

Improving Performance under Process and Voltage Variations in Near-Threshold Computing Using 3D ICs

SANDEEP KUMAR SAMAL, Georgia Institute of Technology
GUOQING CHEN, Advanced Micro Devices
SUNG KYU LIM, Georgia Institute of Technology

Near-threshold computing (NTC) circuits have been shown to offer significant energy efficiency and power benefits but with a huge performance penalty. This performance loss exacerbates if process and voltage variations are considered. In this article, we demonstrate that three-dimensional (3D) IC technology can overcome this limitation. We present a detailed case study with a 28nm commercial-grade core at 0.6V operation optimized with various 3D IC physical design methods. First, our study under the deterministic case shows that 3D IC NTC design outperforms 2D IC NTC by 29.5% in terms of performance at comparable energy. This is significantly higher than the 12.8% performance benefit of 3D IC at nominal voltage supplies due to higher delay sensitivity to input slew at lower voltages. Second, it is well demonstrated that transistor delay is more sensitive to voltage changes at NTC operation. However, our full-chip study reveals that IR drop effect on 2D/3D IC NTC performance is not severe due to the low power consumption and hence lower IR drop values. Third, die-to-die variation impact on full-chip performance is visible in 3D IC NTC designs, but it is not worse compared to 2D IC NTC designs. This is mainly due to the shorter critical path length in 3D IC NTC designs.

CCS Concepts: • **Hardware** → **3D integrated circuits**; *Physical design (EDA)*; *Methodologies for EDA*;

Additional Key Words and Phrases: 3D IC, near-threshold computing (NTC), through-silicon-via (TSV), IR drop, variation

ACM Reference Format:

Sandeep Kumar Samal, Guoqing Chen, and Sung Kyu Lim. 2017. Improving performance under process and voltage variations in near-threshold computing using 3D ICs. *J. Emerg. Technol. Comput. Syst.* 13, 4, Article 59 (June 2017), 18 pages.

DOI: <http://dx.doi.org/10.1145/3060579>

This work is an extension of our previous work [Samal et al. 2015a]. It contains significant new material over the two-page conference proceedings in several aspects. First, we discuss the transistor characteristics and the difference in relative impact of different V_{TH} flavors at different VDD. We focus on the cell-delay sensitivity to input transition time and load capacitance and its relative comparison at different VDD. This key feature lays the foundation of improving performance with three-dimensional (3D) ICs. We added the results of 3D IC design at nominal VDD (1.05V) to compare the performance benefits of 3D IC physical design at nominal vs. near-threshold voltages. We elaborated the analysis of results and comparison accordingly. We studied cell-performance impact at different supply voltage, full-chip power delivery network design, and 3D IR drop analysis and its impact on the power/timing of designs. We also discuss the variation impact on the designs with different die-to-die and within-die variation scenarios.

Authors' addresses: S. K. Samal and S. K. Lim, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332; emails: sandeep.samal@gatech.edu, limsk@ece.gatech.edu; G. Chen (current address), Higon IC Design Co. Ltd. Austin, TX; email: chenguoqing@higon.com.

Permission to make digital or hard copies of part or all 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 and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2017 ACM 1550-4832/2017/06-ART59 \$15.00

DOI: <http://dx.doi.org/10.1145/3060579>

1. INTRODUCTION

Near-threshold computing (NTC) has been researched as one of the most attractive ways to achieve significant energy savings in current VLSI systems ranging from smart low-power sensors and medical devices to high-performance servers. However, excessive performance degradation has prevented the use of NTC in practical applications. On the other hand, the advent of three-dimensional (3D) IC technology has opened up a completely new design exploration space for integrated circuits.

NTC and 3D IC provide mutual benefits to bring the best out of both. While NTC designs have an order-of-magnitude lower power resulting in reduced thermal problems and power delivery demand, 3D ICs help in improving the performance both at the physical design and architecture levels. Architecture-level synergistic benefits have been discussed in prior works on NTC, but the impact of 3D IC physical design itself on full-chip performance boost has not been explored.

The major contributions of this article are as follows:

- We demonstrate 29.5% improvement in operating frequency in NTC 3D IC with similar energy as 2D by carefully choosing the partitioning scheme and block folding techniques. We compare this with 3D IC frequency improvement at nominal voltage as well, with detailed explanation of results (Sections 3 and 4).
- Since delay sensitivity to voltage changes is magnified at lower voltages, we compare the impact of IR drop on the full-chip performance degradation and observe similar impact at nominal and NTC designs due to lower IR drop in NTC designs (Section 5).
- We carry out impact study of die-to-die (D2D) and within-die (WID) variations on the critical path delay of the different design implementations with exact critical path simulations, including interconnect parasitic (Section 6).

To the best of our knowledge, this is the first work that studies full-chip 3D NTC circuits and demonstrates its performance benefits under both deterministic and statistical scenarios. Previous works have mostly focused on power savings under no variations. We summarize our design lesson and guidelines in Section 7 and conclude in Section 8.

2. MOTIVATION AND BACKGROUND

For sub-100nm technologies, maximum energy efficiency occurs near the threshold voltage of the transistor because of the increased proportion of leakage energy at very low sub-threshold voltages. Near-threshold computing offers reduced power dissipation and maximum energy efficiency. It creates a feasible opportunity to successfully tap the advantages of device scaling by utilizing all transistors simultaneously without worrying about thermal issues [Dreslinski et al. 2010; Chang et al. 2010; Chandrakasan et al. 2010]. However, excessive performance degradation is a major issue. In addition to the performance penalty compared to their nominal counterparts, high sensitivity to PVT variations at low operating voltages along with increased process variations at advanced technology nodes add to the challenges and reliability of design. Kaul et al. [2012] observe up to 50% frequency variation at low voltages. Most of the proposed techniques to improve performance for NTC designs are limited to architectural changes. These implement NTC-based parallelism that achieves the desired performance while remaining more energy efficient than its single nominal counterpart. Zhai et al. have demonstrated 70% energy savings over a uni-processor system and 53% over conventional multi-processor scaling by using near-threshold parallelism of 10–50MHz cores [Zhai et al. 2007; Dreslinski et al. 2007]. Device optimization for lower voltage operation and newer device technologies like fully depleted silicon-on-insulator with very low leakage are other explored options [Lo et al. 2015; Corsonello et al. 2015; Beigne et al. 2013].

Three-dimensional ICs offer reduced interconnects, reduced footprint, on-chip memory to logic connections, and shorter paths that reduce power and provide potential increase in performance. Benefits of cluster-based NTC architecture with 3D stacking have already been demonstrated in Centip3De [Fick et al. 2013]. Here the authors show four-core cluster systems to be 27% more energy efficient while providing 55% more throughput than a one-core cluster system. The cores and cache are floorplanned into separate layers and use coarse-grained bus-level connections for design simplicity. However, the individual cores are implemented in 2D. 3D ICs also provide the option of logic on logic folding where logic cells are placed in two or more tiers, thereby reducing the signal wirelength [Jung et al. 2015]. This not only results in lower interconnect switching power but also reduces the timing optimization effort due to shorter paths for the same timing target. Jung et al. studied and quantitatively compared power benefits for various implementations of 3D IC folding at nominal voltages for a multi-core processor [Jung et al. 2015]. Their results show that 3D IC has lower power than 2D IC in general, and the block-folding technique saves more power than core/cache partitioning. In particular, they show that 3D stacking with block folding gives 20.3% power saving over 2D IC while 3D stacking without any block folding shows 13.7% power saving. In our work, we use the superior-quality block-folding design technique in the 3D implementation of a single-core commercial processor and observe 12.8% and 29.5% performance benefits in 3D IC at nominal and NTC voltages, respectively, under similar power-delay products.

The use of 3D ICs is accompanied with issues of degraded thermal behavior and power delivery. This is due to increased power density, complicated power delivery that dies away from the package bumps, and increased sources of variation due to D2D variations along with WID variations. Modeling works have mathematically studied the variability in 3D ICs for various scenarios, including die-to-die paths and within-die paths in the same design [Juan et al. 2013; Garg and Marculescu 2009]. They study the impact of different number of tiers and input variations on maximum critical path delays. They also propose simple techniques to reduce such variations by stacking of properly selected dies.

