Interdie Coupling Extraction and Physical Design Optimization for Face-to-Face 3-D ICs

Yarui Peng, *Member, IEEE*, Dusan Petranovic, *Member, IEEE*, Kambiz Samadi, *Member, IEEE*, Pratyush Kamal, *Member, IEEE*, Yang Du, *Member, IEEE*, and Sung Kyu Lim, *Senior Member, IEEE*

***Abstract*—Interdie coupling in face-to-face-bonded three-dimensional (3-D) ICs is becoming increasingly important for power and signal integrity. For the first time, we conduct a comprehensive study of the coupling impact in all three aspects: extraction methodology, physical design, and technology scaling. We conduct detailed sensitivity analysis of key parameters using full-chip 3-D IC designs built across multiple technologies from 28 nm down to 7 nm. First, we develop a hierarchy-aware design methodology that reduces the total wirelength by 28.1% and interdie coupling by 27.5%. Second, results show that interdie capacitance significantly affects full-chip timing and noise across multiple technology generations. Specifically, clock delay increases by 18% and skew 16%. Moreover, an additional power distribution network (PDN) layer in the 3-D design further reduces interdie coupling by 66%. Third, interdie coupling remains similar in advanced nodes with die-to-die distance scaling. Finally, our extraction methodology named context creation developed to handle design space exploration for logic-memory stacking reduces extraction error to 0.41% and timing error to 0.16%.**

***Index Terms*—Face-to-face, heterogeneous, 3D IC, parasitic extraction, optimization.**

I. Introduction

MOORE's law historically enables designs with higher density at a lower cost. However, it is extremely expensive to scale devices into sub-20 nm technologies in which multiple patterning, EUV, and new device structures must be used. As one of more-than-Moore technologies, 3D ICs enable next-generation systems with much higher device density without the need for technology scaling. To connect two dies in the vertical direction, face-to-face (F2F) bonding, which directly bonds the top metal layers of both dies with vertical connections, is more power-efficient than face-to-back (F2B) bonding, because of physical design quality improvement and the higher F2F via density [1].

Manuscript received January 6, 2017; revised May 8, 2017; accepted June 28, 2017. Date of publication August 2, 2017; date of current version July 9, 2018. The review of this paper was arranged by Associate Editor M. Borg. This work was supported by Mentor, A Siemens Business. *(Corresponding author: Yarui Peng.)*

Y. Peng is with the Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, AR 72704 USA (e-mail: yrpeng@uark.edu).

D. Petranovic is with Mentor, A Siemens Business, Fremont, CA 94538 USA (e-mail: dusan_petranovic@mentor.com).

K. Samadi, P. Kamal, and Y. Du are with Qualcomm Research, San Diego, CA 92121 USA (e-mail: ksamadi@qti.qualcomm.com; pkamal@qti.qualcomm.com; ydu@qti.qualcomm.com).

S. K. Lim is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA (e-mail: limsk@ece.gatech.edu).

Digital Object Identifier 10.1109/TNANO.2017.2735361

Traditional F2F designs use microbumps to connect two dies that are separated by more than 5 μm. In the future, technology nodes that support more functionalities integrated into the same chip, means a increasing number of F2F vias need to fit into a smaller die footprint. Direct copper bonding [2] eliminates the bonding layer and microbumps between dies, allowing for a much smaller F2F via pitch and a closer die-to-die (D2D) distance. According to the ITRS road map [3], 3D IC bonding technology advances one generation every four years. This generally results in the F2F via pitch and D2D distance shrinking, which leads to stronger noise coupling between dies, especially if bumpless F2F vias are used [4]. Studies have shown that a 3 μm D2D distance can enable a 6 μm pin pitch using direct copper bonding process [5], which potentially creates a strong E-field coupling across dies.

The F2F-bonded designs have many advantages over their F2B counterparts. The direct copper bonding is more scalable, and provides a much higher 3D via pitch [6]. Also, the contact resistance can be reduced with optimized pre-bonding passivation [7]. Even though inter-die coupling is usually undesirable, it can be used to implement via-less inter-die communication channels [8], [9]. However, such applications also require an extremely close D2D distance, which means inter-die coupling analysis is critical to ensure a robust cross-die signal channel. Inter-die coupling also has a large impact on wafer-on-wafer (WOW) technology. To provide the smallest pin pitch, WOW must thin the silicon substrate down to a few microns [10], which results in strong inter-die E-field interactions. The WOW can also enable heterogeneous 3D ICs with a low temperature bonding [11].

