

Multiple Si Layer ICs: Motivation, Performance Analysis, and Design Implications

Shukri J. Souri

Kaustav Banerjee

Amit Mehrotra¹

Krishna C. Saraswat

Department of Electrical Engineering, Stanford University, Stanford, CA, 94305.

E-mail: [ssouri, kaustav, saraswat]@stanford.edu

Abstract

Continuous scaling of VLSI circuits is reducing gate delays but rapidly increasing interconnect delays. Semiconductor Industry Association (SIA) roadmap predicts that, beyond the 130 nm technology node, performance improvement of advanced VLSI is likely to saturate unless a paradigm shift from present IC architecture is introduced. This paper presents a comprehensive analytical treatment of ICs with multiple Si layers (3-D ICs). It is shown that significant improvement in performance (more than 145%) and reduction in wire-limited chip area can be achieved with 3-D ICs with vertical inter-layer interconnects (VILICs). This analysis is based on dividing a chip into separate blocks, each occupying a separate physical level. A scheme to optimize interconnect distribution among different interconnect tiers is presented and the effect of transferring the repeaters to upper Si layers has been quantified in this analysis. Furthermore, thermal analysis of ICs with two Si layers is presented. It is demonstrated that using a thermally responsible design and/or a high-performance heat sinking technology, die temperatures for ICs with two Si layers can be reduced well below present die temperatures. Finally, implications of 3-D architecture on several circuit designs are also discussed.

1 Motivation for Multiple Si Layer (3-D) ICs

1.1 Interconnect Limited VLSI Performance

In single Si layer (2-D) ICs, chip size is continually increasing despite reductions in feature size made possible by advances in IC technology such as lithography, etching etc., and reduction in defect density [1]. This is due to the ever-growing demand for functionality and higher performance, which causes increased complexity of chip design, requiring more and more transistors to be closely packed and connected [2]. Smaller feature sizes have dramatically improved device performance [3-5]. The impact of this miniaturization on the performance of interconnect wires, however, has been less positive [6-9]. Smaller wire cross-sections, smaller wire pitch and longer lines to traverse larger chips have increased the resistance and the capacitance of these lines resulting in a significant increase in signal propagation (RC) delay. As interconnect scaling continues, RC delay is increasingly becoming the dominant factor determining the performance of advanced ICs [1,6-9]. Therefore, as feature sizes are further reduced and more devices are integrated on a chip, the chip performance will degrade, reversing the trend that has been observed in the semiconductor industry thus far.

At 250 nm technology node, Cu with low-k dielectric was introduced to alleviate the adverse effect of increasing interconnect delay [10-14]. However, below 130 nm technology node, substantial interconnect delays will result in spite of introducing these new materials, which in turn will severely limit the chip performance [1]. Further appreciable reduction in interconnect delay cannot be achieved by introducing any new materials. This problem is especially acute for global interconnects, which typically comprise about 10% of total wiring, for current architectures. Therefore it is apparent that material limitations will ultimately limit the performance improvement as the technology scales. Also the problem of long-lossy lines cannot be fixed by simply widening the metal lines and using thicker interlayer dielectric since this conventional solution will lead to a sharp increase in the number of metallization layers. Such an approach will increase the complexity, reliability, and cost, and will therefore be fundamentally incompatible with the industry trend of maximizing the number of chips per wafer, and 25% per year improvement in cost per chip function.

Greater performance and greater complexity at lower cost are the drivers behind large scale integration. In order to maintain these driving forces it is necessary to find a way to keep increasing the number of devices on a chip, yet limit or even decrease the chip size to keep interconnect delay from affecting chip performance.

Permission to make digital/hardcopy of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copying is by permission of ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC 2000, Los Angeles, California

(c) 2000 ACM 1-58113-188-7/00/0006..\$5.00

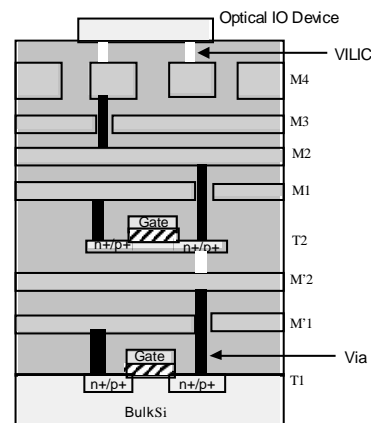


Figure 1. Schematic representation of 3-D integration with multilevel wiring network and VILICs. T1: first active layer device, T2: second active layer device, Optical I/O device: third active layer I/O device. M'1 and M'2 are for T1, M1 and M2 are for T2. M3 and M4 are shared by T1, T2, and the I/O device.

A decrease in chip size will also assist in maximizing the number of chips per wafer, thus maintaining the trend of decreasing cost function. Therefore innovative solutions beyond mere materials changes are required to meet future IC performance goals [15]. We need to think beyond the current paradigm of system architecture.

1.2 3-D Architecture

3-D integration (schematically illustrated in Figure 1) to create multilayer Si ICs is a concept that can significantly alleviate interconnect delay problem, increase transistor packing density and reduce chip area.

The entire chip is divided into a number of blocks and each block is placed on a separate layer of Si, which are stacked on top of each other. Each Si layer in the 3-D structure can have multiple layers of interconnect. Each of these layers are connected together with vertical inter-layer interconnects and common global interconnects as shown in Figure 1. The 3-D architecture offers extra flexibility in system design, placement and routing. For instance, logic gates on a critical path can be placed very close to each other using multiple active layers. This would result in a significant reduction in RC delay and can greatly enhance the performance of logic circuits. This technology can be exploited to build systems-on-a-chip by placing circuits with different voltage and performance requirements in different layers. One such example will be to have logic circuits in the first Si layer and then have memory (SRAM) circuits in the second layer to realize distributed memory systems in a microprocessor. The advantages of using this technology are elaborated in Sections 4 and 6.

2 Overview of 3-D IC Technology

2.1 3-D Technology Options

Although the concept of 3-D integration was demonstrated as early as in 1979 [16], and was followed by a number of reports on its fabrication process and device characteristics [17-26], it largely remained a research technology, since microprocessor performance was device limited. However, with the growing menace of RC delay in recent times, this technology is being viewed as a potential alternative that can not only maintain chip performance well beyond the 130 nm node, but also inspire a new generation of circuit design concepts. Hence, there has been a renewed spur in research activities in 3-D technology [27-31] and their performance modeling [32-36].

Presently, there are several possible fabrication technologies that can be used to realize multiple layers of active-area (single crystal Si or recrystallized poly-Si) separated by inter-layer dielectrics (ILDs) for 3-D circuit processing. A brief description of these alternatives is given below. The choice of a particular technology for fabricating 3-D circuits will depend on the requirements of the circuit system, since the circuit performance is strongly influenced by the electrical characteristics of the fabricated devices as well as on the

¹Department of ECE, University of Illinois, Urbana-Champaign, IL 61801.

manufacturability and process compatibility with the relevant 2-D technology.

Beam Recrystallization: A very popular method for fabricating a second Silicon layer on top of an existing substrate is to deposit polysilicon and fabricate thin film transistors (TFT). To enhance the performance of TFTs, an intense laser or electron beam is used to induce re-crystallization of the polysilicon film [16-26]. This technique however may not be very practical for 3-D devices because of the high temperature involved during melting of the polysilicon and also due to difficulty in controlling the grain size variations. Beam recrystallized polysilicon films can also suffer from lower carrier mobility and unintentional impurity doping. However, high-performance TFTs fabricated using low temperature processing [37], and even low-temperature single-crystal Si TFTs have been recently demonstrated [38] that can be employed to fabricate advanced 3-D circuits.

Processed Wafer Bonding: Another alternative is to bond two fully processed wafers, on which chips are fabricated on the surface including some interconnects, such that the chips completely overlap [28]. Vias are etched to electrically connect both chips after metallization. A backside of the bonded pair can be back etched to allow further processing or the bonding of more pairs in this vertical fashion. Other advantages of this technology lie in the similar electrical properties of devices on all active levels and the independence of processing temperature since all chips can be fabricated separately and later bonded. The major limitation of this technique is its lack of precision (best case alignment $\pm 2 \mu\text{m}$) which restricts the inter-chip communication to global metal lines. However, for applications where each chip is required to perform independent processing before communicating with its neighbor, this technology can prove attractive.

