Journal of information and communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 13(3): 172-179, Sep. 2015 Regular paper

Die-to-Die Parasitic Extraction Targeting Face-to-Face Bonded 3D ICs

Taigon Song and Sung Kyu Lim*, ***Member, KIICE***

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Face-to-face (F2F) bonding in three-dimensional integrated circuits (3D ICs), compared with other bonding styles, is closer to commercialization because of its benefits in terms of density, yield, and cost. However, despite the benefits that F2F bonding expect to provide, it's physical nature has not been studied thoroughly. In this study, we, for the first time, extract cross-die (inter-die) parasitic elements from F2F bonds on the full-chip scale and compare them with the intra-die elements. This allows us to demonstrate the significant impact of field sharing across dies in F2F bonding on full-chip noise and critical path delay values. The baseline method used is the die-by-die method, where the parasitic elements of individual dies are extracted separately and the cross-die parasitic elements are ignored. Compared with this inaccurate method, which was the only method available until now, our first-of-its-kind holistic method corrects the delay error by 25.48% and the noise error by 175%.

Index Terms: 3D IC, Capacitance, Coupling, Face-to-face (F2F), Full-chip

I. INTRODUCTION

Face-to-face (F2F) bonding is a bonding style in three-dimensional integrated circuits (3D ICs) where two dies are bonded on their top-metal surfaces. Thus far, many researchers have reported its advantages [1]. Thus, F2F bonding is now considered a promising technology that can provide better power/performance with denser I/Os. To provide high-density I/Os in F2F bonding, two important technologies must be scaled down: bump diameter and chip-to-chip height. With respect to the bump diameter, several researchers have reported F2F bumps as small as 1.6 μm [2] and a chip-to-chip distance (the gap between tiers) as small as 1 μm [3]. Above all, F2F bonding with more than 3000 I/O pads using these small bumps (<5 μm) has proven to be reliable [4], showing the possibility for mass production. Thus, we anticipate that F2F bonding with relatively small bumps and a short chip-to-chip distance will be the key technology for dense I/Os. However, as technology scales for F2F bonding, one of the issues that must be studied is coupling. Inter-die parasitic elements are almost negligible when the distance between the dies is more than few tens of microns. However, since this distance has scaled down to less than a few microns, the inter-die parasitic elements in F2F bonding will be in a non-negligible range. Despite such significance, no studies have reported the significance of F2F coupling or the impact of F2F coupling on system performance. Previous studies on F2F 3D designs extracted the parasitic elements of each die separately and then stitched them together, assuming that the impact of inter-tier coupling was not significant [1].

In this research, we study the impact of inter-tier parasitic

Received 20 March 2015, **Revised** 09 April 2015, **Accepted** 11 May 2015
***Corresponding Author** Sung Kyu Lim (E-mail: limsk@ece.gatech.edu, Tel: +1-404-894-0373)
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

Open Access http://dx.doi.org/10.6109/jicce.2015.13.3.172 print ISSN: 2234-8255 online ISSN: 2234-8883

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ⓒ The Korea Institute of Information and Communication Engineering

elements in F2F-bonded 3D ICs. We provide a methodology for extracting both intra-die and inter-die parasitic elements in a single run on the full-chip scale. Then, we analyze how significant the impact of the F2F parasitic elements is. The main contributions of this work include the following. 1) To the best of our knowledge, we are the first to provide a holistic methodology for designing full-chip-level F2F-bonded 3D ICs and extracting their parasitic elements. Integrated with commercial computer-aided design (CAD) tools, our methodology facilitates a sign-off quality timing/power analysis. Field-solver-based tools may be able to extract segments of F2F-bonded structures. However, on the full-chip scale, this is the first work providing such a methodology and visualizing the full-chip impact using various metrics. 2) We reveal that F2F bonding leads to significant inter-die capacitance and a considerable reduction in the top-metal-to-top-metal capacitance in the same die. 3) F2F bonding causes a major timing/noise error on single nets. However, its impact on the total power consumption is minor. 4) The power distribution network (PDN) significantly reduces the F2F capacitance between tiers.