In addition to fabrication technology innovations, there are some previous studies [12] on CAD flows to enable heterogeneous 3D IC designs. These multi-die systems are ideal for mobile application and Internet-of-things, because of their low power consumption and minimum footprint size [13]. They can be enabled by using 2.5D [14] or 3D packaging [15]. Several studies on CAD methodologies for 3D IC physical designs [16] have been published with interface design techniques demonstrated with heterogeneous 3D IC design prototypes [17], [18]. However, none of these works considered inter-die coupling at the full-chip level.

To accelerate the time-to-market, dies in a 3D IC system are designed separately. In previous work, the inter-die coupling capacitance could only be considered during sign-off

1536-125X © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

TABLE I
TECHNOLOGY NODE COMPARISON IN KEY FEATURE SIZES

Node	Fin Pitch	Poly Pitch	M1 Pitch
Our45	–	190	190
Our28	–	116	116
Our14	40	64	64
Our7	36	54	54
Intel22	60	90	90
TSMC16	48	90	64
IBM14	42	80	64
Samsung14	40	78	64
Intel14	42	70	60

verification stage [19], where netlists and layouts of all dies are known. However, ignoring inter-die coupling, timing, and power analysis for a single die potentially increases the risk of violating timing constraints, which may require a redesign of the whole chip. To alleviate this issue, designers must leave large design margins and make worst-case assumptions, which increases total cost and power by inserting many buffers to meet timing targets. In addition, previous study ignores some critical nets, such as clock and power supply nets. These nets are routed on the top metal layer, where routing environments are cleaner and wire resistivity is low. As a result, with future device scaling and bonding technology advancing, the impact of inter-die coupling remains unknown.

II. INTER-DIE COUPLING ANALYSIS

A. F2F Bonding Technology Settings

To study technology trends, we use three technology nodes in this paper: A commercial FD-SOI 28 nm technology, an open source 14 nm FinFET technology [20], and a 7 nm FinFET technology from an industrial IP vendor. We chose these three nodes because they cover a wide range of designs, which provides a thorough examination of the advancement in technology scaling roughly every four years. With every two technology nodes, the interconnect dimension shrinks approximately by half, and cell density increases by about 3.5x. Detailed dimensions regarding interconnect technologies are listed in Table I. To ensure a realistic and representative study, we also compare results of the interconnect dimension and cell density with those of commercial foundries and IDMs to ensure our design matches state-of-the-art designs Table I.

With advanced technology nodes, the bonding technology must also scale accordingly as well to provide smaller pin dimensions and a higher landing-pad density. We assume a F2F via pitch of 2 μm in 28 nm technology as the baseline. If the same pitch is used in a 7 nm technology, one sample LDPC design will have 2,983 F2F vias and a die size of 90 μm × 90 μm. All F2F vias will occupy more than 11,932 μm^2 area on the top metal layer, meaning the total F2F via size is larger than the design footprint. Therefore, in advanced technologies, the bonding technology must a scale accordingly to match with the pin density. Since it is difficult to fabricate a F2F via with a high aspect ratio, the D2D distance also needs to decrease. According to the ITRS technology road map, 3D IC bonding technology advances one generation every four years. Therefore, we assume a 0.7 package scaling ratio every three technology nodes and scale all dimensions in F2F vias accordingly, as shown in Table II.

TABLE II
PREDICTED BONDING TECHNOLOGY SCALING TREND OF D2D DISTANCE BASED ON FOUR YEAR PER ONE GENERATION

Technology Node (nm)	45	28	20	14	10	7
Pessimistic (μm)	1	1	0.7	0.7	0.5	0.5
Optimistic (μm)	1	0.7	0.7	0.5	0.5	0.35

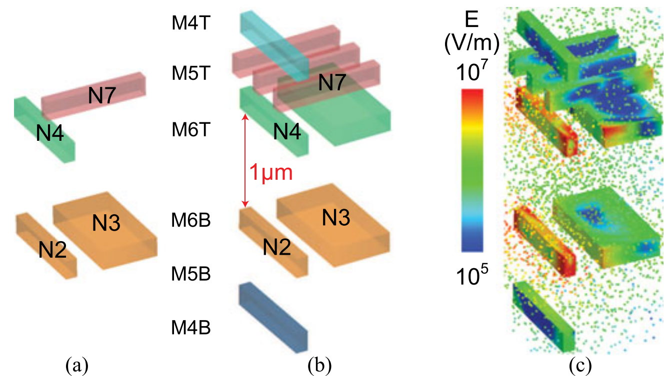


Fig. 1. (a) and (b) are two sample F2F structures in 28 nm technology; (c) shows the E-field distribution map between aggressor N4 and victim N2.