Silicon Epitaxial Growth: Another technique for forming additional Si layers is to etch a hole in a passivated wafer and epitaxially grow a single crystal Si seeded from open window in the ILD. The silicon crystal grows vertically and then laterally, to cover the ILD [30]. In principle, the quality of devices fabricated on these epitaxial layers can be as good as those fabricated underneath on the seed wafer surface, since the grown layer is single crystal with few defects. However, the high temperatures ($\sim 1000^\circ\text{C}$) involved in this process cause significant degradation in the quality of devices on lower layers. Also this technique cannot be used over metallization layers. Low temperature silicon epitaxy using ultra-high-vacuum chemical vapor deposition (UHV-CVD) has been recently developed [39]. However, this process is not yet manufacturable.

Solid Phase Crystallization (SPC): As an alternative to high temperature epitaxial growth, low temperature deposition and crystallization of amorphous silicon, which passivates the lower active layer devices, can be employed. The amorphous film can be randomly crystallized to form a polysilicon film [40-43].

TFT performance can be enhanced by eliminating grain boundaries. For this purpose, local crystallization can be induced using low temperature processes such as using patterned seeding of Germanium [29,44], or by using Metal Induced Lateral Crystallization (MILC) [45-46]. The SPC technique offers the flexibility of creating multiple active layers and is compatible with current processing environments. Recent results prove the feasibility of building high performance TFTs at low processing temperatures which can be compatible with lower level metallization [47]. MILC, for example, can be used to build repeaters above metal lines. It is found that the electrical characteristics of these TFTs (although superior among their peers) are still inferior to single crystal devices. However, technological advances to overcome the thermal budget problem have been made to allow fabrication of high-performance TFTs using SPC [48-50].

It is possible to conceive of several 3-D circuits for which SPC will be a suitable technology, such as in upper-level non-volatile memory, or by simply sizing up the TFTs to match their single crystal CMOS counterparts. For example, deep sub-micron polysilicon TFTs [51], stacked SRAM cells [52,53], and EEPROM cells [54] have already been demonstrated. The MILC process can also be used to fabricate islands of single-grain-devices to maximize circuit performance.

2.2 Vertical Inter-Layer Interconnect Technology Options

The performance modeling presented in this study directly relates improved chip performance with increased utility of VILICs. It is therefore important to understand how to connect different active layers with a reliable and compatible process. Upper-layer processing needs to be compatible with metal lines underneath connecting lower layer devices and metal layers. With Cu technologies, this limits the processing temperatures to $< 450^\circ\text{C}$ for upper layers. Otherwise, Cu diffusion through barrier layers, and the reliability and thermal stability of material interfaces can degrade significantly. Heavily doped polysilicon can be used as a VILIC material, however, p-type to n-type contacts limit the use of this material. Polysilicon will also significantly increase the interconnect resistance. Tungsten is a refractory metal that can be used to withstand higher processing temperatures. Recently, a metallization scheme for 3-D VILICs has been demonstrated [55].

3 Scope of This Study

A 3-D solution at first glance seems an obvious answer to the interconnect delay problem. Since chip size directly affects the interconnect delay, therefore by creating a second active layer, the total chip footprint can be reduced, thus shortening critical interconnects and reducing their delay. However, in today's microprocessors, the chip size is not just limited by the cell size, but also limited by how much metal is required to connect the cells. The transistors on the silicon surface are not actually packed to maximum density but are spaced apart to allow metal lines above to connect one transistor or one cell to another. The metal required on a chip for interconnections is determined not only by the number of gates, but also by other factors such as architecture, average fan-out, number of I/O connections, routing complexity etc. Therefore, it is not obvious that by using a 3-D structure, the chip size will be reduced.

In this work we study the possible effects of 3-D integration on chip area and performance by modeling the optimal distribution of the metal interconnect lines. To better understand how a 3-D design will affect the amount of metal wires required for interconnections, we applied a stochastic approach for estimating wiring requirements derived for a single layer in [56] and modified it for 3-D ICs to quantify effects on interconnect RC delay. Unlike previous work [57], wire-pitch limited chips are considered. The study is based on Rent's Rule [58] that relates the number of I/O pins to the number of blocks in a processor. Rent's rule is suitably modified to be applicable to all the possible 3-D scenarios.

The results obtained in Section 4 indicate that when critically long metal lines that occupy lateral space are replaced with effective VILICs to connect logic blocks on different Si layers, a significant chip area reduction can be achieved. VILICs are found to be ultimately responsible for this improvement. The assumption made here is that it is possible to divide the microprocessor into different blocks such that they can be placed on different levels of active silicon.

We have also quantified thermal effects in 3-D ICs by estimating the die temperatures for two different 3-D designs in Section 5 and demonstrated the growing need for advancement in IC cooling technology for maximizing 3-D circuit performance.

Throughout this work no differences were assumed in the performance or the properties of the individual devices on any layer. Also the treatment is independent of the 3-D technology used. However, even if the properties of the devices on the upper Si layers are different, these layers can be used for memory devices or repeaters. Some of these applications are discussed in Section 6.

4 Performance Estimation of 3-D ICs

We now present a methodology, which can be used to provide an initial estimate of the area and performance of high-speed logic circuits fabricated using multiple Silicon layer IC technology. This approach is inspired by [56,59], where a delay and area estimation scheme for conventional IC technologies was presented. A brief summary of their approach is as follows: Based on Rent's rule, a stochastic wire-length distribution function was proposed. Using a three tier interconnection structure (local, semi-global and global, illustrated in Figure 2), the semi-global tier pitch that minimizes the wire limited chip area was determined. The maximum interconnect length on any given tier was determined by the interconnect delay criterion. This distribution was used to enhance an existing critical path model to estimate clock frequencies.

In the analysis presented below, we modify Rent's rule for 3-D ICs and use it to obtain wire length distributions and compare our results with the 2-D case. The analysis is carried out for different technology nodes, with primary focus on the 50 nm technology node based on [1].

4.1 Rent's Rule Modification

To modify Rent's rule so that it is applicable to 3-D ICs, we first note that this rule is also applicable to individual logic blocks. For the case of 3-D ICs, different blocks can be physically placed on different Silicon layers and connected to each other using VILICs. The area saving by using VILICs can be computed by modifying Rent's rule suitably and extending the stochastic wire-length distribution analysis to 3-D ICs. For simplicity, we consider the case where two Silicon layers are available. The extension to M-layer case is straightforward. An N gate design is divided into two $N/2$ gate blocks. It is assumed that the routing algorithm and overall logic style is the same for both layers. This ensures that Rent's constant and Rent's exponent are the same for both layers. Applying Rent's rule to both the blocks, we have,

$$T_1 = T_2 = k \left(\frac{N}{2} \right)^p \quad (1)$$

Here T_1 and T_2 are I/Os of each block, p is Rent's exponent and k is the average number of I/Os per gate. Moreover, for the overall design,

$$T = k N^p = T_1 + T_2 - T_{int} \quad (2)$$

where T is the number of I/Os of the entire design and T_{int} represents the total number of I/O ports connecting the two blocks (Figure 3). Each block will have ($T_{int}/2$) dedicated I/O ports for connection to the other block.

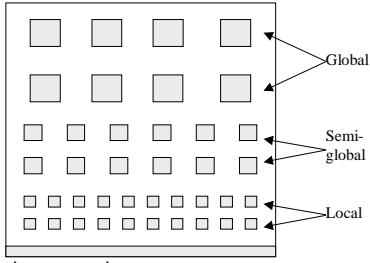


Figure 2. A three-tier interconnection structure.

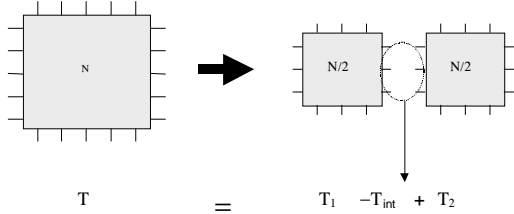


Figure 3. Total number of external I/O ports is conserved to maintain constant functionality of chip.

Hence the number of external I/O ports per block, i , is given by,

$$T_{ext,i} = T_i - \frac{T_{int}}{2} = k 2^{p-1} \left(\frac{N}{2} \right)^p \quad i=1, 2 \quad (3)$$

Comparing this with Rent's equation, we find that for each block, the Rent's exponent is unchanged, the number of gates is halved and the effective number of I/Os per gate used for connecting other gates on the same layer is given by $k_{eff,int} = k (1-2^{p-1})$. Similarly we have $k_{eff,ext} = k 2^{p-1}$ which is the effective number of I/Os per gate used to connect to gates on the other active layer. Now an analysis based on the technique presented in [56] can be carried out with these modified values of k for each layer and wire limited areas can be estimated. Furthermore, if the area occupied by VILICs is assumed to be very small compared to the chip area, additional area savings can be achieved.

