System-Level Comparison of Power Delivery Design for 2D and 3D ICs

Nauman H. Khan, Syed M. Alam*, and Soha Hassoun
Department of Computer Science, Tufts University Medford, MA, *Everspin Technologies, Inc. Austin, TX
nauman@cs.tufts.edu, soha@cs.tufts.edu, *salam@alum.mit.edu

***Abstract*—Three-dimensional integrated circuits (IC) promise high bandwidth, low latency, low device power, and a small form factor. Increased device density and asymmetrical packaging, however, render 3D power delivery design a challenge. In this paper, we provide a system-level comparison of power delivery for 2D and 3D ICs. We investigate various techniques that can impact the quality of power delivery in 3D ICs. These include through-silicon via (TSV) size and spacing, controlled collapse chip connection (C4) spacing, and a combination of dedicated and shared power delivery. Our evaluation system is composed of quad-core chip multiprocessor, memory, and accelerator engine. Each of these modules is running representative SPEC benchmark traces. Our findings are practical and provide clear guidelines for 3D power delivery optimization. More importantly, we show that it is possible to achieve 2D-like or even better power quality by increasing C4 granularity and selecting suitable TSV size and spacing.**

I. Introduction

Three-dimensional (3D) integration is promising technology to design integrated circuits (IC) with higher speed and smaller footprint than the ones designed by the traditional 2D IC design technologies. The length of global wires can be reduced by as much as 50%. The wire-limited clock frequency can be increased by 3.9x [1] and the wire-limited area and power can be reduced by 84% [1] and 51% [2] respectively. Moreover we can integrate heterogeneous technologies to make a system-in-package (SiP) IC**.** Dielectric bonding [3] and metallic bonding [4] are the two common techniques to integrate multiple dies forming a 3D system.

In dielectric (oxide or polymer based) bonding, the vertical connections are completed after the bonding process and through-strata or 3D vias are formed to connect multiple dies. In metallic bonding, the vertical connections are formed by bonding conductive microconnects of Copper (Cu) or Cu with a plating of Tin on each bonding surface. If the bonding orientation is face-to-back, where face refers to the metal interconnect side and back refers to the Si substrate side, a through-silicon via (TSV) is also required to connect a signal to the microconnect. TSVs are filled with metal and separated by dielectric liners from the Si substrate. In this work, we assume a metallic-bonded 3D system where Cu-filled TSVs are used for vertical interconnects in conjunction with the microconnects at the bonded interface as shown in Fig. 1. While this is a critical bonded 3D integration scheme in pursuit by many leading manufacturers, our analysis methods

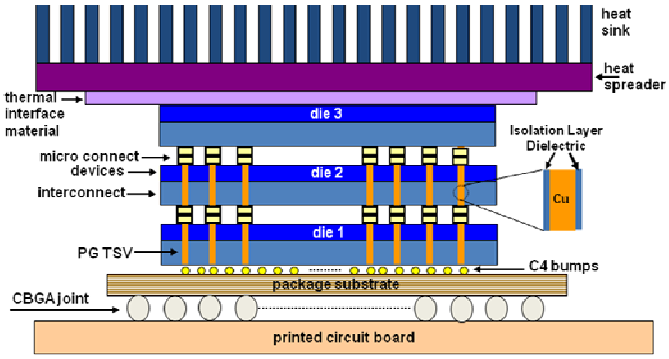


Fig. 1. Illustrative 3D system assuming face-to-back metallic bonding with microconnects.

can be applied to other forms of stacked 3D chips using their parasitic elements for vertical interconnects.

Power delivery in traditional 2D IC design has already become challenging due to increasing operating frequency and power density, and decreasing supply voltage. The average wire length and power dissipation (assuming that the design is interconnect dominated) for a 3D IC drops by a factor of $N^{1/2}$, where N is the number of dies stacked in 3D [5]. Assuming that the power density for each die is same, the power density for the stacked 3D chip will be $N^{1/2}$ times the power density for the corresponding 2D die. The power delivery requirements thus increase with the number of dies in the stack making the problem even harder. The 2008 International Technology Roadmap for Semiconductor (ITRS)[1] predicts that by the year 2015, industry will have 14 and 5 dies stacked in a single package for low cost hand-held and high performance chips, respectively.

In this paper, we study several 3D power delivery configurations with the goal of understanding the major factors that impact 3D power delivery network (PDN) quality. The PDN quality is measured in terms of maximum, average, and standard deviation of IR drop and Ldi/dt droop; metrics that quantify local and global PDN characteristics in both DC and transient analyses. Our evaluation framework consists of a four-core chip multiprocessor (CMP), a memory (MEM), and an accelerator engine (ACCL). We use realistic workloads from SPEC benchmarks for each functional module in the system. We analyze the impact of TSV size and spacing, C4 spacing, and combination of dedicated and shared power

[1] http://www.itrs.net/Links/2008ITRS/Home2008.htm

978-1-4244-4512-7/09/$25.00 ©2009 IEEE

delivery in our 3D PDNs. The major contributions of our work are as summarized below:

- We perform the first comparative study of system-level power delivery for 2D and 3D ICs utilizing realistic workloads, and investigate methods for achieving 2D-like PDN quality in 3D PDNs.
- As TSVs occupy valuable die real-estate, we analyze the impact of TSV size and spacing to optimize area and PDN quality trade-off.
- We summarize our findings in the form of "*Best Practices for 3D PDN Design and Optimization*".

