Compact Modeling and Optimization of fine-pitch interconnects for silicon interposers

Vachan Kumar, Li Zheng, Muhannad Bakir, and Azad Naeemi
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Abstract —**This paper presents the first optimization methodology for silicon interposer interconnect technology. The dimensions of these fine-pitch interconnects are roughly a few microns, because of which they can neither be treated as on-chip RC interconnects, nor as conventional off-chip interconnects. 3D extraction tools can provide an accurate estimate of the circuit parameters, but they prove to be very slow and tedious for design space exploration and optimization. Thus, the novel analytical models developed here for the frequency dependent resistance of fine-pitch interconnects are essential to efficiently optimize these interconnects. The error in the model is shown to be less than 15% for interconnect dimensions and frequency range of interest. The analytical models developed are then used to optimize the data-rate and cross-sectional dimensions to maximize the bandwidth-density and minimize the energy-per-bit, simultaneously.**

Keywords – Silicon interposer, fine-pitch interconnects, analytical modeling, optimization.

I. INTRODUCTION

As the minimum feature size in integrated circuits is scaled down to tens of nanometers, the transistors inside a chip continually improve in performance and energy consumption. However, the improvement in I/O performance cannot keep up with this rapid improvement in on-chip performance. This results in creating a severe I/O bottleneck [1]. 3D stacking of ICs [2] provides an innovative solution to tackle this problem. However, thermal issues have slowed down the use of 3D stacking for high performance ICs. This paves the way for integration of multiple dies on a silicon interposer [3]. Integration of ICs on a silicon interposer eases the I/O bottleneck, while avoiding the thermal issues of 3D stacking.

The primary advantage of integration on a silicon interposer is the use of fine-pitch interconnects, that improve the aggregate I/O bandwidth significantly. The dimensions of these fine-pitch interconnects are roughly a few microns. Thus, the frequency dependant resistance of these interconnects cannot be described by either the DC resistance that is used for on-chip interconnects, or the high frequency skin-effect resistance used for PCB interconnects. 3D resistance extractors can be used to accurately estimate the resistance of these fine-pitch interconnects. However, design space exploration and optimization using 3D resistance extractors is slow and tedious. As a result, a simple analytical model is developed to approximate the resistance of fine-pitch interconnects. To ensure that the model is accurate, it is compared against 3D resistance extraction using Synopsys Raphael [4], and the error is shown to be less than 15% in the frequency range of interest.

The analytical model developed is very useful for fast and efficient design space exploration and optimization. The model is applied to optimize the data-rate and interconnect dimensions, to maximize the aggregate bandwidth and minimize the energy per bit, simultaneously.

II. MODEL DEVELOPMENT AND VALIDATION

A. Mathematical Modeling

Multiple analytical models have been previously developed for the frequency dependent resistance of transmission lines [5], [6]. However, for the fine-pitch interconnects on a silicon interposer, the cross-sectional dimensions are such that the previously developed models have significant errors at the frequencies of interest. The new analytical model developed combines the asymptotic models for resistance in the low and the high frequency regions [7], with a fitting parameter κ to ensure continuity at the transition frequency f_0.

$$R_{low} = \sqrt{R_{dc}^2 + R_{ac}^2}$$
$$R_{dc} = \frac{\rho}{wt} \quad (1)$$
$$R_{ac} = \frac{\rho}{2\delta(w+t)}$$
$$R_{high} = \kappa \frac{\rho}{2\delta(w+t-2\delta)}$$
$$\delta = \sqrt{\frac{\rho}{\pi\mu f}}$$

where ρ is the resistivity of copper, w is the width, t is the thickness, δ is the skin depth, and f is the frequency. The high and low frequency models given above are combined to form

$$R = \begin{cases} R_{low}, f \le f_0 \\ R_{high}, f \ge f_0 \end{cases} \quad (2)$$
$$f_0 = \frac{4\rho}{\pi\mu t^2}$$

where the frequency f_0 is defined as the transition frequency at which the skin depth is equal to half of the conductor thickness t.