4.2 Two Active Layer 3-D Circuit Performance

The above analysis is used to compare area and delay values for 2-D and 3-D ICs. The availability of additional Silicon layers gives the designer extra flexibility in trading off area with delay. A number of different cases are discussed below.

4.2.1 Case I: Wire-length Distribution for Chip Area Minimization with Fixed Interconnect Delay

The model is applied to the microprocessor example shown in Table I (data from [1], NTRS '97 for 50 nm technology node) for the two cases where all gates are in a single layer and where the gates are equally divided between two layers. In this calculation VILICs are assumed to consume negligible area, interconnect line width is assumed to equal half the metal pitch at all times, and the total number of metal layers for 2-D and 3-D case was conserved.

Figure 4 shows the variation in chip area with the normalized semi-global tier pitch for a fixed operating frequency. Local tier pitch is assumed to be twice the minimum feature size and global tier pitch is determined similarly to [56]. The curve for the 3-D case has a minimum similar to the one obtained in [56]. It is observed that the optimum chip area reduces by 50% from 2-D to 3-D case. Moreover, since the total wiring requirement is reduced as shown in Figure 5, semi-global tier pitch is reduced. This increases the line resistance and the line-to-line capacitance per unit length. Hence the same clock frequency, i.e., the same interconnect delay, is maintained by reducing the chip size.

4.2.2 Case II: Wire-length Distribution for Optimization of both Chip Area and Interconnect Delay

As a second alternative, interconnect delay can be improved by increasing the wiring pitch, which causes a reduction in resistance and line-to-line capacitance per unit length. Two scenarios are considered: (a) global pitch is increased to match the global pitch for the 2-D case (b) global pitch is increased to match the chip area (footprint) for the 2-D case. Table 2 shows that performance can be increased by 64% for case (b) where vertical inter-layer interconnects are assumed to use up minimal area. Note that the delay requirement sets a maximum value of interconnect length on any given tier. Therefore, as interconnect lengths are increased, lines which exceeds this maximum length criterion for that particular tier need to be rerouted on upper tiers.

Moreover, an optimization similar to Case I can be carried out for different operating frequencies. Figure 6 illustrates how the optimum semi-global pitch increases for higher operating frequencies. However, as the semi-global tier pitch

PHYSICAL PARAMETER	VALUE
Number of Gates, N	180 million
Rent's Exponent, p	0.6
Rent's Coefficient, k	4.0
Minimum Feature Size, F	50 nm
Max number of wiring levels, n_{max}	9
Metal Resistivity, Copper	1.673×10^{-6} Ohm-cm
Dielectric Constant, Polymer	$\epsilon_r = 2.5$
Wiring Efficiency Factor	0.4

Table 1. Microprocessor example (NTRS '97 for 50 nm technology node).

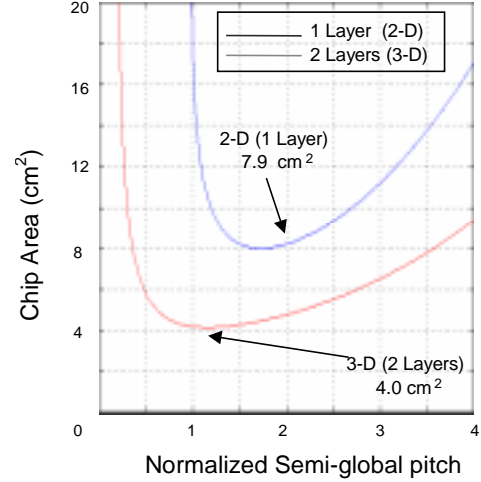


Figure 4. Wire-limited chip area versus normalized semi-global pitch (semi-global pitch/local tier pitch) for 2-D and 3-D ICs at a fixed operating frequency. As the normalized semi-global-pitch reduces, wires are rerouted to the global tiers, which have bigger pitch, and hence the chip area increases. Note that the 2-D area of 7.9 cm² compares well to 7.5 cm² projected by NTRS '97 for the 50 nm node. The estimated global pitches for the 2-D and 3-D ICs are 200 nm and 125 nm respectively. The number of metal layers per tier for 2-D and 3-D ICs is three.

2-Layer Description of Inter-layer Interconnects (ILICs)	Delay Performance Improvement
Horizontal ILICs, equal global pitch	16%
Horizontal ILICs, equal chip area	26%
Vertical ILICs, equal global pitch	49%
Vertical ILICs, equal chip area	64%

Table 2. Summary of delay performance improvement for 3-D ICs. The horizontal ILICs differ from the vertical ILICs in that they consume lateral area.

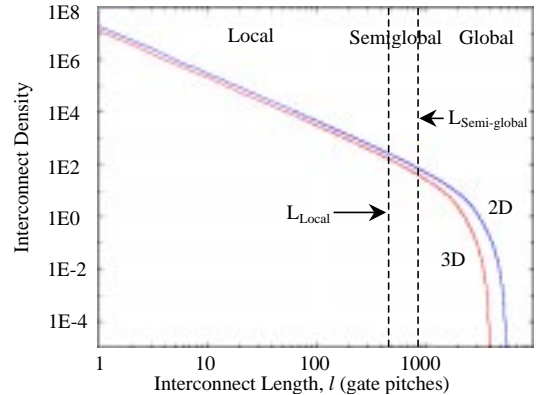


Figure 5. Wire-length distributions for the 2-D and 3-D ICs shown in Figure 4. 3-D significantly reduces requirement for longest wires. Metal tiers determined by L_{Loc} and $L_{Semi-global}$ boundaries.

increases, chip area and therefore, interconnect length also increases. This eventually saturates the reduction in the overall interconnect delay, and therefore, as shown in Figure 7, the clock frequency saturates. Eventually this leads to overcrowding of the global tier. Any further increases in the wiring density in the global tier forces a reduction in the global pitch as shown in Figure 8.

The analysis presented so far was for a 50 nm two Si layer 3-D technology where the number of metal layers was preserved. In the next two sections, we extend this analysis to study the effect of more than two Si layers and also the effect of increasing the number of available metal layers.

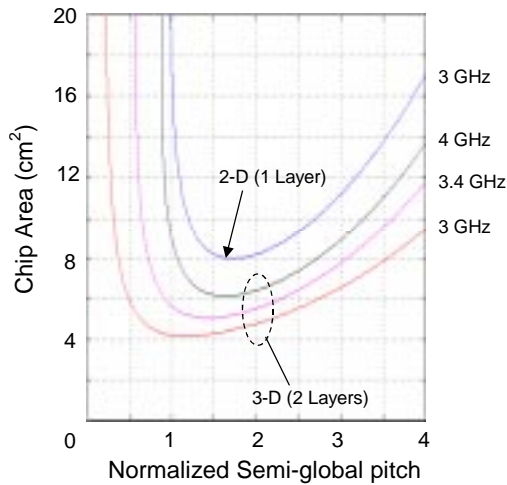


Figure 6. Chip operating frequency increases with increases in wiring pitch. Chip area also increases.

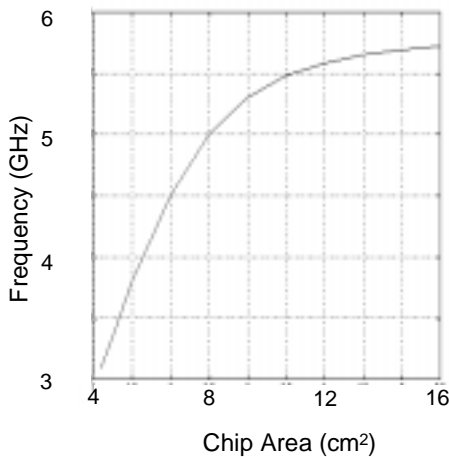


Figure 7. Performance improvement with increasing chip area for two-layer 3-D IC. Chip area is increased due to increasing wire pitch.