TABLE III
EXTRACTION OF STRUCTURES IN FIG. 1 WITH A UNIT OF aF

Case (a)	N2	N3	N4	N7		Total
N2		49.67	6.52	1.84		58.02
N3	49.67		12.01	5.52		67.20
N4	6.52	12.01		31.16		49.69
N7	1.84	5.52	31.16			38.52
Case (b)	N2	N3	N4	N7	Others	Total
N2		42.25	4.16	0.02	15.7	62.13
N3	42.25		5.74	0.03	32.9	80.90
N4	4.16	5.74		12.27	53.2	75.35
N7	0.02	0.03	12.27		76.2	88.50
Others	15.7	32.9	53.2	76.2	201.7	379.7

B. Field Solver Simulation

We illustrate the impact of E-field sharing with two sample wire structures shown in Fig. 1. This test structure uses the same wire dimensions as the M4 to M6 wires in our 28 nm node. There are only a few wires in structure A (Fig. 1(a)) which represent a case, in which the impact of E-field sharing is week. Structure B (Fig. 1(b)) has a much denser interconnects, resulting in a much strong coupling between wire in the same layer. E-field extraction using HFSS is shown in Fig. 1(c), and the coupling capacitance of each wire segment is shown in Table III.

For the selected wires, N2 and N3 are on the bottom die, and N4 and N7 are on the top die. These are selected to illustrate the E-field sharing impact across dies. For example, the inter-die coupling cap between N3 and N4 is 12.01 aF in structure A, but is only 5.74 aF in structure B. The total inter-die coupling capacitance is 25.86 aF for structure A and 32.49 aF for structure B.

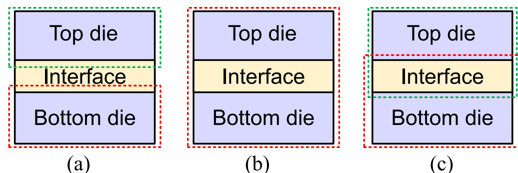


Fig. 2. Three extraction methods for F2F-bonded 3D ICs: (a) die-by-die, (b) holistic, and (c) in-context extraction.

This is because the average distance between neighboring wires on the same layer decreases in structure B. Therefore, even though the total wire length increases significantly, stronger intra-die E-field sharing reduces the total inter-die coupling capacitance. This can be seen from the large increase of intra-die coupling capacitance from 80.82 aF in structure A to 310.8 aF in structure B. In general, if the D2D distance remains the same, with denser 2D wires, the inter-die coupling capacitance percentage decreases. Therefore, all E-field interactions between dies must be extracted and analyzed with full-chip designs for accurate timing, power, and signal integrity analyses.

C. Sign-off Extraction Methodology

For extraction of F2F coupling elements, we adopt the methodologies proposed in [19] for full-chip extraction. As shown in Fig. 2(a), the simplest method for F2F extraction is the traditional die-by-die extraction. This method extracts dies separately, ignoring any coupling between them. It directly combines the extracted individual die netlists together without any consideration of inter-die E-field interaction. The die-by-die can be easily implemented using traditional rule-based extraction engines for 2D ICs, such as Calibre xACT. By extending the 2D vertical metal stack, it can also be implemented with FastCap [21], Random Walk based extraction [22], or fast boundary element methods [23] such as Calibre xACT 3D. Although implementation of die-by-die extraction is simple, it cannot handle any inter-die coupling elements. Therefore, it is only accurate when the D2D distance is large, or both dies are protected by power distribution network (PDN) shielding.

Another extraction method is holistic extraction, shown in Fig. 2(b), in which all layers from both dies are included in the extraction. It is an ultimate solution that provides the highest accuracy, since the extraction engine has all the layout information and extract all layers simultaneously. However, this strategy also results in significant CAD complexity. Moreover, to create an extraction rule deck for a certain technology, designers must have a knowledge of the manufacturing process for both dies. This requirement is extremely difficult to satisfy in reality, since foundries typically do not share their device fabrication secrets, especially to their competitors. Instead, to protect their intellectual properties (IPs), rule decks from foundries are encrypted so that trade secrets are not revealed.

