  values for logic and repeater transistors reduces power by 53% compared to the sub-optimal model, for no performance penalty. There is a 15% wire area penalty, which as we shall see using MINDS, has a negligible impact at the system level. Previous models for repeater insertion such as [4][7] place repeaters based on a certain fixed delay penalty compared to Bakoglu's model. The advantage of using the model derived in this paper is that the delay penalty is optimally increased on scaling to give more benefits than [4][7] (Fig 10(a)).

System level benefits of the above two techniques: Results of MINDS in Fig. 11 show that *use of the new model and separate  $V_t$  values for logic and communication transistors reduces power of the logic block by 33.6%*. There is a 5.3% wire area overhead but no performance penalty.



Repeater insertion model	No. of rep.	Rep. size	Wire pitch (nm)	Delay (ps)	Power (uW)
Bakoglu [1]	19	31	65	69	4.6
Sub-optimal model [4]	8.5	35	72	69	2.6
New model	6.6	21	82	69	1.6
New model with optimal $V_t=0.21V$	7	27	83	69	1.2

Figure 10: (a) Comparison of different repeater insertion models in the power-performance space (b) Power savings for a 1mm wire at 22 nm. Size effects neglected in analysis. Frequencies at 65 nm and 22 nm are 500MHz and 1.4 GHz respectively. All wire cross-sectional dimensions assumed to scale with wire pitch

(C) 3D integration

Previous work [8][9] has analyzed the benefits of 3D integration in technologies upto 50 nm where repeater leakage power is not a significant portion of the system power. In this work, a 3D stochastic wire length distribution [8] and MINDS are used to show that 3D integration gives

more benefits in a 22nm technology with considerable repeater power. Fig. 11 indicates that 3D integration reduces die area by 33%, since die area is wire/repeater limited. Logic block power is reduced by 33% due to the smaller die area, shorter wires and smaller gate sizes that arise due to the reduced wire load. This 33% reduction in power, 33% reduction in die area and 66% reduction in chip dimensions caused by 3D integration has tremendous benefits for battery life, fabrication cost and form factor of future low power portable microchips. These advantages make 3D integration a key technology direction for many low power applications despite its increased power density. Note that face-to-face bonded 3D technologies have been shown to be manufacturable in [10].

**Results and Conclusions**

Fig. 11 thus shows that total power of a future 22 nm 1.4 GHz low power combinational logic block can be reduced by 55% and repeater power can be reduced by 80% by unique optimization of logic and communication transistor technologies, improved repeater insertion techniques and 3D integration.

**References**

[1] H.B. Bakoglu, J.D. Meindl, *TED*, pp. 903-909, May 1985  
 [2] P. Saxena, et al, *Trans. CAD of ICs and Systems*, Apr'04  
 [3] R. Venkatesan, PhD Thesis, Georgia Tech, 2003  
 [4] R. Venkatesan, et al, *Trans. VLSI Syst.*, Dec. 2001  
 [5] D. Sylvester, "BACPAC"  
[www.eecs.umich.edu/~dennis/bacpac/bacpac\\_models.html](http://www.eecs.umich.edu/~dennis/bacpac/bacpac_models.html)  
 [6] Berkeley Predictive Technology Model website (BPTM)  
 [7] G. Garcea, et al., *Proc. ICCAD'03*  
 [8] K. Banerjee, et al., *Proc. IEEE*, pp.602-633, May 2001  
 [9] A. Rahman, R.Reif, *Proc. IITC'01*  
 [10] P. Morrow, et al, *Proc. AMC'04*

	Sub-optimal model	New model + unique $V_t$ values for logic and repeaters	New model + unique $V_t$ values + 3D with 2 face-to-face bonded layers
Total power	26.2 W 	17.4 W 	11.7 W 
Frequency	1.4 GHz	1.4 GHz	1.4 GHz
No. of metal levels	7.6	8	8
Total repeater area	25 sq. mm	17.5 sq. mm	9.8 sq. mm
Die area	60 sq. mm	60 sq. mm	Two 20 sq. mm dice

Figure 11: Power savings at the system level for a 22nm 100 M gate combinational logic block. Rent's constants  $k=4$  and  $p=0.6$ . Size effects neglected in analysis