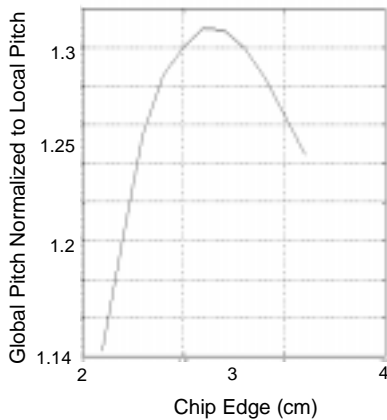


Figure 8. As the chip size increases due to increasing wire pitch, interconnects are re-routed to higher tiers. The global tier becomes over-crowded for large chip areas and global pitch starts to decrease.

4.3 M-Active Layer 3-D Circuit Performance

Technologies providing more than two active layers have also been considered. As the number of Silicon layers increases beyond two, the assumption that all ILICs are vertical and consume negligible area becomes less tenable. For this particular example it is assumed that 90% of all ILICs are horizontal (c.f. Table 2). The area used up by these horizontal ILICs can be estimated from their total length and pitch. As shown in Figure 9, the decrease in interconnect delay becomes progressively smaller as the number of active layers increases. This is due to the fact that area required by ILICs begins to offset any area saving due to increasing the number of active layers.

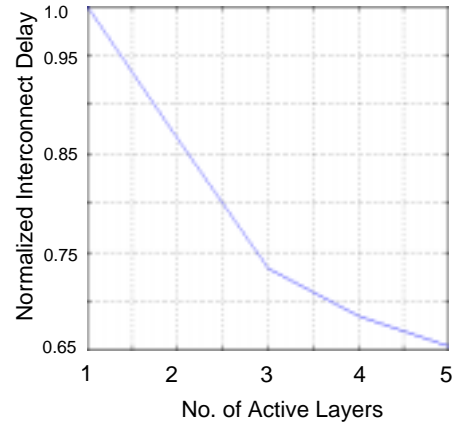


Figure 9. Signal Delay for multiple active Si layers normalized to single layer delay shown for 50 nm node. Note that ILICs here are assumed to consume lateral area.

4.4 Increasing Number of Metal Layers

In the above analyses, the total number of metal layers for 2-D and 3-D case was conserved. However, it is likely that there are local and semi-global tiers associated with every active layer, and a common global tier is used. This results in an increase in the total number of metal layers for the 3-D case. As shown in Figure 10, this causes a further 35% reduction in interconnect delay as compared to the original 3-D case with same total number of metal layers as in 2-D. Figure 10 also shows gate delay and interconnect delay for different technology nodes, both for 2-D and 3-D cases. It can be observed that for more aggressive technologies, the decrease in interconnect delay from 2-D to 3-D case is even more impressive.

4.5 Area Optimization by Rerouting Interconnects to Less Congested Tiers

In estimating chip area, the metal requirement is calculated from the obtained wire-length distribution. The total metallization requirement is appropriately divided among the available metal layers in the corresponding technology. Thus in the example shown in Fig. 4, each tier, the local, the semi-global and the global has three metal layers. The resulting area of the most densely packed tier, the local tier in this example, determines the chip area.

Consequently, higher tiers are routed within a larger than required area. An optimization for this scenario is possible by re-routing some of the local wires on the semi-global tier and the latter on the global, without violating the maximum allowable length per tier. This is achieved by reducing the maximum allowed interconnect length for the local and semi-global tiers (Figure 5) with varying fractions, w_1 and w_2 , respectively. Minimum chip area will be achieved when all the tiers are almost equally congested. The resulting calculations for chip area of the 2-D IC used in Figure 4 are shown in Figure 11. The 2-D chip area is seen to reduce by 9%. This wiring network optimization is also applied to 3-D ICs. The results are shown in Figure 12. The 3-D chip area is reduced by 11%.

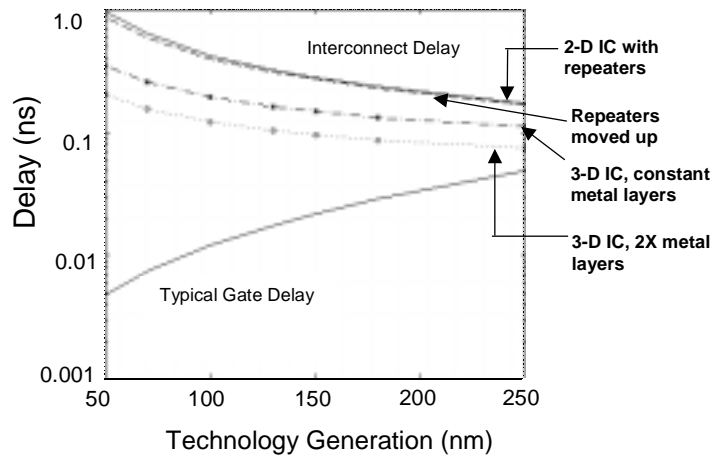


Figure 10. Interconnect delay limits IC performance with scaling. Moving repeaters to upper active tiers reduces interconnect delay by 9%. For the 50 nm node, 3-D (2 active layers) shows significant delay reduction (64%). Increasing the number of metal levels in 3-D reduces interconnect delay by a further 35%. This figure is based on the assumption that 3-D chip area equals 2-D chip area.

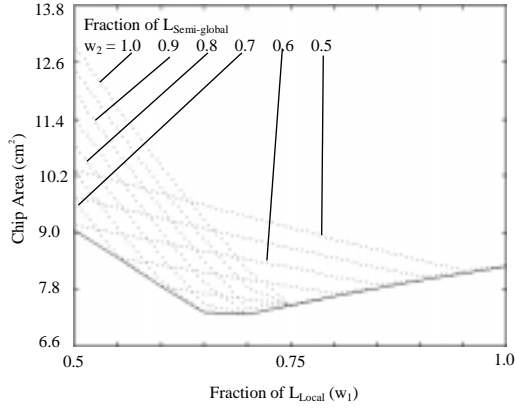


Figure 11. Chip area for 2-D IC with wiring network optimization. Solid line represents points of minimum area. (Based on NTRS '97 data for 50 nm node).

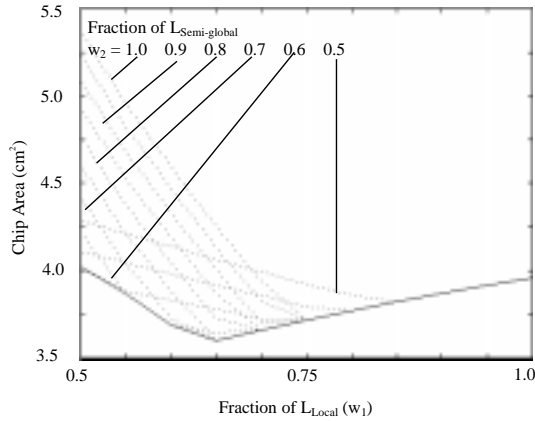


Figure 12. Chip area for 3-D ICs with wiring network optimization. Solid line represents points of minimum area. (Applied to NTRS '97 for 50 nm node).

5 Concerns in 3-D Circuits

5.1 Thermal Issues in 3-D Circuits

An extremely important issue in 3-D ICs is heat dissipation [60]. Thermal effects are already known to significantly impact interconnect and device reliability in present 2-D circuits [61]. The problem is expected to be exacerbated by the reduction in chip size, assuming that same power generated in a 2-D chip will now be generated in a smaller 3-D chip, resulting in a sharp increase in the power density. Analysis of thermal problems in 3-D circuits is therefore necessary to comprehend the limitations of this technology, and also to evaluate the thermal robustness of different 3-D technology and design options.

It is well known that most of the heat energy generated in integrated circuits arises due to transistor switching. This heat is typically conducted through the silicon substrate to the package, and then to the ambient by a heat sink (Figure 13). With multi-layer device designs, devices in the upper layers will also generate a significant fraction of the heat. Furthermore, all the active layers will be insulated from each other by layers of dielectrics (LTO, HSQ, polyimide etc.) which typically have much lower thermal conductivity than Si [62, 63]. Hence, the heat dissipation issue can become even more acute for 3-D ICs and can cause degradation in device performance, and reduction in chip reliability due to increased junction leakage, electromigration failures, and by accelerating other failure mechanisms [61, 64].

In this work, a specific example of using two active silicon layers is considered, and two possible configurations depending on the processing technology (shown in Figure 14) have been analyzed. The principal difference between the two configurations is that the area through which heat can dissipate is twice for configuration (b). Figure 15 shows the total power dissipation (P) and chip area (A) for a high-end microprocessor for various 2-D technology nodes based on NTRS '97 data [1]. It can be observed that as technology scaling continues, chip area and power dissipation is increasing.