However, 3D ICs have smaller footprint per die, reducing spatial variation and also reduced interconnects resulting in relaxed optimization efforts, especially in advanced nodes. Due to reduced length of the nets, 3D ICs offer the unique opportunity to reduce the length of the critical path, resulting in increased operating frequency. This performance boost is higher at low voltages due to cell-delay sensitivity to input transition times of signals. While this does not change the fact that 3D IC is impacted by more sources of variations, the actual physical design and proper full-chip simulation of critical paths with device as well as interconnect impact is essential before reaching any definite conclusions about variability impact for a given 3D IC design.

3. NTC DESIGN INFRASTRUCTURE

In this section, we discuss the details of our design techniques in general and 3D partitioning and folding in specific. We also present the cell-delay sensitivity comparison to different loads and transition times at different supply voltages. This is one of the primary reasons to achieve higher performance boost in NTC 3D IC.

We use full RTL to GDSII block-level implementation of an OpenSPARC T2 single core as our design under study [Oracle 2014]. The T2 single core at the top level consists of 23 blocks with the few largest blocks being load-store unit (lsu), instruction fetch unit (ifu), and floating point and graphics unit (fgu). We use 28nm technology for our design implementations. We design and compare 2D IC and two-tier Through Silicon Vias– (TSV) based 3D IC at nominal (1.05V) and near-threshold (0.6V) voltages. All

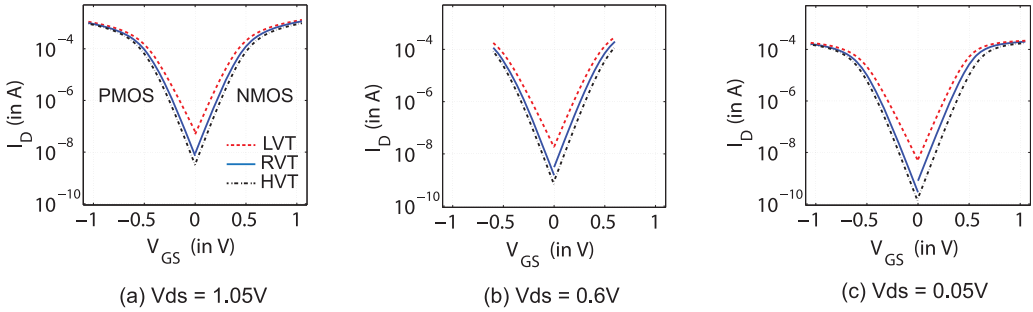


Fig. 1. Transistor characteristics for multi- V_{TH} 28nm technology library: (a) $V_{DS} = 1.05V$, (b) $V_{DS} = 0.6V$, (c) $V_{DS} = 0.05V$ (I_{LIN} curve, which is same for both nominal and near- V_{TH}). Note that the difference in current between multi- V_{TH} transistors is more pronounced at 0.6V (Table I).

Table I. Transistor Current Comparison for 28nm Library for Different V_{TH} Flavors and Supply Voltages. The Relative Difference in Currents among the Three V_{TH} Flavors Magnifies at 0.6V

		I_{ON} (mA/ μ m)	I_{OFF} (nA/ μ m)	I_{LIN} (mA/ μ m)	$\frac{I_{ON}}{I_{OFF}}$
NMOS					
$VDD = 1.05V$					
High-VT	(HVT)	0.98 (1.00)	3.35 (1.00)	0.18	2.93e+05
Regular-VT	(RVT)	1.16 (1.18)	8.78 (2.62)	0.20	1.33e+05
Low-VT	(LVT)	1.32 (1.35)	49.2 (14.7)	0.24	2.69e+04
$VDD = 0.60V$					
High-VT	(HVT)	0.12 (1.00)	0.91 (1.00)	0.04	1.32e+05
Regular-VT	(RVT)	0.20 (1.67)	3.11 (3.42)	0.07	6.43e+04
Low-VT	(LVT)	0.30 (2.50)	18.9 (20.8)	0.11	1.59e+04
PMOS					
$VDD = 1.05V$					
High-VT	(HVT)	0.94 (1.00)	3.13 (1.00)	0.15	2.99e+05
Regular-VT	(RVT)	1.00 (1.06)	7.15 (2.28)	0.16	1.40e+05
Low-VT	(LVT)	1.12 (1.19)	61.3 (19.6)	0.18	1.80e+04
$VDD = 0.60V$					
High-VT	(HVT)	0.08 (1.00)	0.66 (1.00)	0.02	1.21e+05
Regular-VT	(RVT)	0.12 (1.50)	1.52 (2.30)	0.03	7.89e+04
Low-VT	(LVT)	0.18 (2.25)	17.7 (26.8)	0.05	1.02e+04

the designs are pushed to maximum achievable frequency of operation without timing violation in any path.

3.1. NTC Cell and Memory Library

We use a multi- V_{TH} 28nm library for our design with the threshold voltage (V_{TH0}) lying between 0.45V and 0.55V. Figure 1 shows the I_D - V_{GS} characteristics at different drain-to-source voltages, and Table I shows the details of the current values and I_{ON} to I_{OFF} ratio. Figure 1(c) shows the I_{LIN} characteristics used to determine the effect of drain-induced barrier lowering at the different voltages. The relative increase in both on and off currents for the different V_{TH} flavors is more pronounced at near-threshold supply than the nominal 1.05V supply. This implies that a switch from HVT to LVT increases the speed of the transistor by greater than 100% at 0.6V as compared to less

than 35% increase at nominal voltages. While the leakage current also increases in going from HVT to LVT, the gain in on-currents is prominent and the I_{ON} -to- I_{OFF} ratio is similar to nominal voltage. This fact is leveraged later in the NTC designs, since we target maximum performance with minimal power overhead. The change of threshold voltages magnifies the performance improvements at NTC, providing more room for optimization both for 2D and 3D implementations.

Memory reliability is affected at low voltages, and extra design effort is required for proper memory implementations in NTC designs [Hanson et al. 2006]. Hanson et al. [2006] study the 8T SRAM cell for a better static noise margin and reducing the minimum VDD from 0.64V to 0.36V. Therefore, we first analyzed our memory operation through extensive spice simulations and fixed 0.6V as the reliable operating voltage. Though this is higher than the optimal voltage of about 0.5V for maximum energy efficiency, it ensures that memory is not the critical portion of our design. The focus of our research work is to study the performance improvement of NTC voltage with 3D IC physical design. Read time is much slower than nominal voltage conditions, but by setting a 0.6V operating voltage, we ensure that memory is not the most critical part in the full design. Many prior works have demonstrated reliable memory operation at low voltages. Fick et al. use 0.8V for SRAM operation in 130nm technology [Fick et al. 2013]. Hanson et al. show memory operation down to 0.64V and down to 0.36V for 8T SRAM cell designs using 65nm technology [Hanson et al. 2006]. Konijnenburg et al. use memory reliably at 0.4V for 40nm CMOS [Konijnenburg et al. 2013], while Abouzeid et al. operate memory at 0.35V in 28nm technology [Abouzeid et al. 2013]. In our case, the energy benefits of voltage scaling are still significant at 0.6V and above that 3D IC helps in improving frequency. We use the transistor models shown in Figure 1 and characterize our cell libraries for all three V_{TH} flavors at 1.05V and 0.6V using Synopsys SiliconSmart. For our study, we only use typical process and temperature corners during design and analysis. The nominal libraries matched the original library information. Since, we had to characterize libraries at low voltage (0.6V) ourselves, we characterized at both voltages to be fair with the settings.

3.2. Cell Delay Sensitivity

Figure 2 shows the delay sensitivity of an inverter cell in the 28nm technology node to varying input transition times (slew) and load capacitance, respectively. The delay values are normalized with minimum delay of each respective curve at minimum slew (load) to highlight the relative impact of increasing slew (load) at different voltages. The actual cell-delay values at lower supply voltage are much higher ($4\times$ - $5\times$) than that at nominal voltage. It is important to look at this comparison in terms of relative as well as absolute values to properly understand the bigger impact. The key observation is that the delay is much more sensitive to input transition time (input slew) at low voltage supplies. Here input transition time is defined as the time for the input signal to rise from 20% to 80% of the supply voltage (VDD). Alioto et al. have previously studied the delay sensitivity to input rise time and the VDD/V_{TH} ratio and show that delay sensitivity increases by greater than $2\times$ at low voltages (= lower VDD/V_{TH}) [Alioto and Palumbo 2006]. In addition, cells in low-voltage designs operate with higher input transition times values (arrows in Figure 2(a)) due to larger cell delays of the previous stage. For lower VDD (0.6V), the transistor turns ON only after the input has reached 80–90% of the supply but for nominal VDD (1.05V), it is already ON at 40–50% of the supply. Therefore, similar reduction in input transition time will have a larger impact in reducing cell delay at lower supply voltages. The increase in load capacitance has almost similar relative impact on the delay of a single isolated inverter operating at both nominal and near-threshold voltage. However, the case for a chain of cells differs.