Rest of the paper is organized as follows: In section II, we discuss some of the previous work in 3D power delivery. Experimental setup is described in Section III. In Section IV, we perform the analysis to find the optimal TSV size for 3D PDN. In Section V, we present different comparative studies between 2D and 3D power delivery networks (PDNs). We then present the 3D PDN design guidelines in Section VI, and conclude our work in Section VII.

II. Previous Work On 3D Power Delivery Analysis

The previous work on 3D power delivery can be summarized under two main themes: power delivery techniques and power integrity analysis. Kim et al. analyzed a multi-story power delivery technique [6] where a higher than nominal Vdd supply voltage is applied from the package and distributed differentially to subsequent power rails using level conversion. Their work utilized a lumped off-chip and on-chip models with tungsten filled TSVs in bonded SOI technology to assess the impedance response of overly simplified lumped 2D and 3D PDNs. Yu et al. investigated the impact of via stapling, where a 3D mesh is created, on both power and thermal integrity [7]. Zhan et al. proposed a partition-based algorithm for assigning modules at the floorplanning level to reuse currents between Vdd domains, and to minimize power wasted during circuit operation [8]. In the power integrity analysis area, Huang et al. proposed an analytical physical model of 3D grid models, accurate within 4% compared to SPICE, to capture the impact of power supply noise [9]. Most of these works assumes worst case switching currents and overly simplified power grid network. On the contrary, our work utilizes a more detailed off-chip and on-chip power grid model in a realistic design example where we use a workload derived from SPEC benchmarks. We estimate both IR drop and Ldi/dt droop in 2D and 3D PDNs for comparative analysis, and investigate methods for achieving 2D-like PDN quality in 3D.

III. EXPERIMENTAL SETUP

A. 2D Architecture

We use a conventional single-layer die in a flip-chip package to implement our baseline architecture. The die implements three functional modules; quad-core chip-multiprocessor (PROC), memory (MEM), and accelerator engine (ACCL). Fig. 2 shows the floorplan for our 2D system. Each core of the CMP utilizes 10W of maximum power, and is composed of 5 functional blocks: *fpu* (Floating Point Unit), *ooo* (the rename, reg file, result-bus and window units), *int* (Integer ALU), *fetch* (combines the Icache and branch predictor), and *data* (represents the Data cache and Load-Store Queue). MEM and ACCL modules utilize a maximum of 20W and 10W respectively. We assume that each module has an area of 1cm^2 and the total die area is thus 3cm^2. Both Vdd and Gnd have 16x16 C4 connections per cm^2 which are uniformly distributed over the chip.

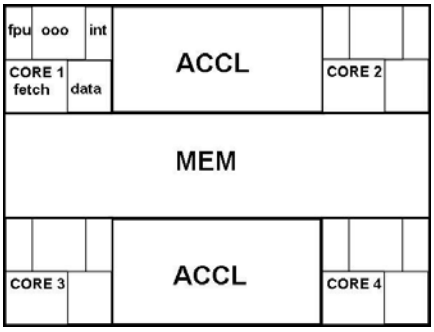


Fig. 2. Floorplan of the 2D Architecture.

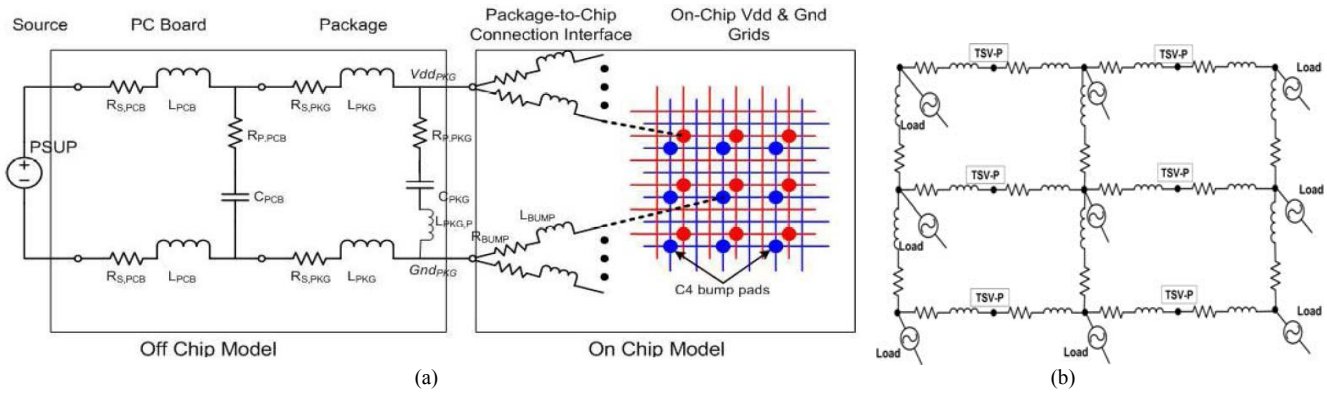
We use an architectural-level power model based on Wattch [10] to estimate the benchmark-specific power dissipation in each functional block. Four SPEC benchmarks (*apsi, bzip, equake,* and *mcf*) were used to collect the power traces. These benchmarks are representative of a wide variety of current patterns [11]. We assume MEM has the same current trace as the L2 Cache, and that the ACCL has the same current trace as the *fpu* block. The Vdd supply voltage is 1.1V.