The relationship between the die temperature rise (ΔT_{Die}) and P can be expressed as,

$$\Delta T_{Die} = (T_{Die} - T_0) = P \cdot R_{\theta} \quad (4)$$

where T_0 is the ambient temperature ($\approx 25^\circ\text{C}$), and R_{θ} is the effective thermal resistance from the Si to the heat-sink. Neglecting interface resistances, R_{θ} can be expressed as,

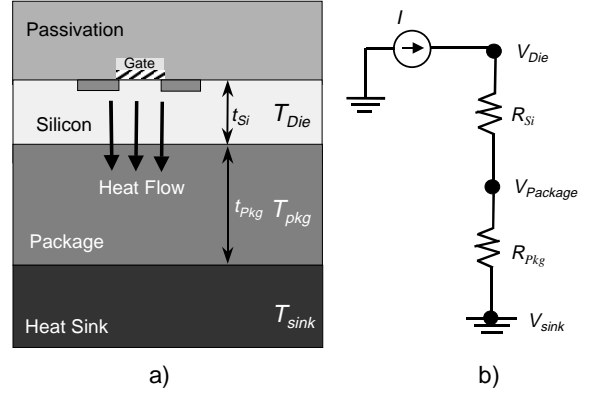


Figure 13. Schematic view of a) heat flow in 2-D circuits and b) equivalent thermal circuit. T denotes temperature of different materials. R_{Si} and R_{pkg} are the thermal resistances of the Si and the package material respectively.

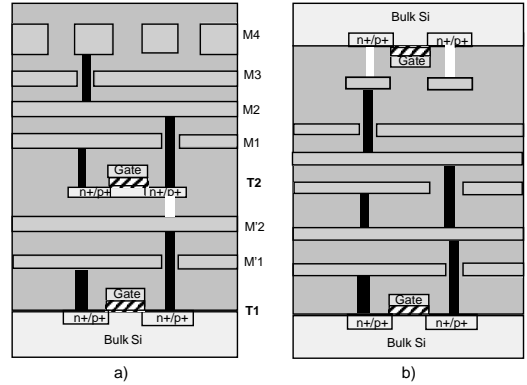


Figure 14. 3-D circuits with two active layers depicting two different processing scenarios. a) TFT or epitaxially grown second layer type technology with two metal layers per layer. M3 and M4 levels are common for both active layers (global wires). b) Wafer bonding type technology with three metal layers per active layer. Note that the metal layers for the second active layer are now upside down and the second layer Si can easily dissipate heat.

$$R_{\theta} = \left(\frac{t_{Si}}{K_{Si}} + \frac{t_{pkg}}{K_{pkg}} \right) \frac{I}{A} \quad (5)$$

Here t_{Si} and K_{Si} are the thickness and the thermal conductivity of the Si substrate, and t_{pkg} and K_{pkg} denote same parameters for the packaging material. A is the effective area through which heat flow takes place. Since the die size (length) is much larger than the thickness of Si, we assume one dimensional heat flow. Hence, from equation (4) and (5), it follows that,

$$\Delta T_{Die} = R_n \frac{P}{A} \quad (6)$$

where R_n is the normalized thermal resistance. Since the typical die temperature for present high-performance 2-D circuits (250 nm technology node) is known to be $\sim 110^\circ\text{C}$, the value of R_n can be calculated. Based on this value of R_n , we have calculated die temperatures for other 2-D technology nodes (based on NTRS '97 data) from Figure 15.

Now, for the 3-D circuits shown in Figure 14, we can calculate the power dissipation from the following equation,

$$P = \frac{1}{2} \alpha C V_{dd}^2 f \quad (7)$$

Here the interconnect dominated capacitance, C , and the chip frequency, f , are determined from our model and the product (αV_{dd}^2), is calculated from the corresponding 2-D technology nodes for consistency. The increase in die temperature (ΔT_{Die}) for the 3-D circuits can then be determined from equation (6), since 3-D chip area can be calculated from our model, and R_n remains constant for both 2-D and 3-D circuits assuming same packaging material for all the technology nodes.

Estimated 3-D die temperature is plotted as a function of technology node in Figure 16 by assuming the same packaging material for all the technology nodes. For case (a) of 3-D circuits, heat is assumed to dissipate from the lower Si substrate only, while for case (b) heat can dissipate from both the Si substrates. It can be observed that the thermal problem is significantly less in case (b) than in

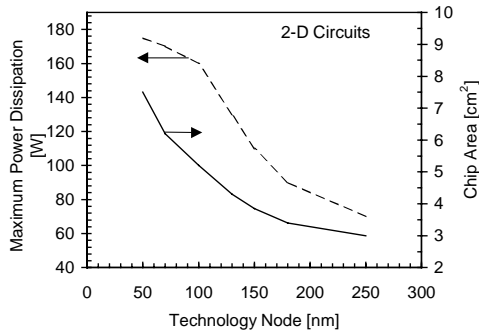


Figure 15. Maximum power dissipation and chip area in 2-D circuits as a function of technology node.

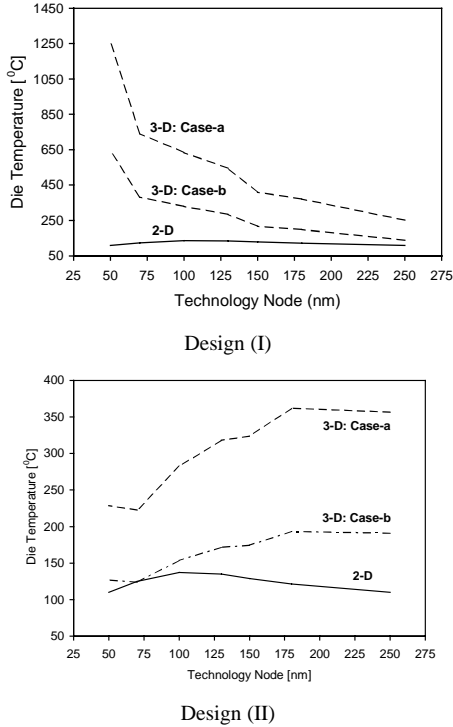


Figure 16. Die temperature as function of technology node for 2-D and 3-D circuits assuming, Design (I): equal chip area for 2-D and 3-D ICs, and Design (II): equal metal wire pitch for 2-D and 3-D ICs. Design (II) indicates that thermal problems can be alleviated in 3-D ICs provided thermal issues are considered at an early design phase.

case (a) of the 3-D circuits due to the presence of an additional heat-dissipating surface.

It should be pointed out that design (I) is desirable for higher performance (see Figure 6). However, it results in higher power dissipation giving rise to higher die temperatures as shown in Fig. 16. This is expected since same chip area between 2-D and 3-D in design (I) is achieved by increasing the metal cross-sectional area for 3-D, which reduces the line resistance (R) and hence increases the operating frequency. However, since, $P \propto Cf \propto 1/R$, the power dissipation increases, resulting in higher die temperatures.

Note that in all calculations Joule heating of the interconnects has been ignored since most of the heat is dissipated by the transistors. However, interconnect Joule heating can increase for 3-D case (a) due to the presence of a second active layer that remains isolated from the substrate, and can cause strong thermal coupling with the neighboring interconnects giving rise to higher interconnect temperatures and hence higher interconnect resistance and also lower interconnect electromigration performance [61,64]. Furthermore, in our die temperature calculations for 3-D case (a), we assumed identical temperatures for devices on the two active layers. However, the temperature of the second active layer in case (a) which is embedded between insulators, will be higher than the one at the substrate, where heat can dissipate more easily through the packaging and the heat sink. Hence, the average die temperatures calculated for case (a) would actually be even higher than what is shown in Figure 16.

It should be noted that advances in heat-sinking technologies would benefit both 2-D and 3-D ICs. In the thermal analysis presented above, the normalized

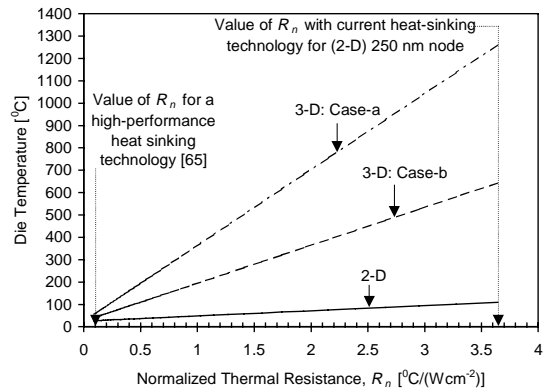


Figure 17. Attainable die temperatures for 2-D and 3-D ICs at the NTRS based 50 nm node for design (I) using advanced heat-sinking technologies that would reduce the normalized thermal resistance, R_n .