In addition, even if device information is not shared, the holistic extraction requires design houses to share the connectivity information of their chips. This leaves room to reverse-engineer the design based on the provided netlist and layout information. With future 3D IC technologies, more dies will be stacked. Therefore, the holistic extraction, which reads metal layer geometries and netlists of all dies, will consume a significant amount of runtime and memory resources. For example, we ran holistic extraction on a commercial quality F2F processor in a 10 nm technology with 10 metal layers. The total runtime took more than 4 days and the total required DRAM space exceeded 700 GB. It is likely that multi-die holistic extraction will be limited by computing resources and trade secrets.

TABLE IV
EXTRACTION ON 28 NM LDPC UNDER VARIOUS NUMBER OF INTERFACE LAYERS

Interface	Type	M1-4B	M5B	M6B	M6T	M5T	M4-1T	Total
Holistic extraction results (pF)								
M1-M6	intra-die	65.6	22.2	20.6	18.4	21.5	59.9	208
	inter-die	0.06	0.24	3.75	3.72	0.25	0.08	8.11
In-context extraction error (fF)								
M6	intra-die	−242	−128	−199	−299	−231	−365	1601
	inter-die	−36.8	−229	182	195	−235	−266	922
M5-M6	intra-die	−133	−70.0	−66.1	−59.0	−120	−200	734
	inter-die	−24.6	3.36	−36.7	−30.5	3.91	−21.2	130
M4-M6	intra-die	−121	−51.7	−34.2	−25.6	−43.2	−171	491
	inter-die	−5.70	−0.42	−41.8	−39.9	1.58	−6.09	98.9

Total Error is Reported in Absolute Sum.

For better IP protection and heterogeneous integration, in-context extraction (Fig. 2(c)) is more appropriate when different foundries fabricate the top and bottom dies. Instead of requiring all layers from the neighboring die, in-context extraction takes only a few extra layers (called "interface layers") from the neighboring die during extraction. The extraction time and the memory required to store the layout significantly decreases, and both top and bottom dies can be extracted in parallel to further reduce extraction time. Moreover, foundries are not required to reveal their device fabrication details but have to share details of a few metal layers from the top metal stack. Design houses only need to share the top metal routing. This interconnect information can be easily found in the technology manuals, which allows design houses to bond chips from different vendors, but still capture most inter-die coupling elements and achieve close-to-optimum accuracy.

D. Full-Chip Inter-Die Coupling Impact

To study the impact of F2F inter-die coupling, we use the 28 nm LDPC design. We perform all three extraction methods, with results shown in Table IV. With the die-by-die extraction, all inter-die coupling capacitance is ignored, resulting in a significantly lower coupling capacitance, especially for the top metal layer. The total inter-die coupling capacitance for M6B is 3751fF and for M6T it is 3721fF. While the absolute value of intra-die coupling capacitance out of the total coupling capacitance is small with a 3.8% portion, inter-die coupling capacitance comprises 19.1% of all coupling capacitance on M6B and M6T. These results mean if the inter-die coupling capacitance is ignored, then wire caps, especially for nets on the top metal layer, are heavily underestimated.

If we focus on 3D nets only, which are routed across dies, we can observe inter-die coupling more efficiently. As shown

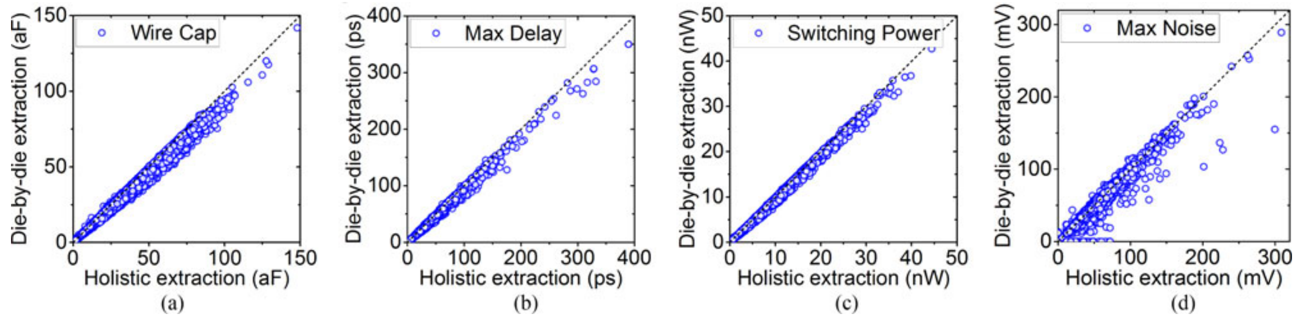


Fig. 3. Die-by-die vs. holistic extraction results on 3D nets of the (a) wire cap, (b) max delay, (c) switching power, and (d) max noise.