B. 3D Stacked Architecture

We use an architectural level 3D stack of three dies in facedown orientation as illustrated in Fig. 1. Each of the three functional modules, PROC, MEM, and ACCL, is fabricated on a separate die. We consider the thermal/electrical profiles of the dies while considering there placement in the 3D chip. Since PROC has the highest power consumption, we place it adjacent to the heat sink. We place ACCL farthest from heat sink due to its lowest power consumption and MEM is placed at the center for shorter access paths from both PROC and ACCL. TSVs and microconnects provide connections for power delivery in the stacked chip and are connected to C4 bumps at one side of the stacked 3D chip. Other vertical connection between the dies would exist for inter-layer signal and thermal management. Each die area is 1cm^2. This configuration has 16x16 C4s per cm^2, and a similar number of TSVs for power delivery. We refer to this configuration as a Normal Stacked Configuration (3D NOR) to differentiate it from the other 3D configurations investigated later in the paper.

C. Power Delivery Network (PDN)

We assume both on-chip and off-chip networks to model the PDN for 2D and 3D systems as illustrated in Fig. 3-a. The off-



(a) (b)

Fig. 3. (a) Modeling Power Delivery Network [11]. (b) A portion of the on-chip power grid for each die.

chip network is the RLC circuit representing motherboard and package. We use the RLC values from the Pentium IV processors as presented in [11]. The on-chip network consists of global level grid-like structure routed in top metal layers. The length of a grid element is such that we have a 16x16 element grid in a 1cm^2 area. Load is connected to the middle of each grid element as shown in Fig. 3-b. The TSV-Ps are connected to the C4 bumps either directly or through other stacked layers depending on the position of the on-chip PDN in the stack. We assume wide metal line widths such that the grid collectively occupies 50% of the total die area. We use the predictive technology model (PTM)[2] to calculate the R & L for grid elements. The off-chip and on-chip networks are connected using series resistors and inductors representing the flip-chip package.

The new interconnect elements in on-chip networks for 3D PDN are TSV and microconnect. These elements are modeled as an RC circuit. According to analytical and electrostatic simulation based study by Alam et al. [12], a TSV of width 5µm and height 50µm has resistance and capacitance of 43mΩ and 40fF, respectively, and a microconnect of width 5µm has resistance and capacitance of 40mΩ and 0.4fF, respectively. We use these values and scale accordingly for different TSV/microconnect sizes while the TSV height is fixed at 50µm. Accurate inductance characterization is much complex as it is essential to include a return path which is directly related to the design and layout of specific interface circuitry. However, good news is that TSV inductance is expected to be very low, in the range of 0.3-0.9pH per µm of TSV length [13], especially in comparison to off-chip inductance.

A fast circuit solver, based on preconditioned Krylov subspace iterative methods [14], is used to solve the SPICE netlist for the modeled configuration. A decoupling capacitance of 33nF/cm^2 is assumed in our study, corresponding to device capacitance implementation with 1nm gate oxide thickness (from ITRS roadmap of 90nm-65nm technology) occupying 20% of die area. The decoupling capacitance is uniformly distributed along the grid elements in our 2D and 3D ICs.

[2] http://www.eas.asu.edu/~ptm

IV. OPTIMAL TSV SIZE FOR 3D PDN

In this section, we examine how TSV size impacts 3D power delivery. The maximum IR drops in different dies in the 3D configuration, normalized to the maximum IR drop in the 2D design, are shown in Fig. 4 for TSV sizes ranging from 5µm to 50µm in the 3D NOR configuration. The following observations can be made:

- The maximum IR drop is worse in the 3D configuration with the worst case degradation of as much as 3.4x the 2D IR drop in the PROC die for the smallest considered TSV size.
- The ACCL, in close proximity to the C4 bumps, exhibits nearly constant maximum IR drops across the different TSV sizes.
- More importantly, the IR drop saturates in PROC and MEM for the TSV sizes of and greater than 20µm. Such saturation suggests the lack of benefit of increasing the TSV size beyond a specific value. A TSV size of 25µm is therefore used for all analyses in the following section.