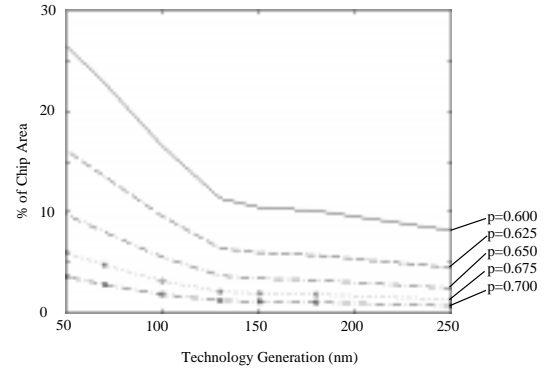


Figure 18. Fraction of chip area used by repeaters for different technology nodes and different Rent's exponents. As much as 27% of the chip area at 50 nm node is likely to be occupied by repeaters.

thermal resistance used for all the technology nodes was based on the current heat-sinking technology for 250 nm 2-D circuits. Advancements in the heat-sinking technology would significantly help lower the chip operating temperatures as shown in Figure 17, where the die temperatures for 2-D and 3-D circuits have been estimated at the 50 nm technology node for design (I). It can be observed that employing a cooling design similar to the one suggested in [65] can significantly alleviate the thermal problem in all cases.

5.2 Interconnect Capacitance and Cross Talk

In 3-D devices an additional coupling between the top layer metal of the first active layer and the devices on the second active layer would be present [66]. This needs to be addressed at the circuit design stage. However, for deep submicron technologies, the aspect ratio of interconnects is approximately 1.5 to 2. Therefore line-to-line capacitance is the dominant portion of the overall capacitance. Therefore, the presence of an additional Silicon layer on top of a metal level will not have an appreciable effect on the line capacitance per unit length. For technologies with very small aspect ratio, the change in interconnect capacitance due to the presence of an additional Silicon layer would be significant, as reported in [66].

6 Implications on Circuit Design and Architecture

6.1 Buffer Insertion

For deep submicron technologies, interconnect delay is the dominant component of the overall delay, especially for circuits with very long interconnects where the delay can become quadratic with line lengths. To overcome this problem, long interconnects are typically broken into shorter buffered segments. In [67] it was shown that for point-to-point interconnects, there exists an optimum interconnect length and an optimum buffer size for which the overall delay is minimum. Buffer sizes for various metal layers for different technologies have been presented in [64,67]. For top layer interconnect, the corresponding inverter sizes were approximately 450 times the minimum inverter size available in the relevant technology. These large buffers present a problem since they take up a lot of active silicon and routing area. The vias that connect such a buffer from the top global interconnect layers to the inverters block all the metal layers present underneath them, hence taking up substantial routing area. It has been predicted [68] the number of such buffers can reach 10,000 for high

performance designs in 100 nm technology. Inserting these buffers will cause unacceptable area increase as shown in Figure 18.

However, this problem can be easily tackled using 3-D technology with just two Silicon layers. The repeaters can be placed on the second Silicon layer thereby saving area on the first Silicon layer. Furthermore, if the second Silicon layer is placed close to the common global metal layers, the vias connecting the global metal layers to the repeaters will not block the lower metal layers thereby freeing up additional routing area.

Previously Figure 10 had also included delay simulation results for an otherwise single active layer IC except that the repeaters had been moved to a second active layer. A conservative value of Rent's exponent ($p=0.65$) was used to estimate the reduction in chip area and therefore reduction in overall interconnect delay. At 50 nm node, an additional reduction of 9% in the overall interconnect delay results from the resulting area reduction.

6.2 Layout of Critical Paths

In typical high performance ASIC and microprocessor designs, interconnect delay is a significant portion of the overall path delay [69]. Logic blocks on a critical path need to communicate to other logic blocks which, due to placement and other design constraints, may be placed far away from each other. The delay in the long interconnects between such blocks usually causes timing violations. With the availability of a second active layer, these logic blocks can be placed on different Silicon layers and hence can be very close to each other, thereby minimizing interconnect delay. Even if TFT devices are used on the second active layer, the decrease in interconnect delay can be more than the increase in gate delay due to inferior transistor characteristics.

6.3 Microprocessor Design

In microprocessors and DSP processors, most of the critical paths involve on-chip caches [70]. The primary reason for this that on-chip cache is (physically) located in one corner of the die whereas the logic and computational blocks, which access this memory, are distributed all over the die. By using a technology with two Silicon layers, the caches can be placed on the second active layer and the logic and computational blocks on the first layer. This arrangement ensures that logic blocks are in closer proximity to on-chip caches.

Consider a microprocessor of dimensions $L \times L$. In typical current generation microprocessors, about half the physical area is taken up by on-chip caches. Hence the worst case interconnect length in a critical path is $2L$ (see footnote 1 below). If on-chip caches are placed on the second active layer and the chip is resized accordingly to have dimensions $\frac{L}{\sqrt{2}} \times \frac{L}{\sqrt{2}}$, then the worst case

interconnect length is $\sqrt{2}L$, a reduction of about 30%. Even though this analysis is very simplistic compared to the more elaborate one presented in Section 4, and does not perform any optimization of the interconnect pitch, it demonstrates that going from single silicon layer to two layers results in nontrivial improvement in performance. Recent studies [33] have shown that by integrating level one and level two cache and the main memory on the same Silicon using 3-D technology, access times for level 2 cache and main memory can be decreased. This coupled with an increase in bandwidth between the memory, level 2 cache and level 1 cache, reduces the level 2 cache/memory miss penalty and therefore reduces average time per instruction and increases system performance.

6.4 Mixed Signal Integrated Circuits

With greater emphasis on increasing the functionality that can be implemented on a single die in the system-on-a-chip paradigm, more and more analog, mixed-signal and RF components of the system are being integrated on the same piece of Silicon. However, this presents serious design issues since switching signals from the digital portions of the chip couple into the sensitive analog and RF circuit nodes from the substrate and degrade the fidelity (or equivalently, increase the noise) of the signals present in these blocks [71]. However, with the availability of multiple Silicon layers, RF and mixed signal portions of the system can be realized on a separate layer thereby providing substrate isolation from the digital portion. A preliminary analysis shows a 30 dB improvement in isolation by moving the RF portions of the circuit to a separate substrate. Moreover, since the second Si layer is not continuous, good isolation between different analog and RF components (such as LNA and power amplifier) can also be achieved.

6.5 Implications on Physical Design and Synthesis

At an abstract level, physical design (placement and routing) can be viewed as a graph embedding problem. The circuit graph (synthesized and mapped circuit) is embedded on a target graph which is planar (which corresponds to the physical substrate of the conventional single Silicon substrate technology). However, with more than one Silicon layer available, the target graph is no longer planar, and therefore placement and routing algorithms need to be suitably modified.

¹Typically the data transfer from cache takes more than one clock cycles but we assume single clock cycle transfers for simplicity.

Moreover, since placement and routing information also affects synthesis algorithms, which in turn can affect the choice of architectures, this modification needs to be propagated all the way to synthesis and architectural level.

7 Conclusions

In this paper we have motivated the discussion of 3-D IC technologies with multiple Silicon layers, as a promising alternative to the present single Si layer technologies, to alleviate the ever increasing problem of interconnect delays in high performance logic circuits. An overview of some of the processing techniques under investigation, which can be used to fabricate these circuits, was provided.

A methodology was presented to obtain the wire length distributions for 3-D ICs which can be used to accurately predict area, delay, and power dissipation, and provide examples of some of these trade-offs which result in area and/or delay reduction over the 2-D case. Our analysis predicts significant performance improvements of up to 145% over the 2-D case. The primary target technology for this analysis has been the NTRS '97 based 50 nm node with two active layers of Silicon. Other technology nodes with more than two active layers have also been considered. It was shown that the availability of additional Silicon layers gives extra flexibility to designers which can be exploited to minimize area, improve performance and power dissipation or any combinations of these.