in Fig. 3, each dot represents one 3D net, whose wire cap, max delay, switching power, and max noise of holistic extraction are compared to those of die-by-die extraction. The maximum underestimated 3D net has a 26% smaller total wire cap with die-by-die coupling extraction, resulting in a large underestimation in both power and noise as well. However, gate capacitance is larger than regular wire coupling capacitance, especially for a highly optimized design in which long wires are segmented by buffers. Therefore, we only observe a small impact from inter-die coupling on the longest path delay of many 2D nets. On the other hand, on some 3D nets, we observe as much as 23.5% max change in delay resulting from inter-die coupling because these nets are not on the critical path. As non-critical 3D nets have fewer and weaker buffers, the wire capacitance portion is larger. Also, holistic extraction captures neighbor aggressors of 3D nets, leading to a significant increase in the worst-case noise. On some nets, the die-by-die extraction fails to extract any aggressors, and only forms ground capacitance, which results in a zero noise. With holistic extraction, noise on these nets can be accurately captured.

With in-context extraction, we capture most inter-die coupling using significantly fewer resources. As shown in Table IV, in-context extraction is almost as accurate as holistic extraction. Previous work [19] uses a 45 nm technology and concludes that two interface layers are sufficient for in-context extraction to provide comparable results to holistic extraction in terms of accuracy. However, for an advanced technology, since the thickness of wires and distance between layers are shrinking, the E-field from the neighboring die can affect more metal layers. Thus, in a 28 nm technology node, adding up to three interface layers further improves in-context extraction accuracy. As results show, adding two interface layers into each in-context extraction die reduces the coupling cap error to -0.39%, while adding three interface layers further reduces the error to less than -0.27%. As general guidance from our experiments, two interface layers are enough for technologies with a top level metal pitch larger than 0.3 μm, while three interface layers are needed to provide the best accuracy for advanced technologies.

III. Physical Design Impact

We select two benchmarks to study the impact of full-chip inter-die coupling with logic-logic stacking. We use a low-density parity-check (LDPC) design that is a widely-used encryption engine, and an OpenSPARC T2 processor core. The LDPC design is a pin-dominated design with 4105 IO pins, while the T2 core is a cell-dominated design with 401k gates. These benchmarks enable us to cover a wide range of applications with realistic layouts. Current designs are much more complicated, so they require careful PDN and clock tree analyses for high performance and design yields, especially for advanced technologies, where mask expenses are so high that ensuring a high probability of first-time success is crucial. Since both PDNs and clock nets that are global nets and they are usually routed on top metal layers, these nets are more likely to be affected by inter-die coupling as well as other coupling elements in the F2F stack.

A. Design Hierarchy Choice

Since no standard design flow exists for 3D ICs, designers may choose various CAD tools and flows for design partition, floorplanning, and placement, which leads to significant variation in final design metrics. Depending on design implementation, inter-die coupling also varies significantly, especially for large-scale designs with detailed architectural hierarchies. We use the T2 core to study the impact of the design floorplan on wirelength and inter-die coupling. The traditional gate-level design flow flattens the whole design and uses min-cut as the partition scheme. However, since no optimum partitioner exists, the heuristic partitioner, unaware of the design hierarchy, may separate standard cells that belong to the same block into several dies. Such partitioning results in more 3D vias as well as longer overall wirelength.

As the T2 core consists of several blocks, a careful partition and floorplan should take hierarchical information into consideration. As shown in Fig. 4(a), while the gate-level design uses a partitioner to obtain a heuristic min-cut solution based on the flattened netlist, the block-level design uses a manual partition based on the block hierarchy. The wirelength and coupling capacitance are compared in Table V. As results show, block-level design significantly reduces the total wirelength by 28.1%, which leads to a significant reduction of 27.5% in all coupling capacitance, especially for inter-die coupling capacitance on top metal layers.