II. MOTIVATION

For dense I/Os in F2F-bonded systems, smaller bumps at a lower chip-to-chip height (=H_{C2C}) are inevitable. If the bump size scales but H_{C2C} does not, the aspect ratio (height) of the bumps increases, causing yield problems. However, closer H_{C2C} introduces inter-die capacitance, which is significant in advanced interconnect technologies. Fig. 1 illustrates our motivation. The two boxes on the bottom (B) and the top (T) represent the top metal of the bottom tier and that of the top tier, respectively. All metal layers have a width/spacing/thickness of 1.8 μm/1.8 μm/2.8 μm that represents an industrial interconnect of the top metal. We use Synopsys Raphael for our simulations.

In Fig. 1(a), capacitance is generated only between the same tier (C_H, intra-die capacitance) because the distance between the two dies is significantly large. In Fig. 1(b), when H_{C2C} = 10 μm, inter-die capacitances C_{V1} and C_{V2} are generated between tiers. Here, C_{V1} and C_{V2} are relatively small compared to C_H. However, in Fig. 1(c), when H_{C2C} becomes very low (H_{C2C} = 1 μm), C_{V1} is larger than C_H (4.59 fF > 3.45 fF), implying that the inter-tier capacitance becomes significant as H_{C2C} scales. In addition, notice that C_H decreases from 5.4 fF to 3.45 fF. This can be attributed to the E-field sharing between the top tier and the bottom tier. When new aggressors (e.g., top-to-bottom) come close to the original aggressors (e.g., bottom-to-bottom) as shown in Fig. 1(b) and (c), the E-field distributes from the original aggressors to the new aggressors because of the change in the distance. Thus, C_H decreases and C_V increases. Thus, we

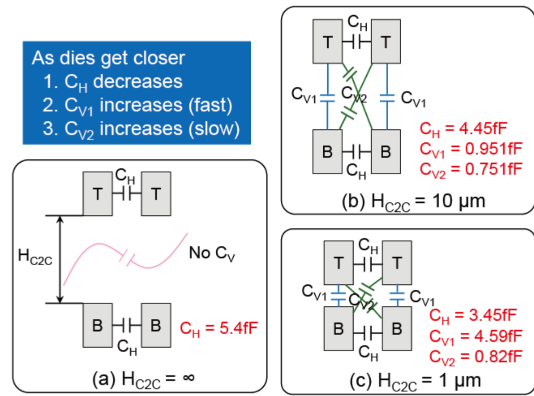


Fig. 1. Capacitance change with chip-to-chip distance change from ∞ to 1 μm. Metal dimensions: width = 1.8 μm, pitch = 1.8 μm, thickness = 2.8 μm. C_H and C_V denote the horizontal and vertical capacitances, respectively.

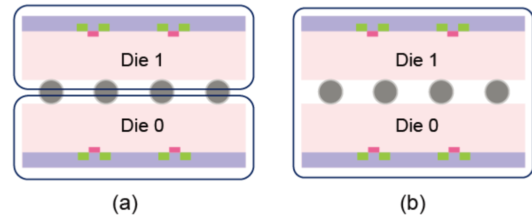


Fig. 2. Two capacitance extraction methodologies: (a) die-by-die extraction and (b) the proposed holistic extraction.

conclude that 1) C_V increases as H_{C2C} scales. In particular, C_V becomes significant when H_{C2C} scales to the most advanced F2F bonding technologies (e.g., H_{C2C} = 1 μm). 2) C_H decreases with a decrease in H_{C2C}.

Conventional (=die-by-die) parasitic extraction extracts the intra-die parasitic elements from each die and then stitches them together as shown in Fig. 2(a) [1]. However, if die-by-die extraction is done in 3D designs where H_{C2C} is small, C_H is significantly overestimated. Comparing Fig. 1(a) and (c), we find that C_H is overestimated by 56.5%. In addition, die-by-die extraction cannot extract C_V, which can become larger than CH. Thus, F2F parasitic elements should be extracted in a holistic manner, as shown in Fig. 2(b).

III. EXTRACTION ANALYSIS

We use Synopsys 28 nm as our baseline process design kit (PDK) [6]. Table 1 describes the two different interconnect structures that we use. We will refer to these structures as Type 1 (thick) and Type 2 (regular), respectively. Both Type 1 and Type 2 structures consist of six metal layers. The Type 1 structure uses an M6 width/thickness of 1.8 μm/2.8 μm, and the Type 2 structure uses