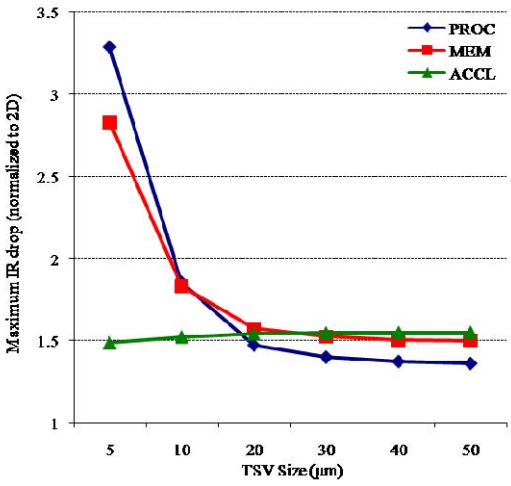


Fig.4. Maximum IR drop for various TSV sizes normalized to the same in 2D design

V. POWER DELIVERY ANALYSIS FOR 3D PDN

A. Normal Stacked Configuration

For static analysis, we remove the inductive and capacitive components in the PDN, and solve the circuit for IR drops for each of the 2048 cycles of each benchmark. The die footprint for the 3D IC is one third of that of the 2D IC. Hence, the number of C4s in the 3D NOR design are reduced accordingly. More current flows through a single C4. We observe the maximum, average, and standard deviation in IR drops, and report the numbers normalized to the values obtained by performing the same analysis for the 2D architecture. The results are presented in the left half of Table I. We observe the followings:

- The 3D NOR power delivery configuration performs worse resulting in higher maximum and in higher average IR drops.
- The ACCL and MEM dies have higher increase in IR drops than the PROC die. This is an important observation that IR drop gets worse even in the die closest to C4 package connections.
- While the standard deviation in IR drop for PROC decreases, the standard deviations in MEM and ACCL increases significantly, indicating a wider distribution of IR drop and resulting in higher variations in power supply.

For Ldi/dt voltage droop analysis, we run the four benchmarks on the 3D NOR architecture. We observe the Ldi/dt voltage droop at each node over the 2048 cycles and report the maximum, average, and standard deviation in each die in Table I (right half of the table). The droop is strongly dependent on each benchmark activity. The followings are major observations from our runs and data in Table I:

- There are few instances where there is a decrease in the maximum Ldi/dt voltage droop due to the fact that the neighborhood in 3D for a node moves from a plane to a cube. A voltage droop at a node in 3D can get current from decoupling capacitors in the vertical neighbors as well as from the ones in the same plane. The resulting behavior is dependent on the locality of the droop as well as the state of the neighboring nodes.
- The decrease in standard deviation also indicates the effect of increased locality for decoupling capacitors in 3D.
- Investigating the general trend using the average voltage droop for the four benchmarks, we notice an increase in voltage droop with a higher impact on the MEM and ACCL dies. The average voltage droop increase is as much as 25% in the ACCL die for the *apsi* benchmark. In a later section, we will investigate the opportunity to mitigate this problem.
- Comparing the IR drop and Ldi/dt voltage droop results in Table I, we notice that 3D stacking has a higher impact on IR drop. 3D stacking inherently increases the resistance of a PDN which directly impacts IR drop. On the other hand, Ldi/dt voltage droop due to the time varying activities in the modules is caused by dominant off-chip inductive components. We, therefore, do not see significant degradation in Ldi/dt voltage droop when compared to the 2D architecture.

TABLE I
IR DROP AND LDI/DT DROOP FOR 3D NOR CONFIGURATION (NORMALIZED TO 2D VALUES).

	IR			LDI/DT		
	PROC	MEM	ACCL	PROC	MEM	ACCL
APSI						
Max.	1.348	1.558	1.457	0.984	1.003	1.029
Avg.	1.338	1.616	1.706	1.164	1.236	1.251
Std.	0.663	2.408	1.248	0.938	0.970	0.966
BZIP						
Max.	1.470	1.596	1.527	1.011	1.136	1.113
Avg.	1.399	1.563	1.639	0.958	0.987	1.018
Std.	0.866	3.393	1.692	0.909	0.964	0.987
EQUAKE						
Max.	1.425	1.542	1.544	0.851	0.984	0.813
Avg.	1.418	1.477	1.738	1.085	1.104	1.120
Std.	0.804	2.773	1.418	0.935	0.933	0.936
MCF						
Max.	1.434	1.604	1.527	1.105	1.130	1.101
Avg.	1.367	1.62543	1.6321	1.105	1.134	1.137
Std.	0.846	3.398	2.101	0.914	0.970	0.959
AVERAGE OF ALL 4 BENCHMARKS						
Max.	1.419	1.575	1.514	0.988	1.064	1.014
Avg.	1.381	1.570	1.679	1.078	1.115	1.131
Std.	0.795	2.993	1.615	0.924	0.959	0.962

