# 3D-ICs created using oblique processing

D. Bruce Burckel
Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM, USA 87106

## ABSTRACT

This paper demonstrates that another class of three-dimensional integrated circuits (3D-ICs) exists, distinct from through silicon via centric and monolithic 3D-ICs. Furthermore, it is possible to create devices that are 3D *at the device level* (i.e. with active channels oriented in each of the three coordinate axes), by performing standard CMOS fabrication operations at an angle with respect to the wafer surface into high aspect ratio silicon substrates using membrane projection lithography (MPL). MPL requires only minimal fixturing changes to standard CMOS equipment, and no change to current state-of-the-art lithography. Eliminating the constraint of 2D planar device architecture enables a wide range of new interconnect topologies which could help reduce interconnect resistance/capacitance, and potentially improve performance.

**Keywords:** 3D-ICs, interconnects, oblique processing

## 1. INTRODUCTION

Moore's law [1], an observation that the cost per transistor decreases as transistor density increases, following a roughly 2 year doubling period, has dominated the landscape of semiconductor research since it was first proposed in 1965. For much of this run, Moore's law was supported by Dennard scaling [2] which posits that processing performance per Watt increases with decreasing device dimension, which also possesses a similar doubling period. Thus the reduced cost per transistor was accompanied by higher performance per Watt, a win-win proposition which served as a self-sustaining feedback mechanism, responsible for today's massive semiconductor and personal electronics industries among other epochal changes. This coincidental scaling of cost, performance and power consumption allows these trends to be conveniently plotted on familiar log-linear graphs showing the breathtaking ascent and, more recently, inevitable stall of these curves as the semiconductor industry has eclipsed the 28 nm processing node.

Transistors with smaller device dimensions are subject to a variety of deleterious effects such as drain induced barrier lowering (DIBL), gate leakage, subthreshold leakage, etc. responsible for the end of Dennard scaling. Even though subsequent lithography nodes add higher transistor density, the failure of these new device designs to scale power requirements means that not all of these transistors can be concurrently active for the same power budget, leading to the notion of dark silicon, a problem which scales non-linearly with shrinking dimensions. In addition to these deterministic effects, small transistors also incur increased variance of the statistical distribution of device performance. Increasing variability in performance from device to device complicates the already enormous task of designing circuits and multi-circuit modules. [3]

Pursuit of Moore's law has required more than just smaller transistors. Adoption of increasingly complicated interconnection strategies has also been necessary in order to allow the dense sea of transistors to communicate with one another. For some time it has been recognized that gate delay and transistor capacitance is only a fraction of the overall delay in switching speed.[4] Interconnect capacitance has dominated gate delay, forcing the adoption of copper damascene and pursuit of low-k dielectrics. The culmination of this trend is the inclusion of air-gaps to further reduce capacitance by filling the space between interconnects with a material with the lowest possible dielectric constant at the cost of manufacturing complexity, yield, and potential reliability concerns.

IP Bridge Exhibit 2010
TSMC v. IP Bridge
IPR2016-01247



DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

Modern integrated circuits are highly optimized to maximize performance per unit area. Shrinking device dimensions and areal scaling are achieved through process node improvements, fueling much of this optimization, however, design process technology co-optimization, the practice of using process-aware design of circuit layout has become increasingly important[5]. Adoption of Manhattan layout geometry and self-aligned double/quadruple patterning (SADP, SAQP) are examples where full 2-dimensional interconnect spatial freedom are sacrificed in order to enable advanced lithography techniques. Optimized module designs are constructed with the optimized transistors. In order to connect these modules together, sophisticated place and route algorithms are used to construct larger computational blocks.

These increasingly disruptive trends have forced the semiconductor industry to adopt new materials, process techniques and device design concepts, with even more drastic changes required for future process nodes.[6] Adoption of a 3D integrated circuit topology is seen by many in the industry as a viable approach to continue density scaling. The shorter average interconnect lengths of 3D-ICs [7] reduce both the resistance and capacitance of the line, and hence reduce ohmic power dissipation and RC time delay. While just on the verge of being adopted as a high volume manufacturing (HVM) solution, the concept and advantages of 3D-ICs have been identified since at least the 1980's [8]. In its current incarnation, 3D-IC architectures are divided into two distinct types: 1) 3D-ICs created by stacking planar 2D chips [9], achieving interconnection in the vertical direction using through silicon vias (TSV); and 2) monolithic 3D-ICs where epitaxial regrowth is performed on the wafer after fabrication of the first layer of devices, yielding a second layer of single crystal silicon for a second layer of silicon devices [10]. Both of these approaches yield ICs with current flow vertical to the wafer surface, and, in that sense are 3-dimensional, however the transistors forming each integrated circuit all have their active regions oriented parallel to the wafer surface.