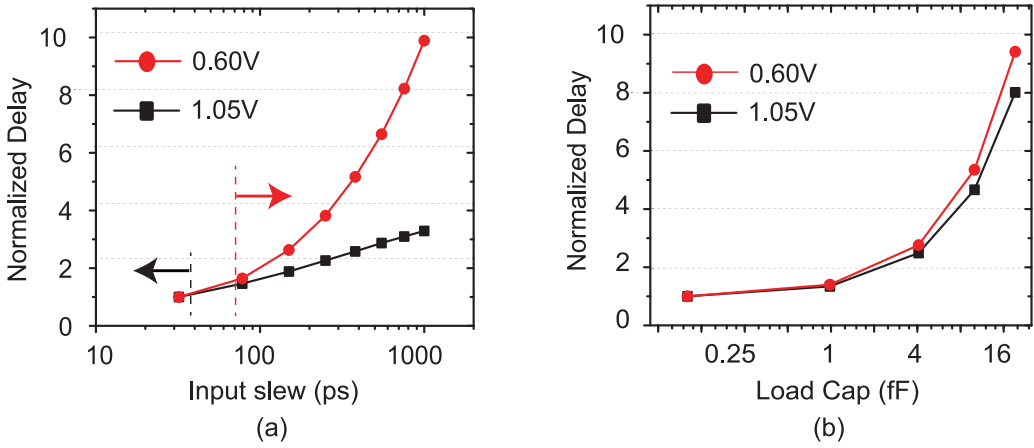


Fig. 2. Delay sensitivity for an inverter cell under nominal and near-threshold voltage supplies for (a) input transition time/slew and (b) load capacitance. The arrows in (a) show the general range of input slew of cells at the respective voltages. The Y-axis denotes delay values normalized with the respective minimum delay at minimum slew (load) to focus on slope (i.e., delay sensitivity) of each curve. The absolute delays at 0.6V are much higher than that at 1.05V.

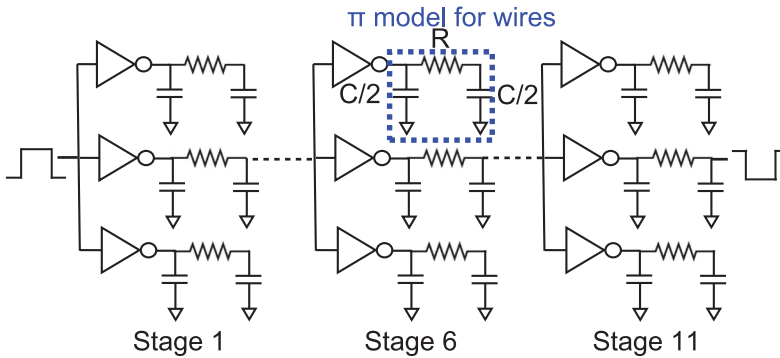


Fig. 3. Inverter Chain Circuit with π -model for interconnects used to demonstrate interconnect impact on delay at different supply voltages.

To compare the relative impact of interconnects on timing at different voltage levels, we conduct a simple experiment with an 11-stage inverter chain where interconnects are represented by the RC π -model. Figure 3 shows the circuit setup. For the interconnects (back end loading), we use $R = 0.5\text{ohm/square}$ and $C = 0.2\text{fF}/\mu\text{m}$ as per 28nm technology specifications. Typically, critical paths have many tens of gates, and the critical path determines maximum frequency of a design. Therefore, we use a higher number of inverters to model a critical logic path. Having very few cells in the chain is not a realistic representation of a critical path. We then simulate the inverter chain with Spice at supply voltages of 1.05V and 0.6V. We measure the full chain delay for average cell-to-cell interconnect lengths varying from $0\mu\text{m}$ (no-interconnect) to $50\mu\text{m}$. The same experiment is repeated for a NAND2 gate chain as well with one input of a NAND gate connected to a supply voltage (Logic 1).

Figure 4 shows the normalized results of this experiment for inverter and NAND chains, respectively. As can be clearly observed, the rate-of-delay increase with interconnects (=slope of curve) is higher at 0.6V. Therefore, even though cell delay (=delay

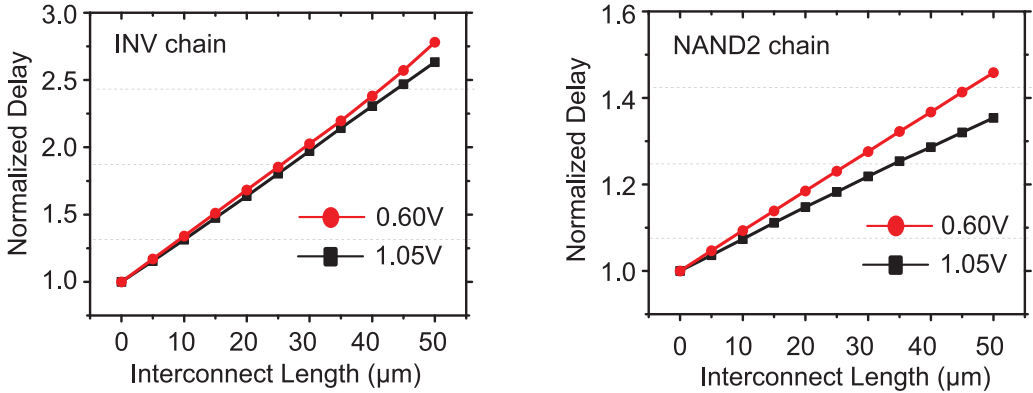


Fig. 4. Normalized total chain delay vs. Interconnect length for Inverter/NAND2 chain at nominal and near-threshold supply voltage. This simplified setup demonstrates that delay degradation with interconnect increase is worse at 0.6V. For interconnect π -model, $R = 0.5\text{ohm/square}$ and $C = 0.2\text{ff}/\mu\text{m}$ (for 28nm technology).

at $0\mu\text{m}$ interconnect) dominates total delay at 0.6V, cell-cell interconnect length change has a considerable impact on total delay. This is due to the cumulative impact of input transition time propagated across the chain. Input transition to one cell (N th cell) is the cell propagation delay of the ($N-1$)th cell in the path and the N th cell's propagation delay is the input transition to the ($N+1$)th cell in the path. As a result, the delay impact magnifies due to the input transition sensitivity as the signal propagates along the path. Even a small reduction in cell delay due to reduced interconnect (RC) will result in cell-delay improvement for the next cell in the path and so on. This relative improvement of path delays will be more at low voltage supplies with higher cell delays and higher sensitivity to input slew.

This is the reason why 3D IC with reduced interconnects can potentially have more performance benefits at NTC voltages. Using this fact and reduced interconnects in 3D IC physical design technique, we study the full-chip performance results of NTC designs vs. nominal designs.

3.3. Full-Chip 2D/3D NTC Design Flow

With the characterized libraries and technology information, we used commercial standard CAD tools with the addition of a few 3D specific in-house tools for all our designs. We used Design Compiler for netlist synthesis followed by Cadence Encounter for place and route optimization. We designed the T2 core in block level based on the top-level architecture. We carried out floorplanning using simulated annealing on soft blocks with area as constraint and inter-block wirelength as the cost function. The block-level 2D IC and 3D IC implementations for NTC designs with placement and routing are shown in Figure 5. We determined the timing budget of blocks in a top-down approach and then designed the individual blocks based on these timing constraints at their I/O pins. For 3D design folding, we incorporated extra steps described in Sections 3.4 and 3.5.