B. Effects of TSV Spacing

The PDN in 3D NOR has the same TSV spacing as that of the C4 connections. To devise a 3D stacked configuration that enables increasing the granularity of TSVs for power distribution in any of the dies in 3D stack, we introduce an interposer die [15] between the C4 connections and the bottom die as illustrated in Fig. 5. The interposer die acts as a redistribution layer that is connected to C4 bumps on one side and bonded microconnects (higher granularity) on the other,

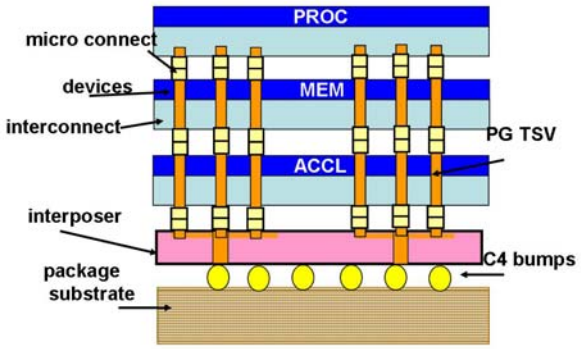


Fig. 5. 3D Stacked Interposer (SI) Configuration.

thus distributing power all the way to the top die via the TSVs in the 3D stack. We can therefore decrease the TSV spacing in the PDN to as low as the minimum allowed microconnect pitch while the C4 pitch can remain unchanged. We refer to this setup as the Stacked Interposer (SI) configuration. For a fair comparison between the 2D and 3D SI configuration, we keep the granularity of C4 bumps and off-chip PDN the same as that in the 2D and the 3D NOR configurations described earlier.

To assess the impact of the TSV spacing on power delivery, we vary the TSV granularity from 16x16 in our 3D NOR configuration to the granularities: 32x32, 48x48, and 64x64. At the highest granularity of 64x64, the TSV spacing is well above the minimum TSV pitch limit of 0.4µm in wafer-to-wafer and 5µm in die-to-wafer or die-to-die 3D bonding technologies [16]. Moreover, the silicon area consumed by TSVs in the 3D PDN for 32x32 and 64x64 granularities are 5% and 20%, respectively, for a TSV size of 25µm. For each increased granularity, the physical dimensions of each grid element are adjusted, and R and L values are re-calculated. The total decoupling capacitance is uniformly redistributed throughout the on-chip grid on each die and remains same throughout.

In Table II, we report the results from static IR drop and transient Ldi/dt voltage droop analysis with various TSV spacing in 3D SI configuration. All data is normalized to those in the 2D architecture. We only report the results from *mcf* benchmark, a representative case for the worst case static and dynamic effects on the 3D PDN. We make the following observations:

- Despite the expectation that the increasing TSV granularity in the 3D PDN would improve the overall quality of power delivery, we notice only marginal improvements in all the metrics for IR drop. The maximum IR drop in the PROC die is improved only 7% by increasing the TSV granularity from 32x32 to 64x64 whereas the silicon area penalty for TSVs rises from 5% to 20%. Similar observations are made for the transient voltage droop where the improvements in maximum and average voltage droop figures are less than 2%.
- The marginal improvement suggests that on-chip grid and TSV granularity of 32x32 reaches a near optimum solution for power grid quality, particularly for IR drops. This observation leads us to consider improving the off-chip network by examining the granularity of C4 bumps, which we explore next.

TABLE II
IR DROP AND LDI/DT VOLTAGE DROOP ANALYSIS FOR DIFFERENT TSV GRANULARITIES IN 3D SI ARCHITECTURE (NORMALIZED TO 2D VALUES).

	IR			LDI/DT		
	PROC	MEM	ACCL	PROC	MEM	ACCL
32 X 32						
Max.	1.144	1.406	1.394	1.075	1.082	1.055
Avg.	1.250	1.544	1.596	1.059	1.103	1.124
Std.	0.606	1.482	1.201	0.897	0.961	0.956
48 X 48						
Max.	1.093	1.363	1.359	1.065	1.073	1.046
Avg.	1.233	1.537	1.597	1.044	1.091	1.115
Std.	0.399	1.043	0.864	0.887	0.953	0.949
64 X 64						
Max.	1.071	1.342	1.339	1.061	1.069	1.042
Avg.	1.221	1.526	1.589	1.038	1.085	1.111
Std.	0.321	0.859	0.710	0.883	0.950	0.946

C. *Effects of C4 Spacing*

We now assume the 3D SI configuration with both C4 and TSV having equal granularity of 32x32 for both Vdd and Gnd supply networks. This is an increase over the 16x16 C4 granularity used earlier. We perform IR drop and Ldi/dt voltage droop analysis, and summarize the results in Table III. We can make following observations from these results:

- Increased C4 granularity results in significant improvement in IR voltage drop. This 4x increase in the number of TSV and C4 results in the 3D PDN performance even better than the equivalent 2D architecture.
- Ldi/dt voltage droop results show that although increasing C4 granularity has significant impact, this impact is not as significant as the one for IR drop. This is due to the fact that the rest of the off-chip PDN components (package and PCB etc.) are still the same.