Furthermore, we also addressed the thermal concerns for 3-D circuits. A simple thermal model for heat dissipation in 3-D ICs was presented. This allows us to analyze thermal effects in these circuits in detail. From the estimated values of the total wiring capacitance, frequency of operation, and chip area, the total power dissipation was calculated and then used in conjunction with our thermal model to estimate die temperatures for two different 3-D circuit design options at various NTRS technology nodes. It was demonstrated that for circuits with two Silicon layers, acceptable die temperatures can be maintained by either employing a thermally responsible circuit design or by using advanced heat-sinking technologies that would lower the thermal resistance between the chip and the heat sink.

Finally, we highlighted some scenarios in current and future VLSI and mixed signal systems, where the use of 3-D circuits will have an immediate and beneficial impact on performance. We believe that 3-D ICs could definitely prove to be beneficial for applications that require integration of disparate technologies, and for system-on-chip (SOC) designs. We also briefly discussed the long-term implications of using this technology, as conventional physical design, gate level and architecture level synthesis algorithms need to be suitably adapted.

Acknowledgements

This work was supported by the DARPA AME Program and the MARCO Interconnect Technology Focus Center at Stanford University.

8 References

- [1] The National Technology Roadmap for Semiconductors, *Technology Needs*, 1997.
- [2] C. R. Barrett, "Microprocessor evolution and technology impact," *Symp. VLSI Technol., Digest*, 1993, pp. 7-10.
- [3] C. Hu, "MOSFET scaling in the next decade and beyond," *Semiconductor International*, pp. 105-114, 1994.
- [4] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high performance and low power-The next ten years," *Proc. of the IEEE*, vol. 83, no. 4, pp. 595-606, 1995.
- [5] G. A. Sai-Halasz, "Performance trends in high-end processors," *Proc. of the IEEE*, vol. 83, no. 1, pp. 20-36, 1995.
- [6] M. T. Bohr, "Interconnect scaling-the real limiter to high performance ULSI," *IEDM Tech. Dig.*, 1995, pp. 241-244.
- [7] J. D. Meindl, "Low power microelectronics: retrospect and prospect," *Proc. of the IEEE*, vol. 83, no. 4, pp. 619-635, 1995.
- [8] S-Y Oh and K-J Chang, "2001 needs for multi-level interconnect technology," *Circuits and Devices*, pp. 16-21, 1995.
- [9] M. T. Bohr and Y. A. El-Mansy, "Technology for advanced high-performance microprocessors," *IEEE Trans. Electron Devices*, vol. 45, no. 3, pp. 620-625, 1998.
- [10] D. Edelstein et al., "Full copper wiring in a sub-0.25 μm CMOS ULSI technology," *IEDM Tech. Dig.*, 1997, pp. 773-776.
- [11] S. Venkatesan et al., "A high performance 1.8V, 0.20 μm CMOS technology with copper metallization," *IEDM Tech. Dig.*, 1997, pp. 769-772.
- [12] E. M. Zielinski et al., "Damascene integration of copper and ultra-low-k xerogel for high performance interconnects," *IEDM Tech. Dig.*, 1997, pp. 936-938.
- [13] N. Rohrer et al., "A 480MHz RISC microprocessor in a 0.12 μm L_{eff} CMOS technology with copper interconnects," *Int. Solid-State Circuits Conf., Tech. Digest*, 1998, pp. 240-241.
- [14] B. Zhao et al., "A Cu/low-k dual damascene interconnect for high performance and low cost integrated circuits," *Symp. VLSI Technology, Tech. Digest*, 1998, pp. 28-29.

