

Performance Trend in Three-Dimensional Integrated Circuits

Author list
Affiliation
xxx@xxx

ABSTRACT

3DICs are motivated by the expectation of better performance over their 2D counterparts; however, non-idealities threaten to diminish the benefit of going vertical. This work develops an automated design flow with 2D CAD tools to design 3DICs with the MIT Lincoln Lab 0.18 μm three-tier fully depleted silicon on insulator (FDSOI) process [10]. This flow uses carefully designed scripts to fill the gap between 2D methodologies and 3D designs. We examine wire-length, timing, clock skew, and total power dissipation, along with temperature, of two benchmark circuits implemented in both 2D and 3D integration. We then extend our observations to the 90nm and 45nm technology nodes with Predictive Technology Model (PTM) [4] and the BSIMSOI model [3]. Experimental results show that the performance of 3DIC, even with the non-idealities, shows up to two-generation advantage over its 2D counterpart with only three tiers.

Categories and Subject Descriptors

[Interconnect Systems]: Novel architectures

Keywords

3DIC, 3D integration, system performance, 3D clock skew, power

1. INTRODUCTION

Interconnect dominates the system performance in deep sub-micron (DSM) ICs [12]. 3D integration has attracted a lot of attention for its ability to reduce interconnect wire-length in DSM ICs [2] [11]. As shown in Figure 1, the long global interconnect can be replaced by a short vertical Inter Silicon-tier Via (ISV) if the circuit is properly folded. Due to the reduction of wire-length, 3D systems are expected to be able to run at a higher clock rate or consume less power at the same clock rate.

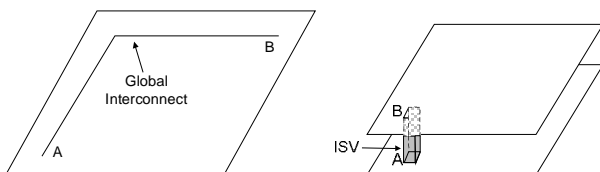


Figure 1: Example of wire-length reduction in 3DIC

Total interconnect capacitance for a system will decrease as the number of tiers increases, up to a maximum of a 40% reduction for 6 tiers [11]. Heat dissipation is a major obstacle to achieving better performance because device running at

higher temperature will have extra delay and leakage power. This work shows a performance analysis and comparison on two test circuits representing both low-power and high-performance designs. The low-power design is based on the 8-point Fast Fourier Transform (FFT) presented in [7], but has been pipelined to increase throughput. The high-performance design is based on the OpenRISC Platform System-on-Chip (ORPSOC) [13], which includes a 32-bit OpenRISC micro-processor, memory controllers, and 40KB of embedded memories. We have modified the ORPSOC to include a second OpenRISC core that communicates with the same memories through a bus arbitration unit. The FFT design has 158K gates, while the ORPSOC design has 120K gates and 320K SRAM cells.

2. CONTRIBUTIONS

The major contribution of this work is that we developed a flow, modified from [5] and shown in Figure 2, to perform 3DIC design and verification. This flow partitions the entire circuit into three tiers with METIS [14] followed by a tier specific physical design. Then the parasitics (in SPEF format) are extracted and merged for performance analysis. Finally, an iterative analysis is carried out to determine the final performance. This flow is described more completely in [7] except that this work supports up to seven metal layers. Another contribution of this work is that we have developed new models to iteratively compute the final 3DIC performance (Figure 2). The models are presented in section 3.

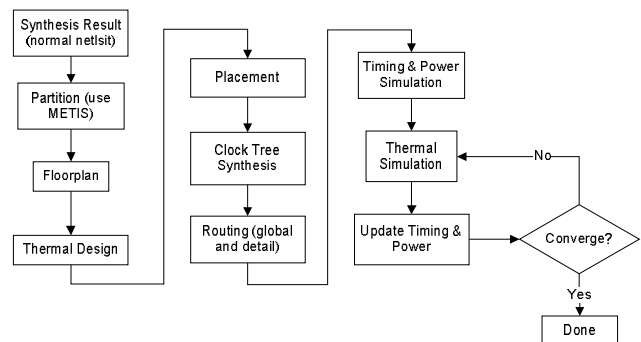


Figure 2: 3DIC design flow

Though thermal vias are useful in reducing temperature in 3DIC [15], we find that they do not help in our particular test cases and packaging method. Therefore, thermal vias are not inserted to save chip area in this study.