D. *Effect of Dedicated Power Delivery in 3D*

The experiments in previous sections assume that TSVs in the 3D PDN are shared among all the dies. In this section we study the effect of adding partially dedicated power delivery to each die through a few TSVs connecting to only select dies. We define a new 3D configuration, named tapered stacked (TAP) 3D configuration, shown in Fig. 6.

In a 3D TAP configuration, dies are progressively sized larger to be able to connect few dedicated vertical connections, called boundary vias, to the PDN in the extended boundary portion of a die. As illustrated in Fig. 6, the boundary vias do not pass through any of the active silicon area and can be formed using advanced package-level routing vias similar to those in redistributed chip packaging [17]. Size of each die is modified such that the tapering ratio is constant between the dies and a total silicon area is still $3cm^2$. Due to die resizing,

TABLE III
IR DROP AND LDI/DT VOLTAGE DROOP ANALYSIS FOR 3D SI WITH BOTH C4 AND TSV GRANULARITIES OF 32X32, (NORMALIZED TO 2D VALUES).

	IR			LDI/DT		
	PROC	MEM	ACCL	PROC	MEM	ACCL
Max.	0.731	0.885	0.868	0.970	0.976	0.952
Avg.	0.810	0.992	1.019	0.920	0.958	0.976
Std.	0.339	0.702	0.531	0.943	1.011	1.006

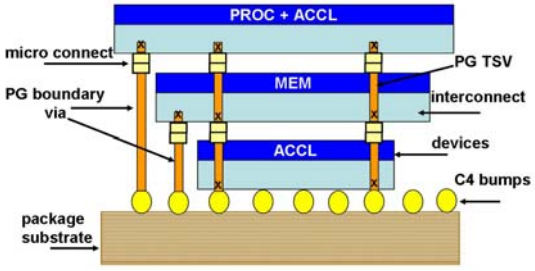


Fig. 6. Tapered Stacked (TAP) 3D Configuration.

we modify the module placements: top die now has PROC and some part of ACCL; middle die has MEM; bottom die has the remaining part of ACCL. All the other off-chip and on-chip parameters remain same.

The results for IR drop and Ldi/dt voltage droop analysis in 3D TAP are presented in Table IV. The results show that partly dedicated power delivery in 3D TAP does not have the same extent of improvement as increasing C4 granularity (comparing results to Table III). However, we noticed that both average IR drop and Ldi/dt voltage droop in 3D TAP are improved compared to those in the 3D NOR (Table I) and the 3D SI (Table II 32x32 TSV granularity case). Although the concept of dedicated or partly dedicated power delivery in 3D is interesting and effectively improves quality of 3D PDNs, there may be an additional risk and considerations associated with non-standard boundary via packaging process. Moreover, the tapered die sizes would only permit die-to-wafer bonding excluding the wafer-to-wafer option which requires same die and wafer sizes. Process considerations aside, the 3D TAP configuration illustrates a method for isolating some of the most active parts of dies by using dedicated delivery in that area. The proposed method also yields improvement in 3D PDN vis-à-vis other 3D PDN configurations. A comparative summary of different 3D PDN configurations is discussed next.

E. Summary of the PDN Studies

We summarize the relative performance of different 3D PDN configurations in the form of graphs presented in Fig. 7 and Fig. 8. Going from Normal to Increased TSV case in Fig. 7 and Fig. 8, we see 8.5% and 4% improvement in IR drop and Ldi/dt droop for the PROC die. This improvement is due to increasing the number of TSVs by 4x. As per results in Section V-B, increasing the number of TSVs by another 4x (16x total) provides additional 3% improvement in IR drop. The trend indicates that further increasing the TSV granularity returns little benefit. Although partially dedicated power delivery does

TABLE IV
IR DROP AND LDI/DT VOLTAGE DROOP ANALYSIS FOR 3D TAP CONFIGURATION (NORMALIZED TO 2D VALUES).