Table 1. Interconnect dimensions used in our design

	Width (μm)	Spacing (μm)	Pitch (μm)	Thickness (μm)	Diel (μm)
Type 1 (thick)					
M6	1.8	1.8	3.6	2.8	–
M5	0.36	0.36	0.72	0.85	0.8
Type 2 (regular)					
M6	0.36	0.36	0.72	0.85	0.8
M5	0.224	0.236	0.46	0.38	0.38
Common in Type 1 and Type 2					
M4	0.112	0.116	0.228	0.19	0.19
M3	0.056	0.056	0.152	0.095	0.09
M2	0.056	0.056	0.152	0.095	0.09
M1	0.05	0.05	0.152	0.095	0.09

an M6 width/thickness of 0.36 μm/0.85 μm. For M5, each width/thickness is smaller than M6 and is scaled accordingly on the basis of the M6 width/thickness used. Note that the top metal in both these types of structures represents the dimensions of the actual industrial 28-nm interconnects. For M4 to M1, we follow the interconnects of the Synopsys 28-nm PDK and use the same for both Type 1 and Type 2 structures. For the 3D stack-up, our F2F bump diameter is 1.6 μm [2] and the chip-to-chip distance is 1.5 μm [3]. We assume that when an F2F design is completed in Type 1 (or Type 2), both dies will have the same Type 1 (or Type 2) interconnect structure.

Fig. 3 shows the proposed extraction and analysis flow. First, we partition a two-dimensional (2D) netlist into two tiers and perform placement on each die. Our placer is based on a force-directed 3D gate-level placement engine [7], and it is modified accordingly to perform placement in our F2F design flow. This gives us the placement results for the two tiers (Die0.def and Die1.def). Once the placement is done, we use our F2F layer generator for generating a two-tier holistic F2F stack for routing and extraction. First, our F2F layer generator assigns the standard cells on the top (Die 1) and the bottom (Die 0) of the stack by using the placement from the previous step (Die0.def and Die1.def). Second, the F2F layer generator creates a platform that models all metal layers of both dies and the F2F interface in a holistic manner for the interconnects. Based on our platform, a holistic full-chip F2F-bonded 3D design (f2f.def) can be created in Cadence SoC Encounter (a commercial P&R tool) for full-chip F2F design and impact study.

Given our 3D F2F platform, we use Synopsys StarRC and extract both intra-tier and inter-tier (F2F) parasitic elements in just one run (.SPEF). Further, the fact that our platform is developed using 2D CAD tools does not deteriorate the accuracy of the F2F extraction results because StarRC is a

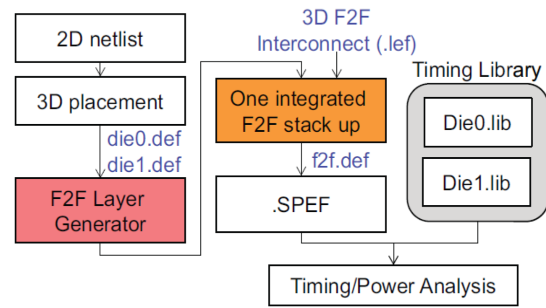


Fig. 3. Proposed extraction flow using our face-to-face (F2F) layer generator.

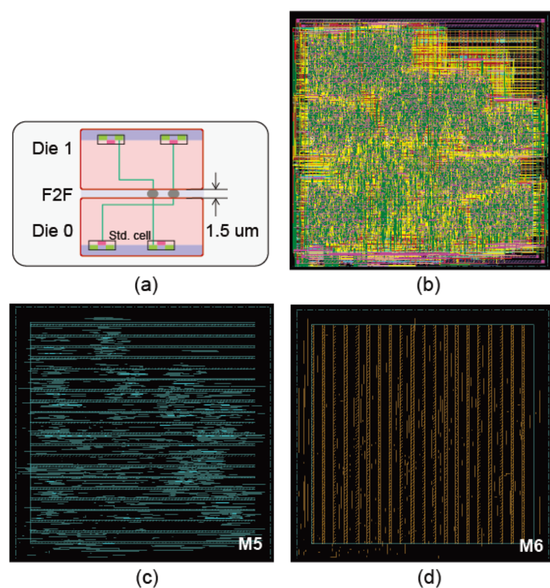


Fig. 4. (a) Face-to-face (F2F) stack-up created by our F2F layer generator. (b) One integrated full-chip layout in Cadence Encounter with power distribution network (PDN). (c) Die_0–M5 with PDN. (d) Die 0 M6 layout with PDN (benchmark: AES).