The purpose of this paper is to point out that another class of 3D-ICs exists with transistors oriented along each of the coordinate axes. Distinct from both TSV-3D-ICs and monolithic 3D-ICs, these take full advantage of the third-dimension at the device and module level, to affect increases in trace width and trace separation, reducing interconnect resistance and capacitance, while capturing the inherent reduction in average interconnect length that comes with 3D interconnection. Each of these help to improve the interconnect delays and power dissipation which threaten to make further lithography node scaling ineffectual. Furthermore, we advance an oblique processing approach, membrane projection lithography (MPL), as a fabrication method capable of fabricating this new class of 3D-ICs, requiring only minimal fixturing changes to current state-of-the-art semiconductor fabrication equipment.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

## 2. ADVANTAGES OF DEVICE-LEVEL 3D-ICS

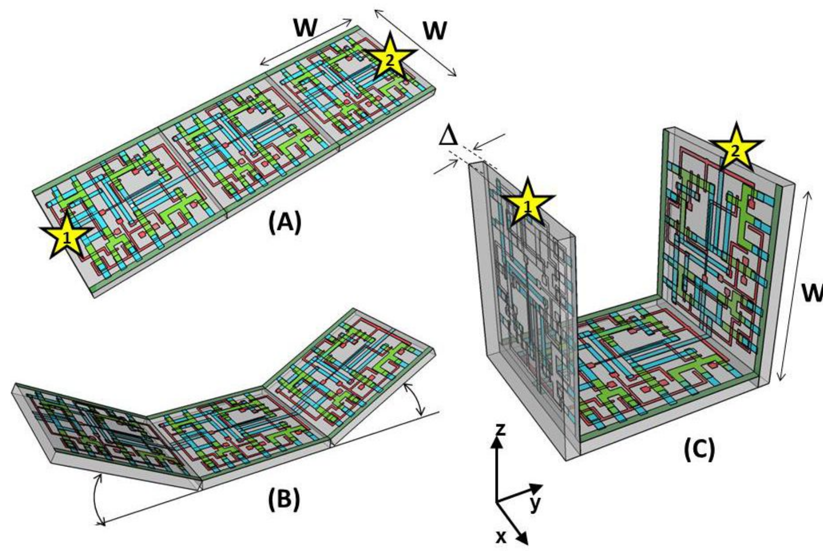


Figure 1. (A) Planar logic module W x 3W in size. Interconnect from point 1 to point 2 is 3W micrometers long; (B) Folding the two edges of the planar module allows for module level three dimensionality; (C) Same module folded into 90° segments. The distance from point 1 to point 2 is only W micrometers long.

Even in two-dimensional topologies, layout, placement and routing are all enormously complex endeavors. In the discussion that follows, no attempt has been made to generate optimal designs or to consider the myriad co-dependent constraints that exist in an actual circuit design, but rather to advance 3-D specific design possibilities that cannot be adopted by a two-dimensional approach, or for that matter, 3D-ICs in either a TSV or monolithic approach.

Consider the planar multi-module layout in Fig. 1(A). This layout block consists of three $W$-by-$W$ micrometer sections with many transistors. The areal footprint of this block is $3W^2$. Furthermore, an interconnect is required from points 1-2 (noted by the yellow stars in the figure). Given the dimensions of the block, this interconnect will be $3W$ micrometers in length, and the trace will be stood off from the metal layers beneath it by the height of the underlying metal layer, typically in the deep sub-micrometer regime. Without addressing how this is to be done for the moment, assume that the two end sections of the block are folded up (Fig. 1(B)) so that they form right angles with the middle section as shown in Fig. 1(C). The device layer of silicon for the two edge sections has a thickness $\Delta$, and the current flow for the transistors in the two edge sections is contained in the *xz*-plane rather than the *xy*-plane. The block in Fig. 1(C) has an areal footprint of $W^2+2\Delta W$, so that for $\Delta << W$, the configuration in Fig. 1 (C) has nearly a factor of 3 increase in areal transistor density versus the planar case in Fig. 1(A).