3. THERMAL AND PERFORMANCE MODELS

3.1 Thermal Models

Assuming an adiabatic boundary condition on the sides of the die not connected to the heat-sink, the 3D system fits the thermal model shown in Figure 3, where $R_{2,1}$, $R_{3,2}$ and $R_{1,0}$ are the equivalent thermal resistance between active layers of tier 1 and 2, tier 2 and 3, and active of tier 1 and top of silicon handling layer respectively; $R_{Silicon}$ and R_{paste} represent the thermal resistance of the silicon handling layer and the thermal paste. The power consumed on each tier ($P_{tier1,2,3}$) is modeled as constant current sources ($I_{tier1,2,3}$) and the temperature ($T_{tier1,2,3}$) is modeled as voltages. With such assumption, the temperature rise on tier i can be written as (1) [1].

$$\Delta T_i = R_{i,i-1} \sum_{j=1}^3 P_j \quad (1)$$

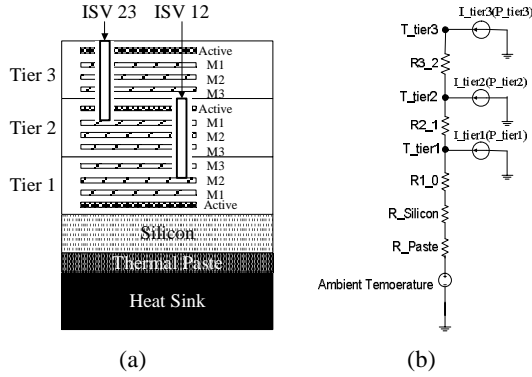


Figure 3: (a) package and (b) thermal model

3.2 Performance Models

The transistor delay-temperature dependency can be expressed as (2)-(4) [16]. According to [6], k is also temperature dependent. Since its variation is very small, we treat it as a linear function (5) for simplicity. Substituting (3) and (4) into (2), the normalized delay can be expressed as (6). The temperature range we studied in this work is $0^\circ\text{C} \sim 150^\circ\text{C}$ because the 45nm device breaks down when temperature exceeds 150°C . For temperature exceeding 150°C , we will assume the circuit fails to operate properly. Dynamic power also varies as delay changes. We assume that the dynamic power, comprised of the short circuit power and switching power, is linearly proportional to clock frequency (7).

$$\text{Delay} \propto \frac{CV_{DD}}{I_D} \propto \frac{CV_{DD}}{m(T)(V_{DD} - V_{TH}(T))^a} \quad (2)$$

$$V_{TH}(T) = V_{TH}(T_0) - k(T - T_0) \quad (3)$$

$$m(T) = m(T_0) \left(\frac{T}{T_0}\right)^{-b} \quad (4)$$

$$k = k_0 + g(T - T_0) \quad (5)$$

$$\text{delay}(T) = \frac{\text{delay}(T_0)(V_{DD} - V_{TH}(T_0))^a T^b}{T_0^b (V_{DD} - V_{TH}(T_0) + k(T - T_0))^a} \quad (6)$$

$$P_{dynamic}(f) = P_{dynamic}(f_0) \cdot f / f_0 \quad (7)$$

Table 1 shows k_0 , γ , β , and α values for our cell library implemented in different technologies. We have performed extensive simulations on the standard cell library used in this study to test the accuracy of our model (6). The largest disparity between SPICE and our model in the 180nm technology node is an XOR gate with an overestimated delay error of 11.6%. In 90nm and 45nm technologies, the largest error between SPICE and our model is still the XOR gate, with overestimated delay errors of 10% and 12.8% respectively.