3.4. 3D IC Clock Tree Design

Clock tree is a critical part of any digital circuit. For 2D ICs, we used Encounter for clock tree synthesis after the preCTS optimization stage. However, for 3D ICs, the clock has to travel across both dies, which makes 3D clock tree synthesis more challenging. We use a single 2D clock tree per die with clock nets in both dies connected by only

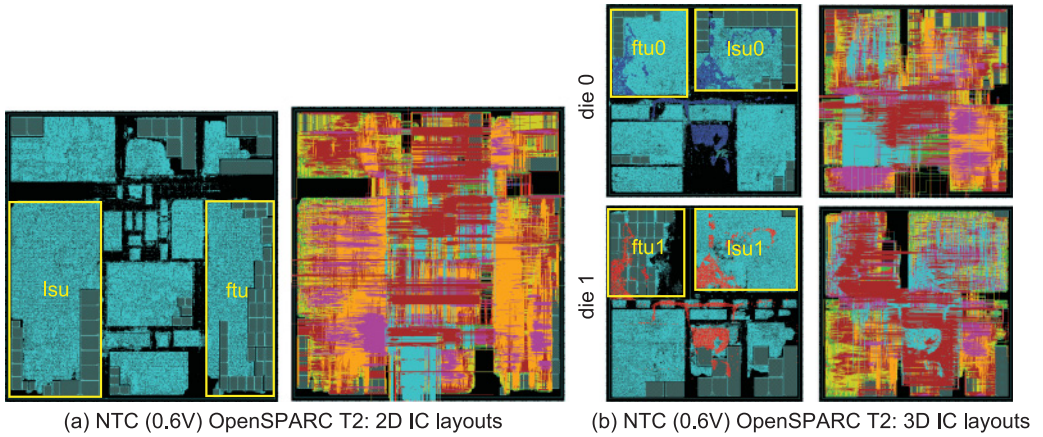


Fig. 5. Near- V_{TH} ($V_{dd} = 0.6V$) OpenSPARC T2 single-core placement and routing views. (a) Two-dimensional implementation (footprint $1.75 \times 1.64mm$) and (b) 3D implementation (footprint = $1.2 \times 1.2mm$). Folded blocks (lsu and ftu) are highlighted in yellow. There are 3,381 TSVs shown in blue in die0 and the corresponding landing pads are in red in die1 in the placement view. Top-level, lsu, and ifu_ftu have 1531, 1132, and 718 TSVs, respectively. All layouts are shown to scale.

Table II. Distribution of Power Consumption in Single Core T2

module	ifu	lsu	fgu	tlu	exu	mmu	others	top
% of total power	29.5	27.4	10.6	10.4	9.4	5.9	4.1	2.7

one TSV [Jung et al. 2015]. This is not the best method, but we can use commercial tools for high-quality clock tree designs per die by treating the clock TSV as a sink for die0 (the die connected to package I/Os) and as a clock source for die1. Individual 2D clock tree per die is also essential for pre-bond testability of 3D ICs. Prior works have studied 3D IC CTS and developed sophisticated algorithms to build clock tree topology with multiple TSVs considering pre-bond testability, TSV-coupling impact, clock power optimization, and so on [Yang et al. 2011; Liu et al. 2013]. However, their analysis is limited to Spice simulations of clock tree models and TSVs with very few clock sinks. Moreover, it does not include actual routing and parasitic extraction. Our design has more than 40K clock sinks and, therefore, we use the simple approach using commercial tools. Using this approach for just one iteration will increase clock skew, but we carry out multiple iterations of die-by-die design with updated boundary conditions for timing delays at TSVs. More details are discussed in Section 3.5.

3.5. 3D NTC Design with Block Folding

While multi- V_{TH} optimization helps in improving speed in 2D IC OpenSPARC T2, the presence of long nets affects the overall timing and also increases power due to increased wirelength. 3D IC implementation facilitates shortening of nets in general. To reduce the net lengths further, we implement a two-stage design-folding strategy [Jung et al. 2015]. First, we select the most power-hungry blocks in the design and fold them into two tiers. The folding is carried out based on the intra-block architecture such that the highly connected sub-modules remain in the same tier. For our design case, lsu and ifu_ftu are the largest and most power-consuming blocks (Table II). These folded blocks have their own intra-block 3D TSV connections and communicate with the other blocks in the design through their block pins similar to the 2D IC implementation. TSVs have a diameter of $4\mu m$ with $R = 40m\Omega$ and $C = 10fF$ [Katti et al. 2010].

Table III. Design Summary of the Four Different Implementations of OpenSPARC T2 Single-Core Designed at Maximum Achievable Frequency. The Number in Brackets Denotes the Percentage of Respective Total Cell Metric to the Nearest Integer

	Footprint (mm ²)	Max Freq. (MHz)	Cell Area (mm ²)	Buffer Area (mm ²)	# HVT-Cells (×1000)	# RVT-Cells (×1000)	# LVT-Cells (×1000)	WL (m)
Nominal								
2D IC	1.64 × 1.75	813.0	1.22	0.12 (10%)	278.6	71.7 (20%)	15.4 (4%)	14.8
3D IC	1.20 × 1.20	917.4	1.31	0.17 (13%)	282.2	78.4 (19%)	52.1 (13%)	16.2
NTC								
2D IC	1.64 × 1.75	116.3	1.20	0.09 (8%)	253.5	103.7 (28%)	9.6 (3%)	14.6
3D IC	1.20 × 1.20	150.6	1.22	0.10 (8%)	257.6	102.9 (27%)	25.9 (6%)	14.7

Based on this folded netlist of the blocks, we carry out top-level partitioning and 3D IC floorplanning to reduce the inter-block wirelength. The folded blocks are kept at the same location in both dies (Figure 5). Using the 3D IC folding results and the x-y-z location of the blocks, we use the netlist connectivity in each die to partition the pins of the folded blocks (lsu and ifu_ftu) into the two separate dies. This pin partitioning strategy not only ensures reduced wirelength and enhanced connectivity but also reduces the addition of too many TSVs. All block pins are placed at the boundaries of the respective blocks. For the folded blocks, the internal TSV locations are inside the block area. The top-level TSVs are only placed at inter-block whitespace.

With wirelength driven floorplanning of the blocks along with architectural partitioning of two blocks, our 3D IC design has a total of 3,381 TSVs of which 1,132 are for lsu only and 718 are for ifu_ftu only. The final utilization of silicon area is more in 3D IC than in 2D IC because of more cells and TSVs with increased performance. However, block folding with pin partitioning and 3D IC floorplanning helps in reducing top-level routing congestion. Another important design feature is the intentional use of large white space between blocks in die0 to facilitate optimized TSV insertion and ensure short connections between blocks. TSVs are treated as standard cells in die0 during TSV planning, and the TSV insertion algorithm minimizes the 3D wirelength. Therefore extra whitespace allows more optimized planning. However, in the process of allocating white space, we maintain the overall silicon area to be the same in 2D IC and 3D IC implementations (Table III).

Top-level die timing constraints are obtained by context characterization using Synopsys PrimeTime followed by budgeting of block-level timing within each die using a top-down approach. The exact global die constraints ensure that block timing budgets are based on the whole 3D IC design including all dies and not just that particular die. All design implementations are targeted for maximum achievable clock frequency. Multiple design iterations are carried out as convergence of 3D IC timing requires accurate timing constraints at the die boundaries (TSV interface) where signal goes from one die to another using TSVs. The TSV parasitic also needs to be included while obtaining these constraints. This is followed by individual die-by-die design. Since current commercial tools cannot handle 3D IC timing optimization and 3D IC co-design, these multiple iterations ensure that the die I/O delays are set correctly including TSV impact while designing the individual dies.

4. 3D NTC PERFORMANCE BOOST

4.1. Power-Performance Comparison

As discussed in the previous section, all our designs are targeted to achieve maximum attainable frequency. Based on this design and optimization approach, we observe that nominal 2D IC reaches up to 813MHz (1.23ns clock) while the best frequency of NTC 2D IC is 116.3MHz (8.6ns clock). Two-tier 3D IC, on the other hand, beats its 2D counterpart by a good margin by going up to a frequency of 917.4MHz (1.09ns clock)

Table IV. Power-Performance Comparison under No Variations. Numbers in Brackets Denote Percentage Relative to Respective 2D IC Design. All Power Numbers Are in mW

	Frequency (MHz)	Switching Power	Internal Power	Leakage Power	Total Power	Power-Delay Product (pJ)
Nominal						
2D IC	813.0	224.3	633.2	16.7	874.2	1075.3
3D IC	917.4	254.4	679.7	18.6	952.8	1038.6 (-3%)
NTC						
2D IC	116.3	9.7	23.9	1.2	34.8	299.3
3D IC	150.6	9.9	31.6	1.4	42.9	284.9 (-5%)

and 150.6MHz (6.64ns clock) for nominal and NTC voltage supplies, respectively. The relative performance improvements in 3D IC are 12.8% at nominal voltages and a significant 29.5% at NTC voltages. Table III presents the details of all four design implementations. Since all designs are pushed to maximum limits, the number of buffers are a significant portion of the total cell count and differ for different design implementations.