3D-based EM solver. As long as we feed the correct details of the full-chip F2F design that we have into the solver, the results of our holistic extraction are accurate in the commercial grade. In addition, commercial tools have been proven to be accurate for the extraction F2F parasitic elements [8]. Fig. 4(a) shows the results of our F2F layer generator, and Fig. 4(b) shows a layout shot of the final result (benchmark: AES) after the completion of the 3D design. Once the parasitic elements are extracted, we provide the timing/power library of the standard cells in each die (Die0.lib and Die1.lib) and perform a timing/power analysis by using Synopsys PrimeTime. To focus on visualizing the impact of F2F capacitance on circuits, we do not include any I/Os in our designs.

Next, we will introduce the new capacitances formed in F2F 3D ICs. We define these inter-tier capacitances as

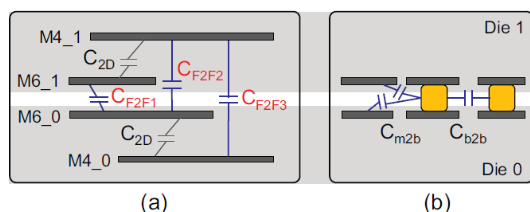


Fig. 5. F2F (3D) capacitances in F2F bonding. (a) Metal-to-metal capacitance. (b) Bump capacitance.

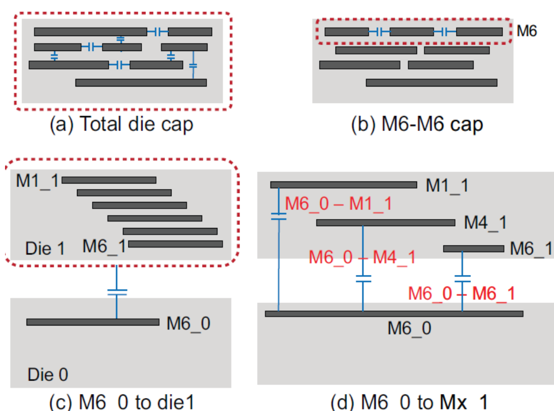


Fig. 6. Parasitic capacitance definitions. (a) Total die capacitance, (b) M6-M6 capacitance, (c) M6 (Die 0) to Die 1 capacitance, and (d) M6_0 (Die 0) to Mx_1 (Die 1) capacitance.

'F2F (3D) capacitances,' and intra-tier capacitances as '2D capacitances.' Fig. 5(a) shows these F2F capacitances when there are no bumps between the top metals of the chip. Note that F2F capacitances are generated not only between the top metal layers (C_{F2F1}) but also between other metal layers (C_{F2F2} and C_{F2F3}). In addition, F2F capacitance not only consists of the inter-metal capacitance but also the capacitance from the bumps to the other structures—bump capacitance, Fig. 5(b). Bump capacitance is of the following two types: bump-to-bump capacitance (C_{b2b}) and metal-to-bump capacitance (C_{m2b}).

We report how significant F2F capacitance is to the other capacitances in the LDPC benchmark. To explain this, we report three capacitances for comparison: total coupling capacitance inside a die (=total die cap) as shown in Fig. 6(a), M6-to-M6 coupling capacitance generated inside the same die (=M6-M6 cap) as shown in Fig. 6(b), and total F2F capacitance generated between the two dies (F2F cap). Table 2 shows our results. We will now explain the results for the Type 1 structure and then for the Type 2 structure. The total F2F capacitance is 259.17 fF. Note that this is a significant value and cannot be extracted by die-by-die extraction.

We note the following points. 1) The total coupling capacitance formed in a die is 38738 fF, and compared with this, the F2F capacitance is only 0.67% of the capacitance formed in a single die. 2) However, M6-M6 capacitance in the same die is 451.06 fF. Compared with this, the F2F capacitance is 57.5% of the M6-M6 capacitance. 3) Bump capacitance (116.54 fF = Cb2b + Cm2b) comprises a significant portion of the F2F capacitance. We see a similar trend in the Type 2 structure. The F2F capacitance is 0.41% of the capacitance generated in a single die, but it is 61.7% of the M6-M6 capacitance. The bump cap is also noticeable, which is 26.4% of the F2F cap. The ratio of the bump cap to the total F2F cap in the Type 1 structure is larger than that of the Type 2 structure. Since the metal dimensions of the Type 1 structure are significantly larger than those of the Type 2 structure (Type 1 M6 is 3.3× thicker and 5× wider than Type 2 M6), Cm2b in the Type 1 structure is greater than that in the Type 2 structure. In brief, the F2F capacitance contributes significantly to the total capacitance, and this impact should not be ignored.