Table 1: Coefficients of delay temperature dependency model

| tech | edge | k_0 (mV/K) | γ (mV/K) | β | α |
|--------|------|--------------|-----------------|---------|----------|
| 180 nm | Rise | 0.12 | 0.003 | 1.5 | 1.3 |
| | Fall | 0.1 | 0.003 | 1.5 | 1.4 |
| 90 nm | Rise | 0.5 | 0.003 | 1.5 | 1.3 |
| | Fall | 0.5 | 0.003 | 1.5 | 1.4 |
| 45 nm | Rise | 0.8 | 0.003 | 1.5 | 1.3 |
| | Fall | 0.8 | 0.003 | 1.5 | 1.4 |

According to [17], the leakage-temperature dependency is super-linear, and can be modeled as a polynomial function (8). Using a curve fitting technique, we find values for the coefficients a_1 , a_2 and a_3 in (8) for each standard cell and derive the average values for our library to estimate the leakage power with temperature as a variable. The average values for each technology are given in Table 2.

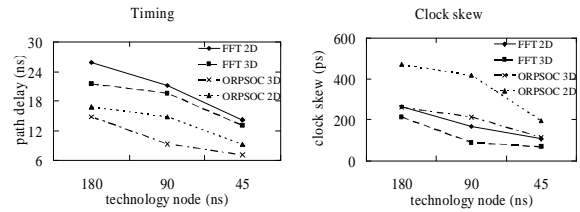
$$\frac{P_{leakage}(T)}{P_{leakage}(T_0)} = 1 + a_1 \cdot (T - T_0) + a_2 \cdot (T - T_0)^2 + a_3 \cdot (T - T_0)^3 \quad (8)$$

Table 2: Coefficients of leakage model

| coefficient | a1 | a2 | a3 |
|---------------|---------|---------|---------|
| value (180nm) | 0.0226 | 0.00033 | 1.77E-6 |
| value (90nm) | 0.03147 | 0.00038 | 1.87E-6 |
| value (45nm) | 0.02728 | -0.0018 | 3.0E-5 |

4. EXPERIMENTAL RESULTS

We have implemented FFT and ORPSOC using same constraints with the flow discussed in section 2, and then carried out a performance analysis with extracted RC parasitics. We used the temperature dependency models



(a) Timing

(b) Clock skew

Figure 4: Performance of test circuits vs feature size

shown in section 3, through a proper iteration technique, to determine the final performance values. Figure 2 shows how the timing and clock skew of the FFT and the ORPSOC change as the feature size shrinks. Table 3 and Table 4 gives more detailed information of the FFT and ORPSOC performance in various technology nodes.

Table 3: Performance summary (a) 2D (b) 3D

| tech | Design | Timing | Power | A(mm ²) | Max T |
|--------|--------|----------|----------|---------------------|-------|
| 180 nm | FFT | 25.81 ns | 1050 mW | 11.61 | 39 °C |
| | ORPSOC | 17.8 ns | 3298 mW | 18.66 | 62 °C |
| 90 nm | FFT | 21.14 ns | 439.0 mW | 4.18 | 41 °C |
| | ORPSOC | 14.82 ns | 1944 mW | 6.76 | 82 °C |
| 45 nm | FFT | 14.16 ns | 333.9 mW | 2.32 | 45 °C |
| | ORPSOC | 9.16 ns | 1215 mW | 3.26 | 98 °C |

(a) 2D integration performance summary

| tech | Design | Timing | Power | A (mm ²) | Max T |
|--------|--------|----------|----------|----------------------|--------|
| 180 nm | FFT | 21.49 ns | 951.8 mW | 4.01 | 57 °C |
| | ORPSOC | 14.78 ns | 2933 mW | 6.72 | 92 °C |
| 90 nm | FFT | 19.54 ns | 394.1 mW | 1.48 | 60 °C |
| | ORPSOC | 10.02 ns | 1551 mW | 2.42 | 112 °C |
| 45 nm | FFT | 13.01 ns | 255.6 mW | 0.82 | 65 °C |
| | ORPSOC | 8.33 ns | 1029 mW | 1.29 | 128 °C |