In the final designs, 3D ICs have more cell area compared to its 2D IC counterparts. With shorter wires in 3D IC, it is expected to have less cell and buffer usage than 2D IC for iso-frequency designs. However, 3D IC designs here successfully run at a much higher frequency compared to 2D IC. During timing optimization, it is possible to insert more buffers into the 3D design to achieve these faster clock periods as the wires are shorter (lower RC), which results in shorter transition times. On the other hand, 2D IC design has longer nets that cannot be pushed faster even with the insertion of many timing buffers and are optimized to the best extent. The designs have many timing paths and each path is optimized to meet timing constraints. We run multiple iterations to achieve fastest frequency per design implementation (Table III). For the final design optimization, we used a target clock skew of 8% of respective clock period with a clock uncertainty factor of 5%. Cadence-Encounter modifies the netlist during timing optimization depending on timing and power constraints and optimization feasibility. The best timing targets differ for different designs. Therefore, the absolute constraint numbers vary, even though they are similar relative to their respective clock period. Buffers are added, and the type and count of cells change, for example, a multi-input AND is replaced with multiple 2-input ANDs. Timing is successfully closed for 3D IC at a faster clock compared to 2D IC and there are more such netlist changes for 3D IC resulting in more cell usage apart from extra buffers.

As demonstrated earlier, delay sensitivity to input transition time for a gate in a path is much higher at low voltages than at nominal voltage supplies and 3D IC helps in improving the transition time by reducing wireload. Also threshold voltage switch has much higher impact at lower VDD (Section 3.1). Such V_{TH} swaps will only happen when the design optimization engine (Cadence-Encounter here) can improve performance further, after considering the back-end loading as well. Therefore, the relative improvement at 0.6V VDD is much higher (29.5%) compared to 1.05 VDD (12.8%).

The detailed power analysis results are presented in Table IV. All post-layout power and timing analysis is carried out with Synopsys PrimeTime. Synopsys Primetime reports internal power that is dynamic plus short-circuit power inside standard-cells due to switching of internal-nodes only. Our 3D IC designs have more cells due to tighter clock constraints, more low- V_{th} cells, and run faster. Therefore, internal power is higher. However, 3D inter-cell net-switching power is not much higher than 2D IC because of shorter individual nets, that is, lower load. This switching power is not very high in 3D IC NTC design even though it runs at faster frequency. Though there are

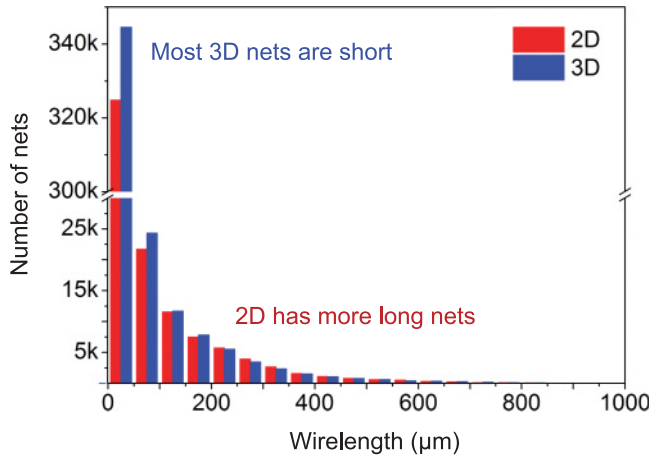


Fig. 6. Number of nets in different wirelength bins for NTC implementation with 2D IC and 3D IC. Total number of nets is 383,599 in 2D and 405,599 in 3D. There is a break in the Y-axis from 30k to 300k.

405.6K nets in 3D IC and 383.6K nets in 2D IC design at 0.6V, the overall wirelength is almost equal, which implies that the average net length is shorter in 3D IC. More LVT cells in 3D IC result in higher leakage but helps in getting the performance boost. The scaling of voltage in 2D IC domain reduces power by $25\times$ and performance by $7\times$, resulting in power-delay product (PDP) savings of $3.6\times$. NTC 3D IC not only increases performance by 29.5% over NTC 2D IC but also reduces PDP by another 5%. The PDP improvement at nominal voltage is 3%.

4.2. Analysis of Results

Here we take a closer look at why and how 3D IC implementation of the same design provides such significant performance benefits over 2D ICs. One of the primary reasons is the reduction of interconnects in 3D IC due to reduced footprint per die with 3D TSV connections. 3D IC design folding reduces the die footprint by 50%. In addition, another level of block folding (for `lsu` and `ifu_ftu`) brings the cells still closer. The cells come closer, and hence most nets become shorter (Figure 6). Even the paths confined in one die become shorter. NTC operation is very sensitive to input transition times. For the same increase in input transition of a gate, the propagation delay increases more at lower voltage than at nominal voltage. As discussed earlier in Section 3.2, the input transition time is a direct function of RC parasitics of previous stage. Therefore, lesser RC parasitics in 3D IC result in lower propagation delays compared to 2D ICs.

Figure 6 shows the distribution of nets in both 2D and 3D designs at NTC based on their lengths. We clearly see that 3D nets are mostly shorter in length and fall in the minimum bin of distribution. Although the total number of nets is higher in 3D due to more cells, their short length has lesser load capacitance and does not degrade the transition time of signal seen by the next cells in the paths, and therefore cell delay is less than 2D IC. In addition, the count of longer nets is higher in 2D than in 3D. The 100 most-critical paths in both NTC designs are shown in Figure 7. It is clear from the highlighted nets that critical paths in 2D IC span a longer length in general compared to 3D IC. The `lsu` block has most of the critical paths and its folding in 3D not only brings intra-`lsu` cells closer but also reduces the inter-block net lengths. It is expected that the effect of interconnects on overall delays will be more and more prominent in advanced technologies and proper 3D IC design helps us reduce that problem to a good extent.

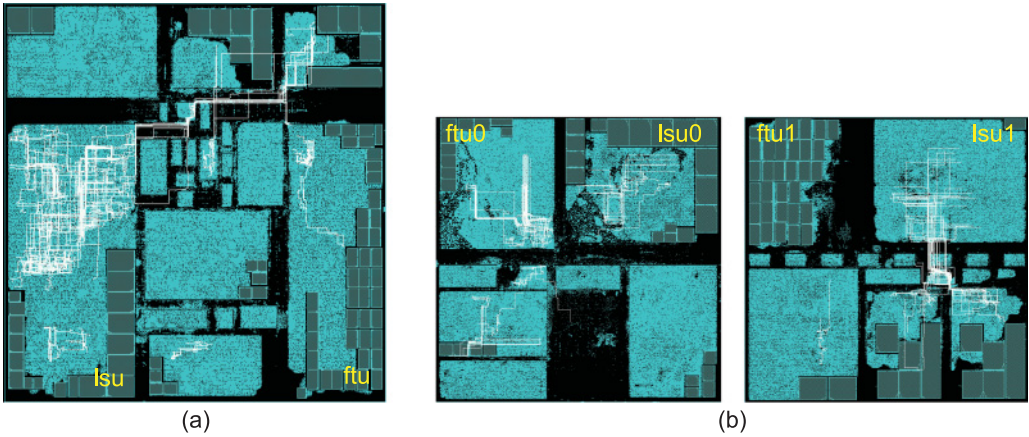


Fig. 7. The 100 most-critical paths for NTC designs highlighted in (a) 2D IC and (b) 3D IC implementations. The 2D IC paths have longer spread.

Table V. Relative Delays for Every 10mV Voltage Drop at 1.05V (Nominal) and 0.6V (NTC) for Different Cells (Figure 8)

	INV	BUF	NAND	XOR	NOR	MUX	DFF
1.05V	1.030	1.022	1.022	1.018	1.026	1.018	1.020
0.60V	1.036	1.059	1.072	1.075	1.077	1.079	1.089

Since transistors' sensitivity to variations increases at lower voltages and 3D IC adds an additional source of die-to-die variations along with power delivery issues, we carry out an impact study of full-chip IR drop and process variations on timing of the designs. We focus only on the NTC (0.6V) designs and compare the full-chip IR drop and variation impact on these with an impact on nominal 2D IC design that represents the common design practice.