Table 2. Comparison of face-to-face (F2F) capacitance to the other capacitances

	Type 1	Type 2
Total die cap (fF)	38738.19	38209.93
M6-M6 cap (fF)	451.06	252.11
F2F cap (fF)	259.17	155.49
F2F % to M6-M6 cap	57.5	61.7
Bump cap (fF)	116.54	41.14

Total die cap and M6-M6 cap are averaged between Die 0 and Die 1 (see Fig. 6(a) and (b) for definitions).

Table 3. F2F capacitance breakdown (see Fig. 6(c) for definitions)

	Type 1		Type 2	
	Total cap (fF)	% to total F2F cap	Total cap (fF)	% to total F2F cap
Die 0–Die 1	259.17	–	155.49	–
M6_0–Die 1	255.64	**98.6**	151.09	**97.1**
M6_0–M6_1	252.06	**97.2**	146.60	**94.2**

We now break down the two types of F2F capacitance. First, we measure the capacitance from one metal (on Die 0) to the other die (Die 1). For example, 'M6_0–Die 1' denotes the total capacitance formed between M6 (in Die 0) and all other metal layers in Die 1 (see Fig. 6(c)). Table 3 shows that most of the F2F capacitance is generated between the top metal (M6) and the other die in both types of structures (98.64% in the Type 1 structure and 97.17% in the Type 2 structure). Second, we measure the F2F capacitance between the metal layers. We see that most of the capacitance is generated between the top metal layers of each die

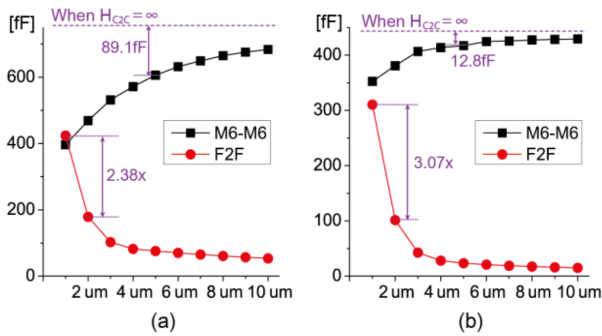


Fig. 7. Face-to-face (F2F) capacitance in different chip-to-chip distance. (a) Type 1 and (b) Type 2. See Table 1 for interconnect dimensions.

(M6_0–M6_1: more than 90%, see Fig. 6(d) for definitions) in both types of structures. This makes sense because M6 is the thickest among all the metal layers, and M6 shields the inter-tier E-field that tries to generate capacitance between the other metal layers. In short, most of the F2F capacitance is generated between the top metal layers.

We verify our motivation discussed in Section II on the full-chip scale. We measure the M6-M6 and M5-M5 capacitances (in the same die) and compare the two extraction methods (die-by-die and holistic). Table 4 shows the results. In both Type 1 and Type 2 structures, we report that the die-by-die extraction overestimates the M6-M6 capacitance significantly (56.2% in the case of the Type 1 structure and 55.4% in the case of the Type 2 structure) because of the inter-tier E-field sharing. We report that 1) the M6 capacitance is significantly overestimated in the die-by-die extraction when the inter-tier interaction between metals is not considered in the F2F designs. Note that when the distance between the tiers decreases, the F2F capacitance (C_V) increases (see Fig. 1) and, at the same time, the capacitance between metals in the same tier (C_H) decreases. 2) The capacitance is significantly overestimated in M6 but not in M5. In short, F2F bonding causes a significant capacitance reduction in the top metal but has an almost negligible impact on the metal below.

Fig. 7 shows how the capacitances change when the chip-to-chip distance (H_{C2C}) changes from 1 μm to 10 μm in both Type 1 and Type 2 structures. It also reports the change in the M6-M6 capacitance in the same die. In both interconnect types, the F2F capacitance converges to 0 and the M6-M6 capacitance saturates to the die-by-die-extracted value as the distance increases ($H_{C2C} = \infty$). First, in the Type 1 structure, the M6-M6 capacitance reduction shows a steeper slope and starts changing more even at a large F2F distance than in the Type 2 structure. For example, when H_{C2C} = 5 μm, the Type 1 structure shows a –89.1-fF reduction, while the Type 2 structure shows only a –12.8-fF reduction.