The interconnect from points 1-2, is now shortened from $3W$ to $W$, a factor of 3 reduction, while the separation of the trace connecting these two points with the underlying metal layers becomes $W$, significantly greater than in the planar case. Shortening the interconnect length reduces ohmic loss and distributed capacitance. Furthermore, the increased separation reduces capacitance and cross talk. In a planar geometry, the separation between traces is fixed by the metal thickness. To lower the capacitance between traces in a 2-D geometry, the only knob to tweak when the distance between traces is fixed is to pursue inter-layer dielectrics (ILD) with smaller dielectric constants. A meaningful reduction in dielectric constant over dense ILDs is only achieved by incorporation of porosity, which necessarily impacts process robustness. Given that the relative dielectric constant of air is 1 and $SiO_2$ is 4, the

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

maximum achievable reduction in capacitance through material changes is 4. In reality, $SiO_2$ has already been replaced by materials with $\varepsilon_r$~2, so that there is much less room for further improvement through material selection.

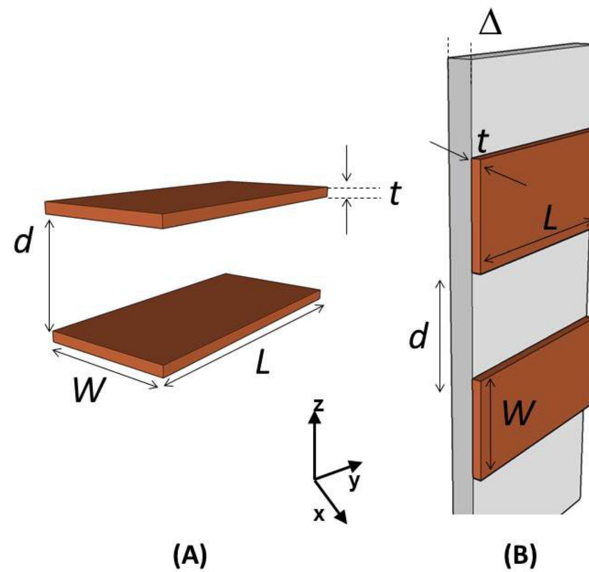


Figure 2. (A) Two horizontally stacked copper interconnect traces with dimensions *W x L x t*, separated by a distance *d*; (B) Two vertically stacked copper interconnect traces with dimensions *W x L x t*, separated by a distance *d*, with significantly reduced capacitance.

Exploring the advantages of 3D topology even further, consider Fig. 2. In Fig 2(A), two parallel traces with thickness, t, are oriented horizontally and stacked vertically. Treating these two traces as a parallel plate capacitor, the capacitance between the two traces is given by

$$C_{1,2} = \varepsilon \frac{A_{1,2}}{d} \qquad (1)$$

where *A* is the cross sectional area of the plates (given by $A_1 = W x L$), *d* is the separation of the plates and $\varepsilon$ is the dielectric constant of the material separating the traces. In Fig 2(B), the same traces are oriented vertically and stacked vertically, separated by the same distance, *d*. In this case the cross sectional area between the two traces is given by $A_2$ =- *t x L*, so that for t << W, the capacitance $C_2 << C_1$. Since the physical dimensions of the traces in both cases are identical, they both possess equivalent current carrying capacity and resistivity. Furthermore, consider the areal footprint of the traces as measured in the xy-plane. The traces in Fig 2(A) have a footprint of W x L whereas the traces in Fig. 2(B) have a footprint of (Δ + *t* ) x *L*, which, if (D + t) < W, has the same current capacity and resistance, smaller capacitance and a smaller footprint.