5. IR-DROP IMPACT ON 2D/3D NTC

5.1. Cell Behavior with Voltage Drop

To measure the impact of IR drop on timing of the designs, we first need to assess the impact of voltage drop on individual cell delay. The normalized delay degradation of seven representative cells of size X2 with voltage drop at 1.05V and 0.6V is shown in Table V with the degradation slope shown in Figure 8. We clearly observe that the degradation is significantly worse when the supply voltage is lowered to 0.6V. Moreover, the degradation gets worse with an increase in complexity of the cells. While the inverter has a relative delay degradation of 1.036X/10mV drop in VDD, D-Flipflop has 1.089 \times delay degradation for a similar voltage drop at 0.6V. This is because of the longer transition times in the internal stages of the cells. Since the NOR, NAND, and so on, gates have similar complexity, the amount of degradation is almost equal for such gates. At nominal supply voltages of 1.05V, delay sensitivity to the internal transition time is not very critical, and all cells behave similarly. We also observe that the relative impact of voltage drop on different sizes of cells of the same functionality is similar. Based on these observations, we categorize the cells based on their complexity and assign delay de-rate as a function of the voltage drop from the respective supply (Table V). We then use these de-rate values to carry out IR drop aware timing analysis of the different designs implemented.

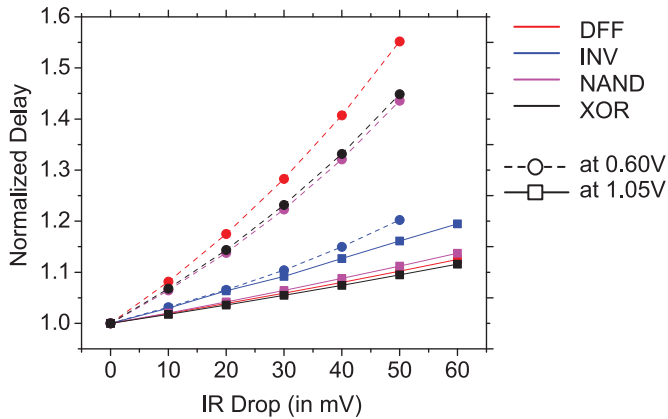


Fig. 8. Impact of voltage drop on cell delay at nominal voltage = 1.05V and near- V_{TH} voltage = 0.60V. Complex cells are much more affected at low voltages due to internal transition times of various stages inside the cell (Table V).

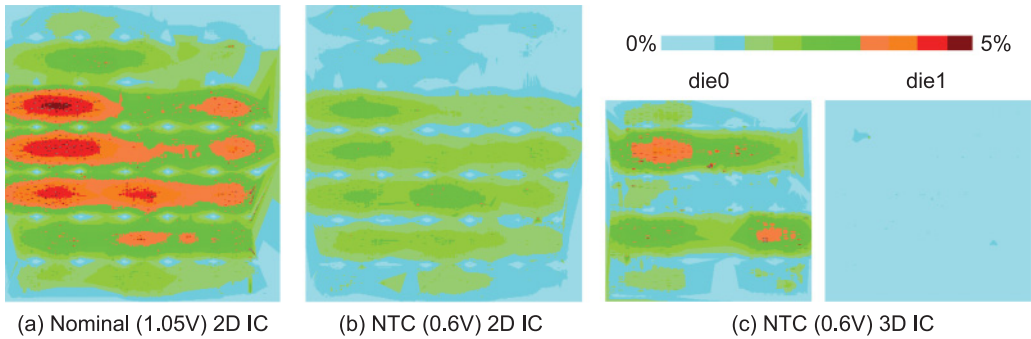


Fig. 9. IR-drop maps of OpenSPARC T2 single core for (a) 2D at 1.05V, (b) 2D IC at 0.6V, and (c) 3D IC at 0.6V with similar PDN and supply bump density. The scale is in % IR drop relative to the actual supply voltage. The maximum absolute IR drop values are 58mV, 20mV, and 27mV, respectively. Nominal supply voltage has more drop as the current tapped from source is much higher.

5.2. PDN Design and IR-Drop Analysis

Uniform power delivery network (PDN) mesh is added to all designs with VDD bump pitch of $200\mu\text{m}$ distributed over the entire footprint area in 2D IC and 3D IC. The bumps are connected to the top metal PDN mesh of 2D IC and die0 of 3D IC (die closer to package bumps). The density of metal usage for PDN is targeted for maximum IR drop limit of around 5% of supply voltage. In our work, we use the same power routing density across all designs. P/G TSVs are added at the same co-ordinates as the bump locations to provide supply to die1 in 3D. Since 2D designs are larger in footprint, they have more supply bumps (8×8) compared to 3D IC (6×6). We use the Encounter Power System to carry out dynamic IR drop analysis at the worst switching timing window of 500ps within the respective clock period. The peak current in this time window is higher than the average current over the entire clock period and therefore results in considerable IR drop at the cell locations.

The IR drop maps are shown in Figure 9. Since 3D IC runs at higher frequency, it has more power demand over half the footprint, therefore increasing the current density by more than double. The P/G TSVs further add to the PDN resistance [Katti et al. 2010]. As a result, 3D IC has more IR drop than 2D IC at 0.6V. The current demand

Table VI. Performance Degradation with IR Drop-Aware Timing Analysis. Numbers in Brackets Represent the Relative Change in Values Compared with the Case without Any IR Drop (Table IV)

	Nominal 2D IC	Nominal 3D IC	NTC 2D IC	NTC 3D IC
Max IR drop (mV)	58	74	20	27
IR drop (w.r.t. VDD)	5.5%	7.0%	3.3%	4.5%
Max Frequency (MHz)	769.0 (-5.4%)	863.9 (-5.8%)	110.0 (-5.2%)	142.8 (-5.2%)
Switching Power (mW)	217.5 (-3%)	245.5 (-4%)	8.9 (-8%)	9.5 (-4%)
Internal Power (mW)	606.5 (-4%)	652.8 (-4%)	22.8 (-5%)	30.4 (-4%)
Leakage Power (mW)	16.7 (0%)	18.6 (0%)	1.2 (0%)	1.4 (0%)
Total Power (mW)	874.2 (-4%)	916.9 (-4%)	32.9 (-5%)	41.3 (-4%)
Power-Delay Product (pJ)	1093 (+2%)	1054 (+1%)	300 (0%)	289 (+1%)

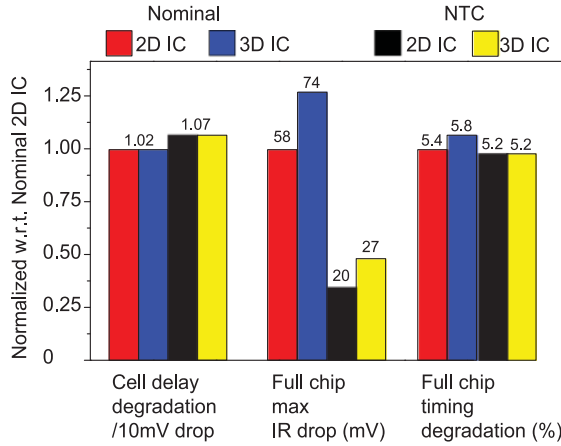


Fig. 10. Comparison of IR drop impact on the various design implementations normalized w.r.t. Nominal 2D IC values (Table VI).

in die0 is more in the timing window with peak current and hence die0 has more IR drop. The 1.05V 2D IC has maximum IR drop in terms of absolute number, since the peak current drawn from supply is very high compared to the current demand at NTC designs.

5.3. Full-Chip Timing and Power Results

Table VI gives the details of the impact of IR drop on the timing and power of the designs at nominal and near-threshold voltages. The relative change with respect to the corresponding values without any IR drop is reported along with the maximum IR drop numbers for each implementation. Even though the cell delay degradation is less at nominal voltage, the high IR drop values result in more degradation in overall timing. For the NTC design, the impact is almost similar in 2D and 3D, since the worst IR drop does not happen at the timing-critical paths. Though this observation is design specific, it shows that 3D IC performance is not necessarily degraded more even though its overall power delivery is worse for the same PDN density. The shorter path lengths in 3D IC also help in keeping delay degradation small. Therefore, on the full-chip scale, the overall timing impact of IR drop on our NTC design is similar to nominal design and not worse as observed for individual gates. This is explained by Figure 10, which shows that relative sensitivity per 10mV drop is higher at low voltages, but final IR drop values are lower. This makes the overall impact equivalent.