Note that, unlike the block level flow used in [1], our flow is still based on the flattened netlist, which allows design tools to further optimize across block boundaries. Traditional block-level flow performs separate optimizations within each

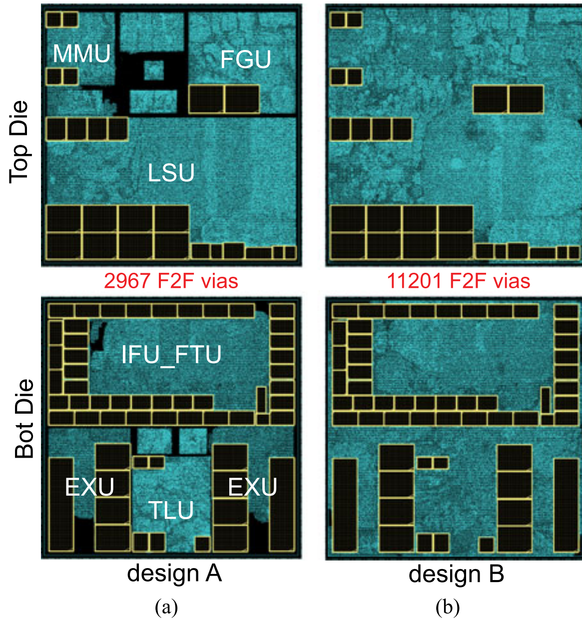


Fig. 4. T2 core design flavors. (a) block-level design, (b) gate-level min-cut design.

TABLE V
INTER-DIE COUPLING COMPARISON OF THE TWO T2 DESIGNS SHOWN IN FIG. 4

Block-level	M5B	M6B	M7	M6T	M5T	Other	Total
WL	1429	1260	0	1434	1860	8411	14394
Intra-die	40.36	51.51	0.12	58.16	55.99	283.1	489.3
Inter-die	0.77	2.93	0.14	2.94	0.78	0.65	8.19
Gate-level	M5B	M6B	M7	M6T	M5T	Other	Total
WL	2742	2166	0	1806	2490	10806	20009
Intra-die	90.51	87.3	0.52	65.4	76.86	354.8	675.4
Inter-die	1.18	4.59	0.53	4.52	1.16	0.27	12.31

Capacitance and Wirelength (WL) Values are in pF and mm, Respectively.

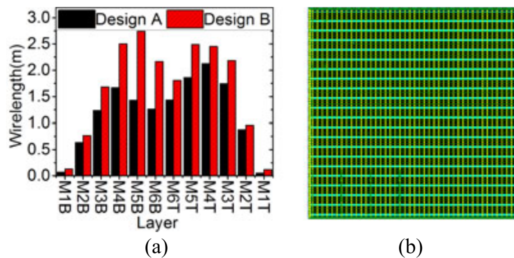


Fig. 5. (a) Wirelength distribution of two T2 designs styles from Fig. 4. (b) PDN designs of the 28 nm T2 core.

block and on the top level. Using our flattened design with block hierarchy awareness, tools can take advantage of all cell information and perform optimization on the entire design. The wirelength distribution is shown in Fig. 5(a), where the hierarchy-aware design has a much shorter top metal layer wirelength. These results demonstrate that, for the best design quality and inter-die coupling reduction, a hierarchy-aware design partition and floorplan are needed.

TABLE VI
IMPACT OF PARTITIONING (LDPC DESIGN). Δ IS WITH RESPECT TO MIN-CUT PARTITIONING

Partition	Wirelength (mm)		F2F Via		M6-to-M6 Cap (fF)	
	Both M6	Δ	Via#	Δ	Cap	Δ
Min-cut	392	–	3,866	–	792	–
Mid-cut	523	33.5%	6,878	77.9%	1,162	46.6%
Max-cut	451	15.1%	19,798	412%	1,038	31.0%

Δ is with respect to min-cut partitioning.

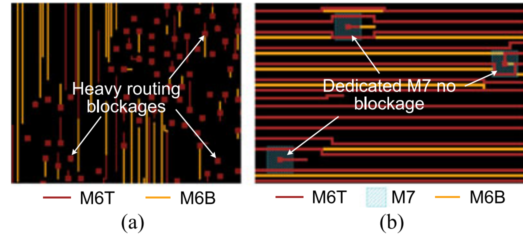


Fig. 6. F2F via options. (a) M6 wires are heavily blocked by F2F via pads, (b) M6 routing is not blocked because of the dedicated M7 for F2F via pads.