(b) 3D integration performance summary

Table 4: Clock skew and power in 2D/3D integration

| tech | Design | Clock skew (ps) | | Clock power (mW) | |
|--------|--------|-----------------|-------|------------------|-------|
| | | 2D | 3D | 2D | 3D |
| 180 nm | FFT | 264.3 | 213.4 | 240.5 | 201.9 |
| | ORPSOC | 469.6 | 260.6 | 838.9 | 482.9 |
| 90 nm | FFT | 167 | 88.8 | 61.0 | 44.6 |
| | ORPSOC | 317.6 | 213 | 694.2 | 419.1 |
| 45 nm | FFT | 106.3 | 68.1 | 66.4 | 48.3 |
| | ORPSOC | 194.9 | 115.1 | 798.1 | 487.9 |

As can be seen, the total power consumption of the 3D integration is significantly reduced due to the power saving on clock tree and switching power. The maximum temperature rises as the feature size shrinks because the power density increases. In the 45nm technology node, the temperature is approaching its break-down point. The leakage power increases significantly when the system becomes hot. In the 45nm ORPSOC design, 13X more leakage power is consumed compared to its nominal value.

5. CONCLUSION

In this paper, we have demonstrated the performance advantage of novel 3D integration over conventional 2D integration. The 3D integration with 3 tiers can improve system timing the equivalent of up to two technology generations. The maximum delay and power reductions are 32% (ORPSOC, 90nm) and 23% (FFT, 45nm), respectively. However, due to the dramatic temperature increase, we cannot integrate more than 3 tiers in high-performance applications unless a new heat removal technique becomes available.

6. ACKNOWLEDGMENTS

The authors would like to thank DARPA for supporting this work. We would also like to thank Cadence, Synopsys, PTC, and Ansoft for generously providing the CAD tools. Thanks to MIT Lincoln Labs for providing access to their FD-SOI library and for their aid in developing our design kit. Lastly, thanks to James Stine at the Illinois Institute of Technology for generously providing access to the IIT-SoC standard-cell characterization scripts.

7. REFERENCES

- [1] A. Rahman *et al*, "Thermal analysis of three-dimensional (3-D) integrated circuits (ICs)", *IITC*, 2001
- [2] K. Banerjee *et al*, "3-D ICs: a novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration", *Proceedings of the IEEE*, 2001
- [3] Berkeley SIMSOI model, available online at: <http://www-device.eecs.berkeley.edu/~bsimsoi/>
- [4] Predictive Technology Model, available online at: <http://www.eas.asu.edu/~ptm/>
- [5] H. Hua *et al*, "The 3DIC Phase 1 Place and Route Flow", available online at <http://www.ece.ncsu.edu/muse/flowdb>
- [6] G. Groeseneken *et al*, "Temperature Dependence of Threshold Voltage in Thin-Film SOI MOSFET's", *IEDL*, 1990
- [7] W. R. Davis *et al*, "Demystifying 3D ICs: the pros and cons of going vertical," *DTOC*, 2005
- [8] S. Im *et al*, "Full chip thermal analysis of planar (2-D) and vertically integrated (3-D) high performance ICs", *IEDM* 2000
- [9] W. Liao *et al*, "Temperature and Supply Voltage Aware Performance and Power Modeling at Microarchitecture Level", *TCAD*, 2005
- [10] V. Suntharalingam *et al*, "Megapixel CMOS Image Sensor Fabrication in Three-Dimensional Integrated Circuit Technology", *ISSCC*, 2005
- [11] R. Zhang *et al*, "Power trends and performance characterization of 3-dimensional integration for future technology generations", *ISQED*, 2001
- [12] International Technology Roadmap for Semiconductors, available online at <http://public.itrs.net>
- [13] *OpenRISC Reference Platform System-on-a-Chip and OpenRISC 1200 IP Core Specification*, available online at <http://www.opencores.org/projects.cgi/web/or1k/orpso>
- [14] G. Karypis and V. Kumar, *The METIS Serial Graph Partitioning Tool*, available online at <http://www-users.cs.umn.edu/~karypis/metis>
- [15] B. Goplen *et al*, "Thermal Via Placement in 3D ICs", *ISPD*, 2005
- [16] K. Kanda *et al*, "Design impact of positive temperature dependence on drain current in sub-1-V CMOS VLSIs", *JSSCC*, 2001
- [17] H. Su *et al*, "Full Chip Leakage Estimation Considering Power Supply and Temperature Variations", *ISLPED*, 2003