As a further example of the possible leverage a fully 3D approach has, consider Fig. 3(A), where a trace with a thickness *t* is deposited over a silicon feature with height *h* and width *Δ*. The resulting trace has a cross-sectional area proportional to 2 x *h*, while occupying an areal footprint of just (*Δ* + 2*t*). Another advantage of a "spine" interconnect such as this is in addition to the large cross-sectional area is that it serves as a common interconnect for devices on the left and right side of the spine simultaneously. It is also possible to create coaxial or core-shell interconnects like those shown in Fig. 3(B), where significant shielding is gained. While modern interconnects bear little resemblance to the parallel plate traces of Fig. 2(A), the principles and potential advantages remain.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

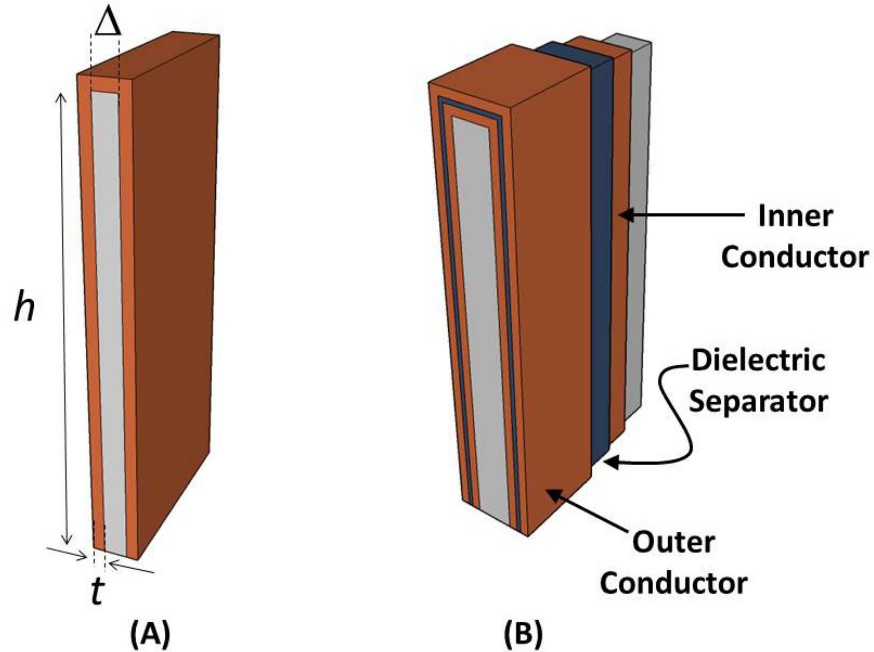


Figure 3. (A) Vertical "spine" interconnect with surface area 2 *x h*; (B) Coaxial vertical spine interconnect for propagating "shielded" signals.

## 3. DEVICE LEVEL 3D-IC GEOMETRY

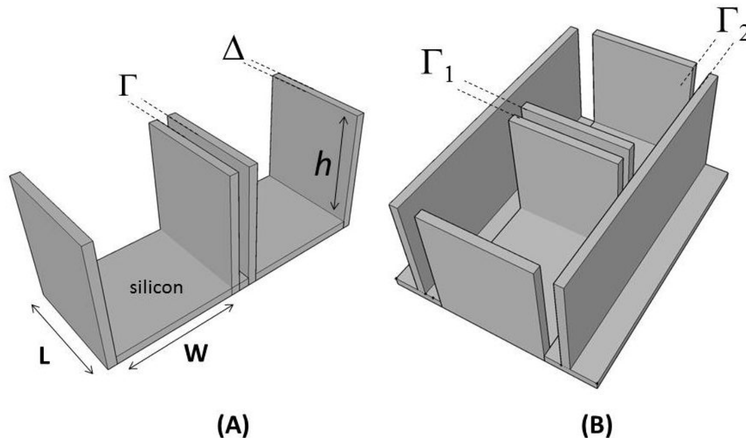


Figure 4. (A) One possible etched, high aspect ratio (HAR) silicon matrix for a "non-folded" fabrication approach to device level 3D-ICs; (B) HAR silicon matrix including orthogonal interconnect spines.

From Section 2, we saw that the folded, 3D version of the logic block has a ~3X increase in areal transistor density, which is attractive. However, aside from the complete lack of manufacturability of such a folding scheme, this approach has at least one more fatal flaw: neighboring modules cannot be densely packed together so that the chip level density remains the same. This can be remedied in a highly manufacturable way, however, as 3D-ICs can be created in a dense array of 3D modules by fabricating the transistors in high aspect ratio (HAR) machined silicon.

DOCKET
ALARM
Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**

Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**

Sync your system to PACER to automate legal marketing.

**WHAT WILL YOU BUILD?** | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase® Smarter legal research.