B. Routing Blockages by F2F Vias

Another physical design impact comes from routing blockages caused by F2F vias. To analyze how much inter-die coupling capacitance is contributed by F2F vias, we build a T2 design that only routes up to M6 and uses M7 purely for F2F via landing pads. Removing top layer routing significantly reduces the inter-die coupling from 18.9 pF to 8.19 pF. The holistic extraction results are shown in Table VI. Most of the coupling capacitance comes from M6, while only a small percentage comes from M7. Therefore, we conclude that the F2F vias do not contribute much to the total inter-die coupling capacitance by themselves.

However, with more F2F vias, connecting these vias requires more routing on the top metal layer. As a result, longer wirelength is routed on the top metal layer, which leads to larger inter-die coupling capacitance. With more routing on the top metal layer and larger caps, inter-die coupling increases with more F2F vias, which are also routing blockages. If too many F2F vias are introduced into the top metal layer, their landing pads heavily block routing tracks. As an example, we build a similar design with a max-cut partition in which we maximize the use of the F2F via. As shown in Fig. 6, because of heavy routing blockage on the top metal layer, the top metal layer wirelength significantly decreases.

To illustrate the impact of F2F vias, we build three variants of the LPDC designs. All three designs are created with the same flow, but different partition schemes: min-cut, max-cut, and mid-cut. Table VI lists holistic extraction results. Both min-cut and max-cut have a shorter top routing wirelength than the mid-cut. Compared with the min-cut partition design, max-cut design increases inter-die coupling by 31.0% due to the 15.1% longer M6 wires. However, for the mid-cut option with 6878 F2F vias, inter-die coupling is the strongest because its M6 wires are 33.5% longer. Therefore, inter-die coupling capacitance is maximized due to long wires on the top metal layers. From

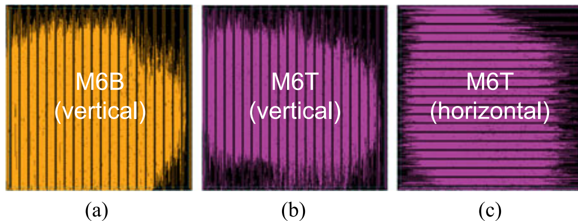


Fig. 7. Two routing directions for M6 (LDPC benchmark). Design XX uses (a) and (b), while design XY uses (a) and (c).

TABLE VII
HOLISTIC EXTRACTION OF LDPC DESIGNS XX AND XY

Design	Layer	M5B	M6B	M6T	M5T	Others	Total
XX	intra-die	22.2	20.6	18.4	21.5	125.5	208.3
	inter-die	0.24	3.75	3.72	0.25	0.15	8.11
XY	intra-die	22.2	20.5	19.6	28.7	123.6	214.6
	inter-die	0.69	4.03	3.78	0.96	0.17	9.63

Values are in pF

these results, the impact of F2F vias on inter-die coupling is indirectly dependent on the F2F via count, mostly because of the correlated wires on the top metal layer that create strong coupling between dies in an F2F 3D IC. As a design guideline, fewer top metal wires and dedicated F2F via layers would be effective to reduce inter-die coupling.

C. Impact of Routing Direction

Another impact comes from routing directions of interface layers. We design two LDPCs with different metal stacks, shown in Fig. 7. One design, XX, keeps the original routing direction for both its bottom and top dies. The other design, XY, has the same bottom die as XX, but the routing directions of all layers on the top die, except for M1, are rotated by 90 degrees, resulting in orthogonal top layer routing directions between both dies. With holistic extraction shown in Table VII, we observe that the XX design has a much larger M6-to-M6 coupling capacitance than the XY design, since its M6 layers have the same routing direction. However, for total inter-die coupling capacitance, design XY is larger because M5 of one die and M6 of the other die are routed in the same direction, so both coupling of M5B-to-M6T and M5T-to-M6B increase, resulting in 2.4x greater inter-die coupling on M5 layers in design XY. Although these coupling components are secondary compared to the coupling between M6 layers on the full-chip level, they still consume a large portion of total inter-die coupling. The routing direction causes these secondary coupling components to increase, leading to larger total inter-die coupling capacitance. As a result, using two interface layers in in-context extraction is recommended, particularly with orthogonal routing direction on top metal layers with stronger non-neighbor coupling.