For the purpose of optimization, the capacitance matrix is computed using the analytical models developed in [8], and the inductance is assumed to be the external inductance. Although the inductance changes by approximately 25% from low to high frequencies, the impact of this change is primarily on the propagation delay, and not the loss. The capacitance and inductance matrices are then combined with the resistance model, to form an ABCD matrix for the system, including a low impedance driver and high impedance receiver [9]. To improve the response of the lossy channel, a 2-tap pre-emphasis filter is assumed to be present at the transmitter end.

978-1-4799-0440-2/13/$31.00 © 2013IEEE

B. Model Validation

The analytical models developed are validated using Synopsys Raphael using the differential stripline structure shown in Fig. 1. The frequency dependant resistance is normalized to the DC resistance of the interconnect, and the frequency is normalized to the transition frequency f_0. This normalization ensures that the resistance curves of Fig. 2 are scale-invariant, thus requiring analysis only for different ratios of (w/t), (s/w), and (h/w). Figure 3 shows that the error in the new model is less than 15% for the dimensions and frequency range of interest. The error in the new model is significantly better compared to the error with the simplistic resistance model given by (3) below. However, when the spacing (s_1 or s_2) is reduced below 0.75w, the error in the model shoots above 25% at certain frequencies due to significant change in the actual current distribution.

$$R = \begin{cases} R_{dc}, f \le f_0 \\ \dfrac{\rho}{2\delta(w+t-2\delta)}, f \ge f_0 \end{cases} \quad (3)$$

III. Optimization of Fine-Pitch Interconnects for Silicon Interposer

The main purpose of developing the analytical model in the previous section for fine-pitch interconnects is the design space exploration and optimization. In this section, we primarily focus on optimizing the data-rate and interconnect width, to maximize the aggregate bandwidth and minimize the energy per bit, simultaneously. The key metrics of interest include wiring density, bandwidth, power dissipation, and latency. However, since the inductance is assumed to be internal inductance, the latency is given by the speed of light in the medium. There exists a very interesting trade-off between the wiring density and aggregate bandwidth. As the width of each wire is reduced, the wiring density goes up, but the conductor losses in the wire also increase, resulting in a lower bandwidth for each wire. Thus, the metric bandwidth density, defined as the bandwidth per unit pitch, is used to optimize the interconnect dimensions. Similarly, an interesting trade-off exists between bandwidth and power dissipation. As the data-rate is increased, the bandwidth increases linearly, but the losses in the interconnect also go up. As a result, the signal voltage swing and hence the power dissipation at the driver are increased, in order to detect the signal reliably at the receiver.

The structure shown in Fig. 1 is used for optimization of data-rate and the cross-sectional dimensions. The height, thickness, and spacing s_1 are assumed to be constant and given by h=1µm, t=1µm, and s_1=2µm [9]. The optimization methodology is based on maximizing the compound metric, bandwidth density/energy per bit, similar to the technique used in [10]. In order to compensate for the frequency dependant losses in the interconnect, a 2-tap pre-emphasis filter is used. The upper limit on the data-rate, which is typically set by the transmitter and receiver circuits, is assumed to be 10Gbps. Figure 4 shows the normalized bandwidth density as a function of data-rate. Bandwidth density is linearly related to the data-rate. Figure 5 shows the energy per bit as a function of data-rate. At very small data-rates, as the data-rate is increased, higher number of bits is transmitted over the channel, for a very small penalty in power dissipation, thus reducing the energy per bit. However, above a certain data-rate, the losses in the channel are so high that the voltage swing at the driver and hence the energy per bit increase with data-rate. Figure 6 shows the compound metric (bandwidth density/energy per bit) as a function of data-rate for interconnect widths of 4 and 6 microns. For an interconnect length of 10mm, the optimal data-rate which maximizes the compound metric is approximately 9Gbps.

The optimization of interconnect width follows a similar procedure where the compound metric bandwidth density/energy per bit is maximized. Figure 7 shows the bandwidth density as a function of interconnect width at data-rates of 4Gbps and 8Gbps. For a given data-rate, as the interconnect width is increased, the bandwidth density goes down. Figure 8 shows the energy per bit as a function of interconnect width. As the interconnect width is increased, the losses in the interconnect decrease, and hence the energy per bit goes down. Figure 9 shows the optimum width at which the compound metric bandwidth density/energy per bit is maximized. Figure 10 shows the dependence of the optimal data-rate on the interconnect length. As the interconnect length is increased, the losses go up and hence the optimal data-rate decreases. Similarly, Fig. 11 shows the optimal width as a function of interconnect length. As the interconnect length is increased, the losses go up. To compensate for the increase in these losses, the optimal width increases with interconnect length.