	IR			LDI/DT		
	PROC	MEM	ACCL	PROC	MEM	ACCL
Max.	1.226	1.403	1.344	1.077	1.084	1.057
Avg.	1.196	1.491	1.538	1.031	1.087	1.107
Std.	1.805	2.788	1.432	0.897	0.960	0.954

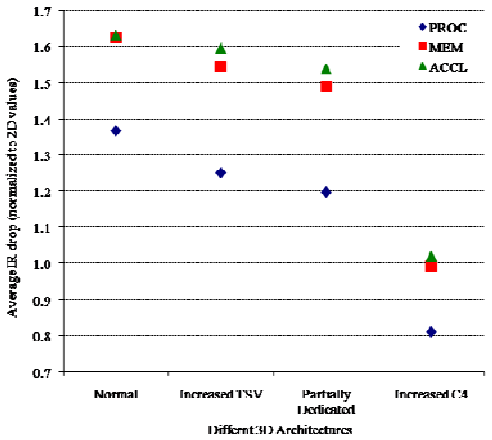


Fig. 7 Comparison of IR Drop from four 3D PDN configurations representing different improvement methods.

not have any area penalty over 3D NOR, it improves average IR drop and Ldi/dt voltage droop by 12% and 7%, respectively, in the PROC die. This is due to the isolation of some of the most active parts of dies by using dedicated delivery in that area. Increased C4 granularity has the maximum impact on the 3D PDN improvement. As shown in Fig. 7 and Fig. 8, a 4x increase in C4 granularity, in the Increased C4 case, provides 41% and 17% relative improvement for IR drop and Ldi/dt voltage droop respectively, when compared to the Normal case.

VI. BEST PRACTICES FOR 3D PDN DESIGN AND OPTIMIZATION

Based on our findings, we present a set of guidelines for designing and optimizing power delivery networks in future 3D designs:

- Locality in the vertical dimension impacts both IR drop and Ldi/dt voltage droop trends in a 3D PDN. A voltage droop at a node in 3D can get current from decoupling caps in the vertical neighbors as well as from the ones in the same plane. The resulting behavior is dependent on the locality of the droop as well as the state of the neighboring nodes. Therefore, a detail 3D PDN analysis with

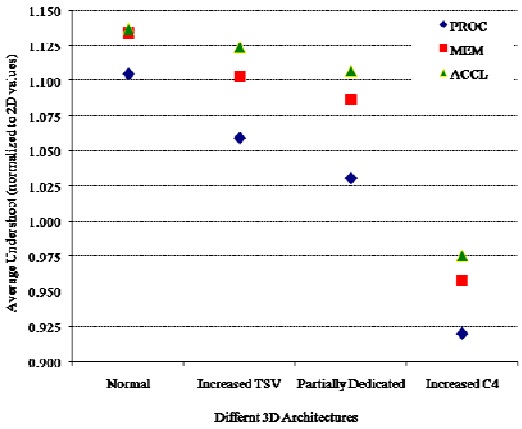


Fig. 8 Comparison of Ldi/dt voltage droop from four 3D PDN configurations representing different improvement methods.

architecture or module level placement using representative workloads is necessary during 3D chip design.

- A critical observation in our work is the saturation trend of IR drop in 3D PDNs with increased TSV size. This suggests the need for first finding the optimal TSV size given the on-chip grids in 3D stacked layers such that the least amount of silicon area penalty is incurred.
- While it is generally expected that the power delivery would be affected most in the die stacked furthest away from the C4 connections, we report that percentage degradation in power delivery is in fact worse in lower level dies closer to C4s. This is particularly true when a highly active module, such as PROC, is placed next to heat sink for thermal concerns and furthest away from C4 connections. Therefore, 3D PDN analysis needs to carefully consider the impact in all the dies while optimizing the grid.
- Increasing the TSV granularity or equivalently decreasing the TSV spacing in 3D PDN improves the standard deviation in IR drop and Ldi/dt voltage droop most, with marginal improvements in maximum and average values. Therefore, physical design for 3D PDN must consider this impact and choose TSV granularity accordingly.
- Despite selecting the optimal TSV size and TSV spacing, 3D PDN performs worse in both IR drop and Ldi/dt voltage droop compared to 2D PDN if the package connection, such as C4, pitch or granularity is maintained the same as in the 2D case. Our study shows that improving off-chip component of the 3D PDN, for example through reducing C4 pitch for a higher number of C4s, has the highest relative impact on power grid metrics that enables 2D like or even better quality 3D PDN.
- A combination of shared and dedicated TSV power delivery can be used, as illustrated in 3D TAP configuration, to achieve improvements in both IR drop and Ldi/dt voltage droop.

VII. CONCLUSION

Power delivery is expected to be a major physical design concern in the 3D ICs due to higher power density and package asymmetries. In this paper, we compared the power delivery networks for 2D and various 3D configurations; 3D NOR, 3D SI, and 3D TAP, that represent different techniques for improving power delivery in 3D. We performed the first detailed architectural analysis to study the IR drop and Ldi/dt voltage droop in the context of various design parameters in 3D PDNs: TSV size, TSV and C4 granularity, and partially dedicated TSV. Interestingly, it is possible to achieve 2D-like or even better power delivery by improving the off-chip component (C4 granularity) in a 3D PDN. Based on our findings, we presented a set of design and analysis guidelines for 3D PDNs.