Table 4. Capacitance overestimation in die-by-die extraction because of the face-to-face (F2F) cap in the LDPC benchmark

	Type 1		Type 2	
	M6-M6	M5-M5	M6-M6	M5-M5
Holistic (fF)	451.06	2890.5	252.11	1875.8
Die-by-die (fF)	702.67	2870.9	392.77	1882.2
Error (%)	56.2	–0.7	55.4	0.3

Table 5. Full-chip timing and noise analysis in the LDPC benchmark

	Holistic	Die-by-die	Δ	Δ (%)
Type 1. Net: decoded_block_1666				
Cap (fF)	188.902	173.232	15.67	9.05
Tran. time (ns)	0.150	0.141	0.014	11.02
Delay (ns)	0.045	0.038	0.007	18.42
Noise (V)	0.11	0.06	0.05	83.33
Type 2. Net: decoded_block_2_				
Cap (fF)	123.025	118.155	4.87	4.12
Tran. time (ns)	0.132	0.124	0.008	6.45
Delay (ns)	0.034	0.031	0.003	9.68
Noise (V)	0.0345	0	0.0345	New

Comparing the two interconnect types, we find that the Type 1 M6 has a wider pitch (3.6 μm) than the Type 2 M6 (0.72 μm). Because of this, M6-M6 is relatively loosely coupled (as compared to that in the Type 2 structure) in terms of the E-field strength. Therefore, F2F coupling starts affecting capacitance even from a far distance. Further, the Type 1 structure has an F2F distance to metal pitch ratio of 1.38× (5/3.6) and the Type 2 structure has a ratio of 6.94×. This indicates that the relative F2F distance of the Type 1 structure is 5× closer than that of the Type 2 structure. This is why the M6-M6 capacitance drops faster in the Type 1 structure.

Second, the F2F capacitance increase at a shorter distance (1–2 μm) is higher in the Type 2 structure (3.08×). Type 2 designs are always packed with more M6 objects than the Type 1 designs because of the closer metal pitch in the same area. Therefore, when the chip-to-chip distance is shorter than a certain value where its capacitance increase ratio is significantly high (e.g., 2 μm to 1 μm), the Type 2 structure shows a higher F2F capacitance because it has more M6 objects than the Type 1 structure to generate the capacitance. In fact, note that when H_{C2C} = 1 μm, the F2F capacitance is significant in both types of structures. This implies that the F2F-bonded 3D ICs will suffer more from the F2F capacitance at shorter chip-to-chip distances.

Table 6. Results for all benchmarks

	Area (μm × μm)	M6-M6 cap (fF)	F2F cap (fF)	F2F % to M6-M6 cap	Bump cap (fF)	M6 Δ (%)	Cap Δ (fF)	Delay Δ (ps)	Noise Δ (mV)
Type 1 interconnect (thick)									
LDPC	700 × 900	451.06	259.17	57.5%	116.54	56.2%	15.67	7.0 (18.4%)	50.0 (83.33%)
AES	150 × 150	141.67	39.18	27.7%	36.07	39.0%	1.91	1.6 (12.6%)	51.0 (31.25%)
VGA	170 × 170	38.30	10.30	26.9%	6.82	45.8%	0.73	0.76 (13.2%)	32.0 (27.27%)
JPEG	700 × 900	557.36	395.01	70.9%	99.35	43.6%	14.45	5.4 (9.2%)	137.5 (175.0%)
M256	900 × 1100	353.99	371.04	104.8%	50.16	53.4%	10.64	13.9 (15.2%)	36.3 (34.78%)
Avg.	–	–	–	57.56%	–	47.6%	–	(13.7%)	(70.33%)
Type 2 interconnect (thin)									
LDPC	700 × 900	252.11	155.49	61.7%	41.14	55.4%	4.87	9.7 (9.6%)	34.5 (NEW)
AES	150 × 150	60.81	25.91	42.6%	17.83	14.9%	0.51	1.1 (17.5%)	10.0 (20.01%)
VGA	170 × 170	9.78	8.26	84.5%	2.52	14.3%	1.16	0.3 (8.8%)	10.1 (19.99%)
JPEG	700 × 900	345.50	232.66	67.3%	57.05	14.8%	4.65	11.2 (25.4%)	40.2 (New)
M256	900 × 1100	291.32	236.55	81.2%	24.56	26.3%	12.39	15.2 (5.8%)	30.0 (28.57%)
Avg.	–	–	–	67.46%	–	25.14%	–	(13.4%)	(22.85%)

M6 Δ denotes the cap overestimation caused by die-by-die extraction between M6-M6 in the same die as in Table 4. Further, cap Δ, delay Δ, and noise Δ are the worst-case underestimation differences observed in the timing and noise analysis of a single net in die-by-die extraction, as shown in Table 5.