Table VII. Standard Deviation over Mean Ratio (σ/μ) of Delay Distribution of Nominal 2D IC, Near- V_{TH} 2D IC and 3D IC with Different D2D and WID Variations

Input σ_{D2D}	Input σ_{WID}	Delay σ/μ		
		Nominal 2D IC	NTC 2D IC	NTC 3D IC
0.05	0.05	0.04	0.15	0.13
0.05	0.10	0.06	0.26	0.27
0.05	0.15	0.07	0.35	0.36
0.10	0.05	0.08	0.44	0.46
0.10	0.10	0.09	0.49	0.49
0.15	0.05	0.12	0.66	0.61

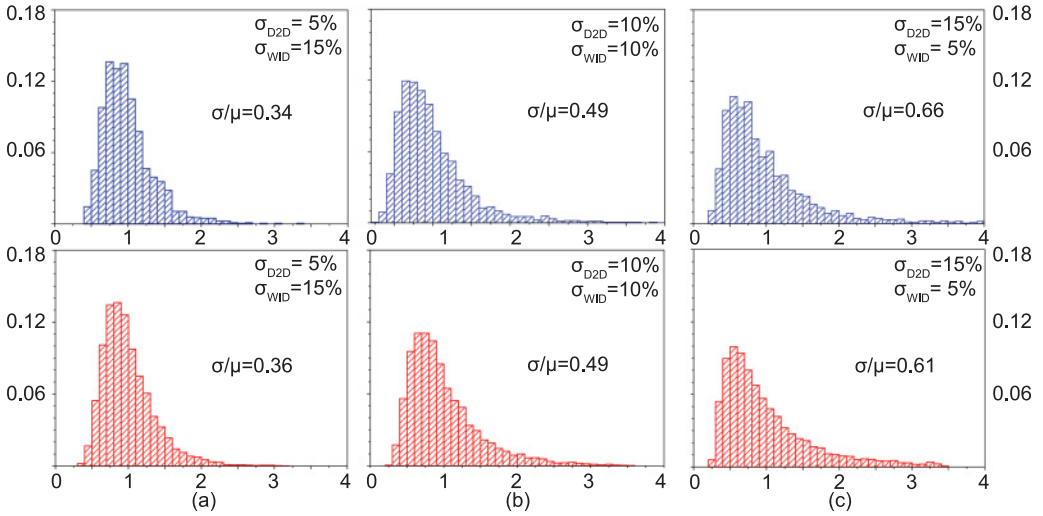


Fig. 11. Delay distribution in different process variation scenarios at 0.6V operation for (a) D2D = 5%, WID = 15%, (b) D2D = 10%, WID = 10%, (c) D2D = 15%, and WID = 5%. Blue histograms are for 2D IC design and red ones for 3D IC. All delay values are normalized w.r.t. the mean of distribution. The Y-axis denotes the fraction of occurrence in total Monte Carlo simulations.

6. VARIATION IMPACT ON 2D/3D NTC

We study the impact of process variations on the timing at nominal 2D IC and NTC 2D IC and 3D IC designs. For simplified analysis and comparison, we use threshold voltage variations to model the D2D and WID process variations. Since the 3D design consists of two separate dies, the D2D variations are independent for the two tiers unlike in 2D ICs where the variation is same across the die. This variation is captured with the use of independent systematic variation input for the different tiers in 3D IC. Random variations are introduced to model within die variations where each transistor in a die experiences an independent variation in addition to the systematic change across that entire die. For our variation analysis study, we choose the 10 most-critical paths in the designs, respectively, and extract accurate spice model netlists for those paths with the help of PrimeTime. These netlists not only have cell information but also contain extracted wire RC parasitics. We then carry out 1,000 Monte Carlo simulations on each of these paths with different combinations of D2D and WID variations. The results are reported in Table VII. Figure 11 shows the normalized delay distributions for NTC 2D IC and 3D IC.

We observe that although 3D IC has additional sources of variation, the reduction in path length results in a delay distribution close to the distribution of the 2D IC

design. In our 3D IC design implementation, most of the critical paths are confined to a single die due to proper floorplanning and design folding. The entire timing path lies in one die as in 2D ICs. Therefore, D2D variations do not have much impact on 3D IC timing path delay. Note that TSVs have 10fF capacitance that cannot be split by buffer addition. At low voltage supplies, this high capacitance leads to large delays. Avoiding TSVs from the critical paths is important especially with low voltage supply. We observe that systematic variation plays a more important role in the overall delay distribution, because it affects the die as a whole. Random variation tend to average out and therefore have less of an impact. NTC designs have $5\times$ more variations than nominal design, which is in agreement with Dreslinski et al. [2010] which shows a $5\times$ delay variation at a 400mV supply voltage.

7. DESIGN LESSONS AND GUIDELINES

7.1. Key Lessons

We summarize our design lessons as follows:

- Cell delay is more sensitive to input transition time at lower voltages. The relative performance difference of multi- V_{TH} transistors are also higher at lower supply voltages.
- 3D IC NTC circuits achieve a significant performance improvement by reducing the critical path length and wire RC parasitic. The power-delay-product is similar to that of 2D IC NTC design.
- The cell delay degradation caused by the supply voltage drop is higher at low voltage operation due to the weaker transistor operation.
- The full-chip IR drop is significantly lower in NTC compared with the nominal voltage case due to the low current demand of the cells.
- The combined effect of high cell delay sensitivity and low full-chip IR drop compensate for each other in NTC circuits. Thus, their overall impact on timing degradation is not always worse in NTC designs. The actual values significantly depend on the length and location of the critical paths. Therefore, a similar PDN design can be used to keep the IR impact at the same level when going from the nominal to the NTC operation.
- 3D IC shows worse IR drop due to the increased current density (= similar current demand in half the footprint) and longer vertical power/ground paths through TSVs. But the overall impact on timing is not necessarily worse than 2D IC and depends on the voltage drop at the cells in the critical path.
- 3D IC is under the influence of additional die-to-die variations due to die stacking. However, its impact on full-chip timing depends on the physical layout of the critical paths. 3D IC is not necessarily worse than 2D IC, because not all the critical paths lie across multiple dies.

7.2. 3D NTC Design Guidelines

We offer the following design guidelines for 3D NTC circuits:

- We suggest that designers keep the critical paths within a single die and closer to the PDN. This helps in reducing the impact of die-to-die variations and IR drop while improving performance.
- We suggest that physical design methods for 3D ICs including block folding, pin partitioning, and 3D floorplanning be used in 3D NTC circuits to further optimize performance.

—We suggest using similar PDN pitches for 3D NTC as in nominal designs along with accurate sign-off analysis. A denser PDN may not be necessary as long as the cells in critical path are not severely affected.

7.3. Near- V_{TH} vs Sub- V_{TH} 3D ICs

In general, NTC designs offer higher energy efficiency than sub-threshold designs [Kaul et al. 2012]. This is due to the significant reduction in design performance, leading to higher leakage energy for sub-threshold designs. NTC designs dissipate lower leakage energy with relatively faster frequency. The detailed design and comparison of low power sub-threshold 3D IC designs with 2D ICs has been studied in Samal et al. [2015b]. While the low power operation is very attractive, the operating speed takes a huge hit with frequencies going down to the kHz range. However, in NTC 3D ICs, we can maintain reasonable frequency while utilizing the advantages of 3D IC architecture [Dreslinski et al. 2010] and physical design as discussed in this work.

8. CONCLUSION

In this article, we demonstrate NTC performance improvement by 3D IC physical design using block folding and pin partitioning and observe 29.5% faster performance than 2D IC NTC design with similar energy for the OpenSPARC T2 single-core processor. This is much higher than the 12.8% performance improvement for 3D ICs at nominal voltages. Even though 3D IC has more variation and worse IR drop for iso-PDN design, we also show that the final impact on delay and hence performance is not necessarily worse and depends on the actual physical design. Lower IR drop values at low voltage operation keep the overall impact of IR drop on timing similar to that of nominal design, which has a higher IR drop. Therefore, 3D IC physical design and optimization can be used to provide performance boost for NTC designs in addition to architectural changes. Our quantitative results are based on OpenSPARC T2 core design case but the design observations and lessons can be qualitatively extended to other design cases as well.