REFERENCES

[1] J. Joyner, R. Venkatesan, P. Zarkesh-Ha, J. Davis, and J. Meindl, "Impact of Three-Dimensional Architectures on Interconnects in Gigascale Integration," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 9, 2001, pp. 922-928.

[2] M. Bamal, S. List, M. Stucchi, A. Verhulst, M. Van Hove, R. Cartuyvels, G. Beyer, and K. Maex, "Performance Comparison of Interconnect Technology and Architecture Options for Deep Submicron Technology Nodes," *Interconnect Technology Conference, 2006 International*, 2006, pp. 202-204.

[3] MIT Lincoln Laboratory "MITLL Low-Power FDSOI CMOS Process Application Notes" [Online]. Available: http://www.ece.umd.edu/~dilli/research/layout/MITLL_3D_2006/3D_PDK2.3/doc/ApplicationNotes2006-1.pdf

[4] P. Morrow, M.J. Kobrinsky, S. Ramanathan, C.M. Park, M. Harmes, V. Ramachandrarao, H. mog Park, G. Kloster, S. List, and S. Kim, "Wafer-Level 3D Interconnects via Cu Bonding," *Proceedings of the 21st Advanced Metallization Conference* October 2004.

[5] W. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. Sule, M. Steer, and P. Franzon, "Demystifying 3D ICs: The Pros and Cons of Going Vertical," *Design & Test of Computers, IEEE*, vol. 22, 2005, pp. 498-510.

[6] P. Jain, T. Kim, J. Keane, and C.H. Kim, "A Multi-Story Power Delivery Technique for 3D Integrated Circuits," *Proceeding of the thirteenth international symposium on Low power electronics and design*, Bangalore, India: ACM, 2008, pp. 57-62.

[7] H. Yu, J. Ho, and L. He, "Simultaneous Power and Thermal Integrity Driven Via Stapling in 3D ICs," *Proceedings of the 2006 IEEE/ACM international conference on Computer-aided design*, San Jose, California: ACM, 2006, pp. 802-808.

[8] Y. Zhan, T. Zhang, and S.S. Sapatnekar, "Module Assignment for Pin-Limited Designs Under the Stacked-Vdd Paradigm," *Proceedings of the 2007 IEEE/ACM international conference on Computer-aided design*, San Jose, California: IEEE Press, 2007, pp. 656-659.

[9] Gang Huang, M. Bakir, A. Naeemi, H. Chen, and J. Meindl, "Power Delivery for 3D Chip Stacks: Physical Modeling and Design Implication," *Electrical Performance of Electronic Packaging, 2007 IEEE*, 2007, pp. 205-208.

[10] D. Brooks, V. Tiwari, and M. Martonosi, "Wattch: A Framework for Architectural-Level Power Analysis and Optimizations," *Computer Architecture, 2000. Proceedings of the 27th International Symposium on*, 2000, pp. 83-94.

[11] M. Gupta, J. Oatley, R. Joseph, Gu-Yeon Wei, and D. Brooks, "Understanding Voltage Variations in Chip Multiprocessors using a Distributed Power-Delivery Network," *Design, Automation & Test in Europe Conference & Exhibition, 2007. DATE '07*, 2007, pp. 1-6.

[12] S.M. Alam, R.E. Jones, S. Rauf, and R. Chatterjee, "Inter-Strata Connection Characteristics and Signal Transmission in Three-Dimensional (3D) Integration Technology," *Proceedings of the 8th International Symposium on Quality Electronic Design*, IEEE Computer Society, 2007, pp. 580-585.

[13] J. Wu and J. del Alamo, "Through-Substrate Interconnects for 3D ICs, RF Systems, and MEMS," *2007 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems, 10-12 Jan. 2007*, Piscataway, NJ, USA: IEEE, 2006, p. 4 pp.

[14] T. Chen and C.C. Chen, "Efficient Large-Scale Power Grid Analysis Based on Preconditioned Krylov-Subspace Iterative Methods," *Proceedings of the 38th conference on Design automation*, Las Vegas, Nevada, United States: ACM, 2001, pp. 559-562.

[15] C.S. Patel, "Silicon Carrier for Computer Systems," *Proceedings of the 43rd annual conference on Design automation*, San Francisco, CA, USA: ACM, 2006, pp. 857-862.

[16] J. UKnickerbocker et al."Three-Dimensional Silicon Integration," *IBM Journal of Research and Development*, vol. 52, Nov. 2008, p. 553.

[17] "Redistributed Chip Packaging (RCP) Technology." [Online] Available: http://www.freescale.com/webapp/sps/site/overview.jsp?code =ASIC_LV3_PACKAGING_RCP