IV. Conclusion

It is shown that compact analytical models developed for low and high frequency regions are sufficient to estimate the resistance of fine-pitch interconnects with an error less than 15% in the frequency range of interest. The optimization of data-rate and width is done to minimize energy per bit and maximize bandwidth density, simultaneously. It is shown that the optimal data-rate and optimal interconnect width are dependant on the interconnect length. For an interconnect length of 1cm, the optimal width is ~6µm and the optimal data-rate is ~9Gbps.

Acknowledgements

This work was supported by Semiconductor Research Corporation (SRC) Global Research Collaboration (GRC) (task ID: 2254.001).

References

[1] O'Mahony, F *et. al. VLSI-DAT '09*.

[2] Lu, J-Q, *Proc. IEEE* 2009.

[3] J. U. Knickerbocker *et al.*, *IBM J. Res. Develop.*, 2005.

[4] http://www.synopsys.com/TOOLS/TCAD/INTERCONNECTSIMULATION/Pages/Raphael.aspx

[5] Krauter, B *et al.*, DAC 1998.

[6] Djordjevic, A.R. *et al*, *IEEE MTT.*, February 1994.

[7] Stephen H. Hall *et al*, John Wiley & Sons, 2009.

[8] Zhao, W *et al* , IEEE *TED* , September 2009

[9] Dickson, T *et al , IEEE JSSC, April 2012*

[10] V. Kumar *et al*, *ECTC* 2011

Figures

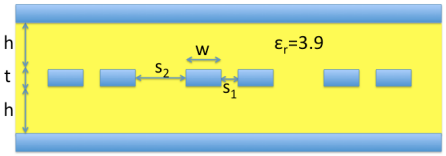

Figure 1: Cross-section of a stripline differential pair.

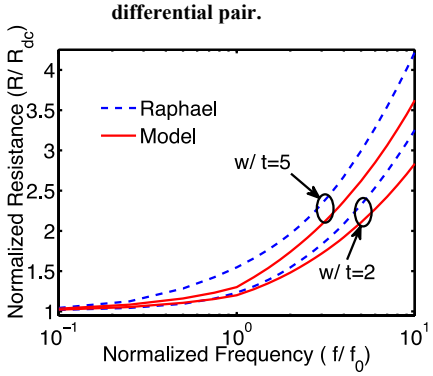

Figure 2: Normalized resistance as a function of normalized frequency for various values of (w/t). The other dimensions are s_1=1.33*t*, *h*=*t*, s_2=*w*

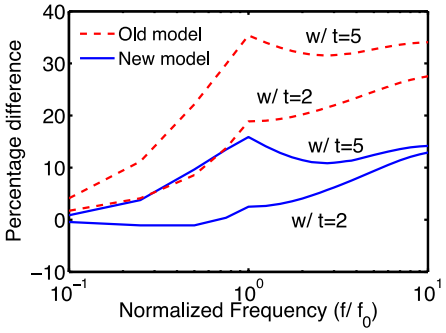

Figure 3: Error in the old model (3), and new model (2). The dimensions are s_1=1.33*t*, *h*=*t*, s_2=*w*

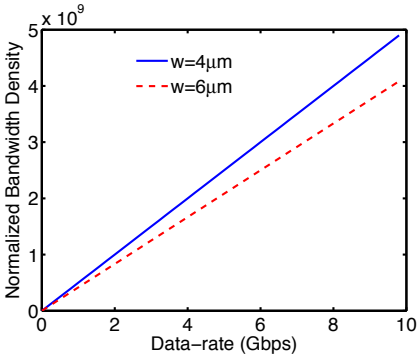

Figure 4: Bandwidth density as a function of data-rate for 2 different interconnect widths.