- [15] W. J. Dally, "Interconnect-limited VLSI architecture," *Int. Interconnect Technology Conf. Proceedings*, 1999, pp. 15-17.
- [16] M. W. Geis, D. C. Flanders, D. A. Antoniadis, and H. I. Smith, "Crystalline silicon on insulators by graphoepitaxy," *IEDM Tech. Dig.*, 1979, pp. 210-212.
- [17] J. P. Colinge and E. Demoulin, "ST-CMOS (Stacked Transistor CMOS): a double-poly-NMOS-compatible CMOS technology," *IEDM Tech. Dig.*, 1981, pp. 557-560.
- [18] G. T. Goeloe, E. W. Maby, D. J. Silversmith, R. W. Mountain, and D. A. Antoniadis, "Vertical single-gate CMOS inverters on laser-processed multilayer substrates," *IEDM Tech. Dig.*, 1981, pp. 554-556.
- [19] S. Kawamura, N. Sasaki, T. Iwai, M. Nakano, and M. Takagi, "Three-dimensional CMOS IC's fabricated by using beam recrystallization," *IEEE Electron Device Lett.*, vol. EDL-4, no. 10, pp. 366-368, 1983.
- [20] S. Akiyama, S. Ogawa, M. Yoneda, N. Yoshii, and Y. Terui, "Multilayer CMOS device fabricated on laser recrystallized silicon islands," *IEDM Tech. Dig.*, 1983, pp. 352-355.
- [21] M. Nakano, "3-D SOI/CMOS," *IEDM Tech. Dig.*, 1984, pp. 792-795.
- [22] K. Sugahara, T. Nishimura, S. Kusunoki, Y. Akasaka, and H. Nakata, "SOI/SOI/Bulk-Si triple level structure for three-dimensional devices," *IEEE Electron Device Lett.*, vol. EDL-7, no. 3, pp. 193-195, 1986.
- [23] Y. Akasaka and T. Nishimura, "Concept and basic technologies for 3-D IC structure," *IEDM Tech. Dig.*, 1986, pp. 488-491.
- [24] S. Tatsuno, "Japan's push into creative semiconductor research: 3-dimension IC's," *Solid State Technology*, March 30, pp. 29-30, 1987.
- [25] T. Nishimura, Y. Inoue, K. Sugahara, S. Kusunoki, T. Kumamoto, S. Nakagawa, M. Nakaya, Y. Horiba, and Y. Akasaka, "Three dimensional IC for high performance image signal processor," *IEDM Tech. Dig.*, 1987, pp. 111-114.
- [26] T. Kunio, K. Oyama, Y. Hayashi, and M. Morimoto, "Three dimensional ICs, having four stacked active device layers," *IEDM Tech. Dig.*, 1989, pp. 837-840.
- [27] S. Strickland, et al., "VLSI design in the 3rd dimension," *INTEGRATION*, Elsevier Science, pp. 1-16, 1998.
- [28] D. Antoniadis, "3-dimensional 25 nm - scale CMOS technology," *Advanced Microelectronics Program Review Proceedings Book*, Sept. 1-2, Lexington, MA, 1998.
- [29] V. Subramanian and K. C. Saraswat, "High-performance germanium-seeded laterally crystallized TFT's for vertical device integration," *IEEE Trans. Electron Devices*, vol. 45, no. 9, pp. 1934-1939, 1998.
- [30] G. W. Neudeck, S. Pae, J. P. Denton, and T. Su, "Multiple layers of silicon-on-insulator for nanostructure devices," *J. Vac. Sci. Technol. B* 17(3), pp. 994-998, 1999.
- [31] K. C. Saraswat, S. J. Souri, V. Subramanian, A. R. Joshi, and A. W. Wang, "Novel 3-D Structures," *IEEE Int. SOI Conf.*, 1999, pp. 54-55.
- [32] S. A. Kuhn, M. B. Kleiner, P. Ramm, and W. Weber, "Performance modeling of the interconnect structure of a three-dimensional integrated RISC processor/cache system," *IEEE Trans. Components, Packaging, and Manufacturing Technology-Part B*, vol. 19, no. 4, pp. 719-727, 1996.
- [33] M. B. Kleiner, S. A. Kuhn, P. Ramm, and W. Weber, "Performance improvement of the memory hierarchy of RISC-systems by application of 3-D technology," *IEEE Trans. Components, Packaging, and Manufacturing Technology-Part B*, vol. 19, no. 4, pp. 709-718, 1996.
- [34] S. J. Souri and K. C. Saraswat, "Interconnect performance modeling for 3D integrated circuits with multiple Si layers," *Int. Interconnect Technology Conf. Proceedings*, 1999, pp. 24-26.
- [35] A. Rahman, A. Fan, J. Chung, and R. Reif, "Wire-length distribution of three-dimensional integrated circuits," *Int. Interconnect Technology Conf. Proceedings*, 1999, pp. 233-235.
- [36] R. Zhang, K. Roy, and D. B. Jones, "Architecture and performance of 3-dimensional SOI circuits," *IEEE Int. SOI Conf.*, 1999, pp. 44-45.
- [37] A. Kohno, T. Sameshima, N. Sano, M. Sekiya, and M. Hara, "High performance poly-Si TFTs fabricated using pulsed laser annealing and remote plasma CVD with low temperature processing," *IEEE Trans. Electron Devices*, vol. 42, no. 2, pp. 251-257, 1995.
- [38] M. A. Crowder, P. G. Carey, P. M. Smith, R. S. Sposili, H. S. Cho, and J. S. Im, "Low-temperature single crystal Si TFT's fabricated on Si-films processed via sequential lateral solidification," *IEEE Electron Device Lett.*, vol. 19, no. 8, pp. 306-308, 1986.
- [39] H-Y. Lin, C-Y. Chang, T. F. Lei, J-Y. Cheng, H-C. Tseng, and L-P. Chen, "Characterization of polycrystalline silicon thin film transistors fabricated by ultrahigh-vacuum chemical vapor deposition and chemical mechanical polishing," *Jpn. J. Appl. Phys., Part 1*, vol.36, (no.7A), pp. 4278-4282, July 1997.
- [40] T. Noguchi, "Appearance of single-crystalline properties in fine-patterned Si thin film transistors (TFTs) by solid phase crystallization (SPC)," *Jpn. J. Appl. Phys., Part 2*, no.11A, vol.32, pp. 1584-1587, Nov. 1993.
- [41] T. W. Little, H. Koike, K. Takahara, T. Nakazawa, and H. Oshima, "A 9.5-in. 1.3-Mpixel low-temperature poly-Si TFT-LCD fabricated by solid-phase crystallization of very thin films and an ECR-CVD gate insulator," *J. Society for Information Display*, 1/2, pp. 203-209, 1993.
- [42] N. Yamauchi, "Polycrystalline silicon thin films processed with silicon ion implantation and subsequent solid-phase crystallization: theory, experiments, and thin-film transistor applications," *J. Appl. Phys.*, 75(7), pp. 3235-3257, 1994.
- [43] D. N. Kouvatso, A. T. Voutsas, and M. K. Hatalis, "Polycrystalline silicon thin film transistors fabricated in various solid phase crystallized films deposited on glass substrates," *J. Electronic Materials*, vol. 28, no. 1, pp. 19-25, 1999.
- [44] J. A. Tsai, A. J. Tang, T. Noguchi, and R. Reif, "Effects of Ge on material and electrical properties of polycrystalline Si_{1-x}Ge_x for thin film transistors," *J. Electrochem. Soc.*, vol. 142, no. 9, pp. 3220-3225, 1995.
- [45] S-W. Lee and S-K. Joo, "Low temperature poly-Si thin film transistor fabrication by metal-induced lateral crystallization," *IEEE Electron Device Lett.*, vol. 17, no. 4, pp. 160-162, 1983.
- [46] S. Y. Yoon, S. K. Kim, J. Y. Oh, Y. J. Choi, W. S. Shon, C. O. Kim, and J. Jang, "A high-performance polycrystalline silicon thin-film transistor using metal-induced crystallization with Ni solution," *Jpn. J. Appl. Phys., Part 1*, pp. 7193-7197, Dec. 1998.
- [47] A. R. Joshi and K. C. Saraswat, "Sub-micron thin film transistors with metal induced lateral crystallization," Abstract no. 1358, *Proc. 196th Meeting of the Electrochemical Society*, Honolulu, HI, 1999.
- [48] J. Nakata and K. Kajiyama, "Novel low-temperature recrystallization of amorphous silicon by high energy beam," *Appl. Phys. Lett.*, pp. 686-688, 1982.
- [49] Y. W. Choi, J. N. Lee, T. W. Jang, and B. T. Ahn, "Thin-film transistors fabricated with poly-Si films crystallized at low temperature by microwave annealing," *IEEE Electron Device Lett.*, vol. 20, no. 1, pp. 2-4, 1999.
- [50] A. Heya, A. Masuda, and H. Matsumura, "Low-temperature crystallization of amorphous silicon using atomic hydrogen generated by catalytic reaction on heated tungsten," *Appl. Phys. Lett.*, vol. 74, no. 15, pp. 2143-2145, 1999.
- [51] R. K. Watts and J. T. C. Lee, "Tenth-micron polysilicon thin-film transistors," *IEEE Electron Device Lett.*, vol. 14, no. 11, pp. 515-517, 1993.
- [52] M. Rodder and S. Aur, "Utilization of plasma hydrogenation in stacked SRAMs with poly-Si PMOSFETs and bulk Si NMOSFETs," *IEEE Electron Device Lett.*, vol. 12, no. 5, pp. 233-235, 1991.
- [53] T. Yamanaka et al., "Advanced TFT SRAM cell technology using a phase-shift lithography," *IEEE Trans. Electron Devices*, vol. 42, no. 7, pp. 1305-1312, 1995.
- [54] M. Cao, T. Zhao, K. C. Saraswat, and J. D. Plummer, "A simple EEPROM cell using twin polysilicon thin film transistor," *IEEE Electron Device Lett.*, vol. 15, no. 8, pp. 304-306, 1994.
- [55] P. Ramm et al., "Three dimensional metallization for vertically integrated circuits," *Microelectronic Engineering*, 37/38 pp. 39-47, 1997.
- [56] J. A. Davis, V. K. De, and J. D. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI) - Part II: Applications to clock frequency, power dissipation, and chip size estimation," *IEEE Trans. Electron Devices*, Vol. 45, no. 3, March 1998.
- [57] L. Robinson, L. A. Glasser, and D. A. Antoniadis, "A simple interconnect delay model for multilayer integrated circuits," *IEEE VMIC Conf.*, 1986.
- [58] B. S. Landman, and R. L. Russo, "On a pin versus block relationship for partitions of logic graphs," *IEEE Trans. Computers*, vol. C-20, no. 12, Dec. 1971.
- [59] J. A. Davis, V. K. De, and J. D. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI) - Part I: Derivation and validation," *IEEE Trans. Electron Devices*, Vol. 45, no. 3, March 1998.
- [60] S. A. Kuhn, M. B. Kleiner, P. Ramm, and W. Weber, "Thermal analysis of vertically integrated circuits," *IEDM Tech. Dig.*, 1995, pp. 487-490.
- [61] K. Banerjee, "Thermal effects in deep submicron VLSI interconnects," Tutorial Notes, *IEEE International Symposium on Quality Electronic Design*, March 20-22, 2000.
- [62] K. E. Goodson and Y. S. Ju, "Heat conduction in novel electronic films," *Annu. Rev. Mater. Sci.*, 29, pp. 261-293, 1999.
- [63] K. Banerjee, A. Amerasekera, G. Dixit, and C. Hu, "The effect of interconnect scaling and low-k dielectric on the thermal characteristics of the IC metal," *IEDM Tech. Dig.*, 1996, pp. 65-68.
- [64] K. Banerjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C. Hu, "On thermal effects in deep sub-micron VLSI interconnects," *Proc. 36th ACM Design Automation Conference*, June 1999, pp. 885-891.
- [65] D. B. Tuckerman, R. F. W. Pease, "High-performance heat sinking for VLSI," *IEEE Electron Device Lett.*, vol. EDL-2, no.5, pp.126-129, 1981.
- [66] S. A. Kuhn, M. B. Kleiner, P. Ramm, and W. Weber, "Interconnect capacitances, crosstalk, and signal delay in vertically integrated circuits," *IEDM Tech. Dig.*, 1995, pp. 487-490.
- [67] R. H.J.M. Otten and R. K. Brayton, "Planning for performance," *Proc. 35th Annual Design Automation Conference*, 1998, pp. 122-127.
- [68] J. Cong and L. He, "An efficient technique for device and interconnect optimization in deep submicron designs," *Int. Symp. on Physical Design*, 1998, pp. 45-51.
- [69] P. D. Fisher, "Clock cycle estimation for future microprocessor generations," *Technical Report*, SEMATECH 1997.
- [70] D. Greenhill et al, "A 330 MHz 4-way superscalar microprocessor," *ISSCC, Digest of Tech. Papers*, 1997, pp. 166-167.
- [71] B. Razavi, "Challenges and Trends in RF Design," *Proc. 9th Annual IEEE Int. ASIC Conf. and Exhibit*, 1996, pp. 81-86.