IV. FULL-CHIP TIMING/NOISE IMPACT

Using the flow from Section III, we use Synopsys PrimeTime for the timing and noise analysis. We perform a static timing analysis (STA) on the basis of the clock frequency of the benchmarks. We analyze the timing and noise results of both the die-by-die extraction and the proposed holistic extraction, and compare each net. Then, we report the worst case nets that show the most discrepancy in terms of the capacitance. Table 5 reports the delay and noise of a net in LDPC when M6 wires are used for its routing. First, for both interconnect types, the die-by-die extraction underestimates the capacitance of a net significantly. In the Type 1 structure, the F2F capacitance is underestimated by 15.67 fF, and this is a 9.05% difference from the value estimated by using the proposed holistic extraction. Because of this, the die-by-die extraction underestimates the transition time and delay by 11.02% and 18.42%, respectively. Similarly, in the Type 2 structure, the capacitance is underestimated by 4.87 fF, and, because of this, both the transition time and the delay are significantly underestimated. Consider that a net on the critical path or a clock net uses the top metal in the F2F design. These nets will have a significant timing error because of the underestimation in the die-by-die extraction, which designers cannot tolerate. Second, die-by-die extraction leads to an inaccurate noise analysis. In the Type 1 structure, the noise voltage of a net is underestimated by 50 mV, and this is 83.3% of the noise missed in the die-by-die method. In the Type 2 structure, die-by-die extraction does not find any effective aggressors near the victim net. However, holistic extraction finds the inter-tier aggressors that die-by-die extraction misses and provides accurate results. Misalignment and process variation between dies (in both the X–Y and the Z direction) causes a significant 3D capacitance change. Stemming from this change, these variations will cause an additional timing/noise error on each net.

The total power consumption by the two different extraction methods is almost the same. For example, Type 1 LDPC consumes 49.5 mW in die-by-die extraction and 49.7 mW in holistic extraction. Type 2 LDPC consumes 49.0 mW in die-by-die extraction and 49.1 mW in holistic extraction. This is a less than 1% difference. This difference can be attributed to the following. (1) Despite the increase in the F2F capacitance due to F2F bonding, the intra-die capacitance (C_H in Fi. 1) decreases at the same time. (2) In terms of the total capacitance in the full-chip, the portion that the F2F capacitance contributes is very small. In addition, since the M6-M6 capacitance decreases at the same time, the total capacitance difference between die-by-die extraction and holistic extraction at the full-chip level is almost negligible (less than 0.1% in total). The dynamic power in digital circuits can be expressed by the following equation:

$$P_{olymer} = CV_{DD}^2 f_{sw} ,$$

where C denotes the capacitance, VDD represents the supply voltage, and f_{sw} indicates the operating frequency.

Since the change in the total capacitance is less than 0.1% in total, which is the only changing parameter between the two extraction methods, the power difference from die-by-die extraction and holistic extraction is almost negligible.

We use five benchmarks (including LDPC) to estimate the impact of F2F parasitic elements in various full-chip designs [5]. The biggest benchmark JPEG consists of 226 K cells, which is more than 1 M transistors, and the smallest benchmark VGA consists of 5.5 K cells. Benchmarks are sized optimally to perform routing without any violations. Table 6 reports our comprehensive results. Through many benchmarks, we report that 1) the portion of F2F capacitance in the M6-M6 capacitance is significant (>67% on average in Type 2 structures), and the bump cap is a big contributor to the total F2F cap. 2) Die-by-die extraction significantly overestimates the M6-M6 capacitance (M6 error, >47% on average in Type 1 structures) but not much on other layers. 3) PDN reduces the F2F capacitance significantly (>47% on average in Type 2 structures). 4) The capacitance error on nets occurs in full-chip designs when using die-by-die extraction. Because of this, the under-estimated total capacitance causes significant timing (25.48%) and noise (175%) errors on the nets.