REFERENCES

- F. Abouzeid, A. Bienfait, K. C. Akyel, S. Clerc, L. Ciampolini, and P. Roche. 2013. Scalable 0.35V to 1.2V SRAM bitcell design from 65nm CMOS to 28nm FDSOI. In *2013 Proceedings of the ESSCIRC (ESSCIRC)*. 205–208. DOI: <http://dx.doi.org/10.1109/ESSCIRC.2013.6649108>
- M. Alioto and G. Palumbo. 2006. Impact of supply voltage variations on full adder delay: Analysis and comparison. *IEEE Trans. VLSI Syst.* 14, 12 (Dec. 2006), 1322–1335. DOI: <http://dx.doi.org/10.1109/TVLSI.2006.887809>
- E. Beigne, A. Valentian, B. Giraud, O. Thomas, T. Benoist, Y. Thonnart, S. Bernard, G. Moritz, O. Billoint, Y. Maneglia, P. Flatresse, J. P. Noel, F. Abouzeid, B. Pelloux-Prayer, A. Grover, S. Clerc, P. Roche, J. Le Coz, S. Engels, and R. Wilson. 2013. Ultra-wide voltage range designs in fully-depleted silicon-on-insulator FETs. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2013*. 613–618. DOI: <http://dx.doi.org/10.7873/DATE.2013.135>
- A. P. Chandrakasan, D. C. Daly, D. F. Finchelstein, J. Kwong, Y. K. Ramadass, M. E. Sinangil, V. Sze, and N. Verma. 2010. Technologies for ultradynamic voltage scaling. *Proc. IEEE* 98, 2 (Feb. 2010), 191–214. DOI: <http://dx.doi.org/10.1109/JPROC.2009.2033621>
- L. Chang, D. J. Frank, R. K. Montoye, S. J. Koester, B. L. Ji, P. W. Coteus, R. H. Dennard, and W. Haensch. 2010. Practical strategies for power-efficient computing technologies. *Proc. IEEE* 98, 2 (Feb. 2010), 215–236. DOI: <http://dx.doi.org/10.1109/JPROC.2009.2035451>
- P. Corsonello, S. Perri, and F. Frustaci. 2015. Exploring well configurations for voltage level converter design in 28 nm UTBB FDSOI technology. In *2015 33rd IEEE International Conference on Computer Design (ICCD)*. 499–504. DOI: <http://dx.doi.org/10.1109/ICCD.2015.7357157>
- R. G. Dreslinski, M. Wiecekowsi, D. Blaauw, D. Sylvester, and T. Mudge. 2010. Near-threshold computing: Reclaiming moore's law through energy efficient integrated circuits. *Proc. IEEE* 98, 2 (Feb. 2010), 253–266. DOI: <http://dx.doi.org/10.1109/JPROC.2009.2034764>

- R. G. Dreslinski, B. Zhai, T. Mudge, D. Blaauw, and D. Sylvester. 2007. An energy efficient parallel architecture using near threshold operation. In *16th International Conference on Parallel Architecture and Compilation Techniques, 2007 (PACT 2007)*. 175–188. DOI : <http://dx.doi.org/10.1109/PACT.2007.4336210>
- D. Fick, R. G. Dreslinski, B. Giridhar, Gyouho Kim, Sangwon Seo, M. Fojtik, S. Satpathy, Yoonmyung Lee, Daeyeon Kim, N. Liu, M. Wiecekowsky, G. Chen, T. Mudge, D. Blaauw, and D. Sylvester. 2013. Centip3De: A cluster-based NTC architecture with 64 ARM cortex-M3 cores in 3D stacked 130 nm CMOS. *IEEE J. Solid-State Circ.* 48, 1 (Jan. 2013), 104–117. DOI : <http://dx.doi.org/10.1109/JSSC.2012.2222814>
- Siddharth Garg and D. Marculescu. 2009. 3D-GCP: An analytical model for the impact of process variations on the critical path delay distribution of 3D ICs. In *ISQED, 2009*. 147–155. DOI : <http://dx.doi.org/10.1109/ISQED.2009.4810285>
- Scott Hanson et al. 2006. Ultralow-voltage, minimum-energy CMOS. *IBM J. Res. Dev.* 50, 4/5 (2006), 469–490.
- Da-Cheng Juan, S. Garg, and D. Marculescu. 2013. Impact of manufacturing process variations on performance and thermal characteristics of 3D ICs: Emerging challenges and new solutions. In *ISCAS*. DOI : <http://dx.doi.org/10.1109/ISCAS.2013.6571900>
- Moogon Jung, Taigon Song, Yarui Peng, and Sung Kyu Lim. 2015. Fine-grained 3-D IC partitioning study with a multicore processor. *IEEE Trans. Comp. Packag. Manuf. Technol.* 5, 10 (Oct 2015), 1393–1401. DOI : <http://dx.doi.org/10.1109/TCPMT.2015.2470124>
- G. Katti, M. Stucchi, K. de Meyer, and W. Dehaene. 2010. Electrical modeling and characterization of through silicon via for three-dimensional ICs. *IEEE Trans. Electron Dev.* 57, 1 (Jan. 2010), 256–262. DOI : <http://dx.doi.org/10.1109/TED.2009.2034508>
- H. Kaul, M. Anders, S. Hsu, A. Agarwal, R. Krishnamurthy, and S. Borkar. 2012. Near-threshold voltage (NTV) design-opportunities and challenges. In *DAC*. 1149–1154.
- M. Konijnenburg, Y. Cho, M. Ashouei, T. Gemmeke, C. Kim, J. Hulzink, J. Stuyt, M. Jung, J. Huisken, S. Ryu, J. Kim, and H. d. Groot. 2013. Reliable and energy-efficient 1MHz 0.4V dynamically reconfigurable SoC for ExG applications in 40nm LP CMOS. In *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*. 430–431. DOI : <http://dx.doi.org/10.1109/ISSCC.2013.6487801>
- Wulong Liu, Haixiao Du, Yu Wang, Yuchun Ma, Y. Xie, Jinguo Quan, and Huazhong Yang. 2013. TSV-aware topology generation for 3D clock tree synthesis. In *2013 14th International Symposium on Quality Electronic Design (ISQED)*. 300–307. DOI : <http://dx.doi.org/10.1109/ISQED.2013.6523626>
- Chien-Wei Lo, Liang Men, J. Brady, and Jia Di. 2015. Asynchronous and synchronous designs for low-power FDSOI CMOS process optimized for subthreshold operation at 0.3V VDD. In *2015 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*. 1–3. DOI : <http://dx.doi.org/10.1109/S3S.2015.7333511>
- OpenSPARC T2 Oracle. 2014. <http://www.oracle.com>. (2014).
- S. K. Samal, Y. Li, G. Chen, and S. K. Lim. 2015a. Improving performance in near-threshold circuits using 3D IC technology. In *2015 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*. 1–2. DOI : <http://dx.doi.org/10.1109/S3S.2015.7333541>
- S. K. Samal, Y. Peng, M. Pathak, and S. K. Lim. 2015b. Ultralow power circuit design with subthreshold/near-threshold 3-D IC technologies. *IEEE Trans. Comp. Packag. Manuf. Technol.* 5, 7 (Jul. 2015), 980–990. DOI : <http://dx.doi.org/10.1109/TCPMT.2015.2441066>
- J. S. Yang, J. Pak, X. Zhao, S. K. Lim, and D. Z. Pan. 2011. Robust clock tree synthesis with timing yield optimization for 3D-ICs. In *2011 16th Asia and South Pacific Design Automation Conference (ASP-DAC)*. 621–626. DOI : <http://dx.doi.org/10.1109/ASPDAC.2011.5722264>
- Bo Zhai, R. G. Dreslinski, D. Blaauw, T. Mudge, and D. Sylvester. 2007. Energy efficient near-threshold chip multi-processing. In *2007 ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*. 32–37. DOI : <http://dx.doi.org/10.1145/1283780.1283789>

Received September 2016; revised December 2016; accepted February 2017