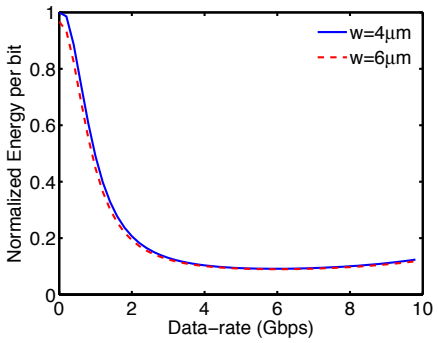

Figure 5: Energy per bit as a function of data-rate for 2 different interconnect widths.

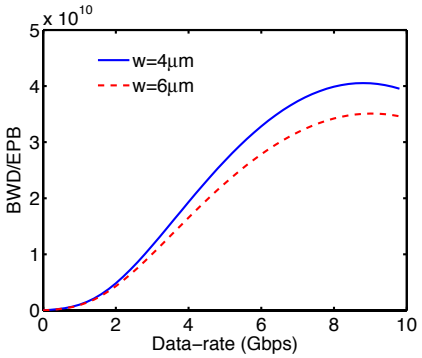

Figure 6: Compound metric (Bandwidth density/Energy per bit) as a function of data-rate for 2 different interconnect widths.

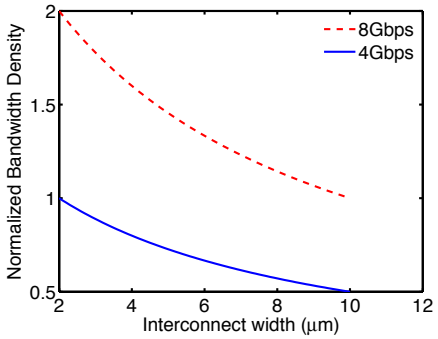

Figure 7: Bandwidth density as a function of interconnect width for 2 different data-rates.

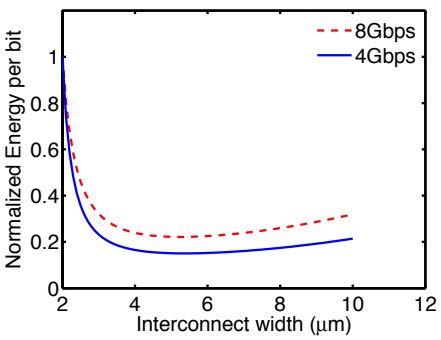

Figure 8: Energy per bit as a function of interconnect width for 2 data-rates.

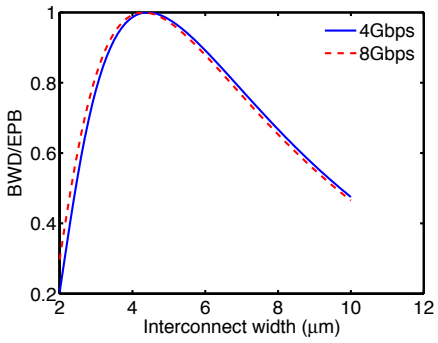

Figure 9: Compound metric (Bandwidth density)/(Energy per bit) as a function of interconnect width for 2 different data-rates.

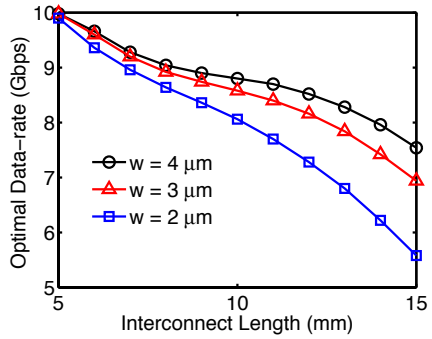

Figure 10: Optimal data-rate as a function of interconnect length, for 3 different values of interconnect width.

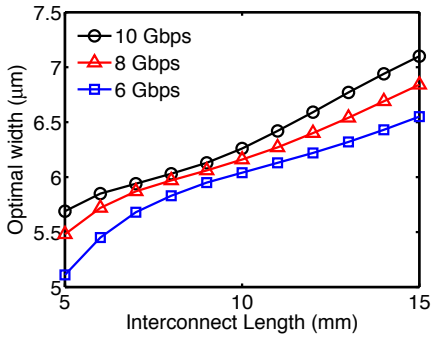

Figure 11: Optimal interconnect width as a function of interconnect length for 3 different values of data-rates.