V. CONCLUSION

In this paper, we proposed a holistic parasitic extraction methodology for F2F-style 3D ICs. We found that these inter-tier parasitic elements become non-negligible, and these parasitics significantly change the capacitance values of the top metal on each die. We demonstrated that a shorter F2F distance causes a significant error in the M6-M6 capacitance (56.2% in LDPC) and a considerable increase in various inter-tier capacitances that the die-to-die extraction cannot extract (104.8% of M6-M6 in M256). Among all these F2F capacitances, we found that the M6_0-M6_1 capacitance is the most significant contributor to the total F2F capacitance. We also found that die-by-die extraction 1) significantly overestimates the M6-M6 capacitance (in the same die) and 2) cannot accurately extract the F2F capacitance. Because of this, a significant timing/noise error occurs (25.48/175%) in the nets. With respect to the reduction in the F2F capacitance, we found that PDN can reduce it significantly (–58.3% in M256).

REFERENCES

[1] M. Jung, T. Song, Y. Wan, Y. Peng, and S. K. Lim, “On enhancing power benefits in 3D ICs: block folding and bonding styles perspective,” in *Proceedings of the 51st Annual Design Automation Conference (DAC)*, San Francisco, CA, pp. 1-6, 2014.

[2] M. Murugesan, H. Kino, A. Hashiguchi, C. Miyazaki, H. Shimamoto, H. Kobayashi, T. Fukushima, T. Tanaka, and M. Koyanagi, “High density 3D LSI technology using W/Cu hybrid TSVs,” in *Proceedings of 2011 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, pp. 1-4, 2011.

[3] M. Motoyoshi, J. Takanohashi, T. Fukushima, Y. Arai, and M. Koyanagi, “Stacked SOI pixel detector using versatile fine pitch μ-bump technology,” in *Proceedings of 2011 IEEE International 3D Systems Integration Conference (3DIC)*, Osaka, Japan, pp. 1-4, 2011.

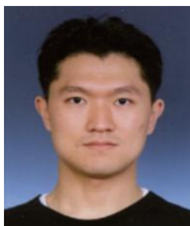
[4] C. K. Lee, C. J. Zhan, J. Lau, Y. J. Huang, H. C. Fu, J. H. Huang, Z. C. Hsiao, S. W. Chen, S. Y. Huang, C. W. Fan, et al., “Wafer bumping, assembly, and reliability assessment of μbumps with 5 μm pads on 10 μm pitch for 3D IC integration,” in *Proceedings of 2012 IEEE 62nd Electronic Components and Technology Conference (ECTC)*, San Diego, CA, pp. 636-640, 2012.

[5] OpenCores [Online]. Available: http://opencores.org/.

[6] Synopsys, 32/28 nm Generic Library [Online]. https://www.synopsys.com/COMMUNITY/UNIVERSITYPROGRAM/Pages/32-28nm-generic-library.aspx.

[7] D. H. Kim, K. Athikulwongse, and S. K. Lim, “A study of Through-Silicon-Via impact on the 3D stacked IC layout,” in *Proceedings of the 2009 International Conference on Computer-Aided Design,* San Jose, CA, pp. 674-680, 2009.

[8] T. Song, A. Nieuwoudt, Y. S. Yu, and S. K. Lim, “Coupling capacitance in face-to-face (F2F) bonded 3D ICs: trends and implications,” in *Proceedings of 2015 IEEE 65th Electronic Components and Technology Conference (ECTC),* San Diego, CA, pp. 529-536, 2015.



Taigon Song
received his B.S. in Electrical Engineering from Yonsei University, Seoul, Korea, in 2007, and his M.S. in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2009. Currently, he is working towards his Ph.D. degree at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA. His research interests include 2.5D/3D IC design and analysis.

Sung Kyu Lim
received his B.S., M.S., and Ph.D. degrees from the Computer Science Department, University of California, Los Angeles (UCLA), in 1994, 1997, and 2000, respectively. He is Dan Fielder Professor at the School of Electrical and Computer Engineering, Georgia Institute of Technology. His research focus includes the architecture, design, test, and EDA solutions for 3D ICs. His research on 3D IC reliability was featured as Research Highlight in the Communication of the ACM in 2014. Dr. Lim received the National Science Foundation Faculty Early Career Development (CAREER) Award in 2006. Further, his work was nominated for the Best Paper Award at ISPD'06, ICCAD'09, CICC'10, DAC'11, DAC'12, ISLPED'12, and DAC’14. He is also an associate editor of the IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.