D. Coupling Impact on Power Net

Unlike other signal nets, power and ground nets are mostly routed on top metal layers to minimize wire resistance. To analyze the inter-die coupling on PDNs, we generate T2 designs with PDNs routed from M4 to M6. The PDN routing is shown in Fig. 5(b). We use 10%, 15%, and 20% of the total area for PDN routing from M4 to M6, respectively, while M1 to M3 are used only for signal nets. Results in Table VIII show that PDN coupling capacitance consumes a large portion of total inter-die coupling, since PDNs are mostly routed in top metal layers and occupy significant area. Thus, a thorough understanding of dynamic power integrity necessitates a careful analysis of inter-die coupling.

However, since PDNs are treated as DC signals and are considered stable most of the time, they act as ground in an AC domain. Also, instead of extracting a coupling capacitor, most extraction tools generate ground capacitance instead, meaning no coupling noise is observed. In addition, as discussed in Section II-B, PDNs can share E-fields between wires, so they reduce the coupling E-field between other signals. Therefore, using more PDN wires on the top metal layer to shield coupling E-field can effectively reduce any coupling noise from the neighboring die.

E. PDN Shielding Impact

Although a PDN significantly affects parasitic extraction, it also provides E-field shielding for other signal nets. In addition, longer PDN wires reduce top metal layer wirelength since the PDN also occupies additional space and reduces available routing tracks for signal wires. Therefore, it provides a perfect means of reducing inter-die coupling, so that aggressive noises from the neighboring die can be minimized. On the other hand, additional PDN wires increase the overall cost, since these wires are mostly routing blockages. Such wires may limit the placement of F2F vias to non-optimum regions, resulting in longer total wirelength. In this section, we provide detailed analysis of using PDNs as protection wires for inter-die E-field shielding.

To demonstrate this effect, we insert an additional M7 on top of the 28 nm T2 design, while keeping the same F2F via location. Extraction results are shown in Table IX. Because of additional D2D spacing, total inter-die signal coupling significantly reduces by 56.5%. We then insert additional PDN wires on the empty space of M7. The PDN occupies 20% of the total M7 area, and the rest of the space is used for F2F via connections. As results show, total inter-die coupling on signal wires further decreases by 20.9% with additional PDN routing. Note that inter-die coupling from the PDNs themselves increases with additional M7 PDN wires. However, it is generally beneficial to have larger capacitance on PDN, as these parasitics can act as local decoupling capacitors to reduce dynamic voltage droop. For example, compared with the original design with M6, the total inter-die capacitance on PDN wires increases from 1.46 pF to 3.6 pF. From these results, we conclude that adding an extra PDN layer can significantly reduce inter-die coupling on signal wires.

F. Coupling Impact on Clock Net

Similar to power nets, the clock network is also heavily routed on top metal layers. These wires are closer to other aggressors from the neighboring die, especially for the trunk of the clock

TABLE VIII
COUPLING CAPACITANCE BREAKDOWN FOR SIGNAL, CLOCK, AND POWER NETS IN T2 (HOLISTIC EXTRACTION USED)

Net	Layer	M1B	M2B	M3B	M4B	M5B	M6B	M6T	M5T	M4T	M3T	M2T	M1T	Total	%
Signal	intra-die	1154	14250	35885	52742	49547	45186	61491	76271	71715	58139	21314	1473	489166	96.6%
	inter-die	0.17	3.29	29.9	276.9	1050	6473	6611	1049	157.4	13.8	1.16	0.23	15667	3.37%
Clock	intra-die	9.9	981	1606	1788	3668	3810	4791	5736	2499	2679	1711	10.2	29288	94.4%
	inter-die	0.01	2.61	2.22	28.5	110.8	727.7	748.3	108.9	9.29	1.22	4.16	0.00	1744	5.62%
Power	intra-die	90.4	17042	2921	14818	2448	5972	5990	3425	18021	4615	27342	103.9	102788	98.6%
	inter-die	0.01	54.3	0.12	191.5	132.8	351.6	457.5	125.9	90.3	0.11	56.6	0.00	1461	1.40%

TABLE IX
IMPACT OF PDN SHIELDING ON SIGNAL NET INTER-DIE COUPLING

Top layer	PDN	M5B	M6B	M6T	M5T	Other	Total
M6	M4-M6	1.29	7.55	7.82	1.28	0.92	18.87
	M4-M6	0.77	2.93	2.94	0.78	0.79	8.2
M7	M4-M7	0.55	2.46	2.52	0.56	0.39	6.5

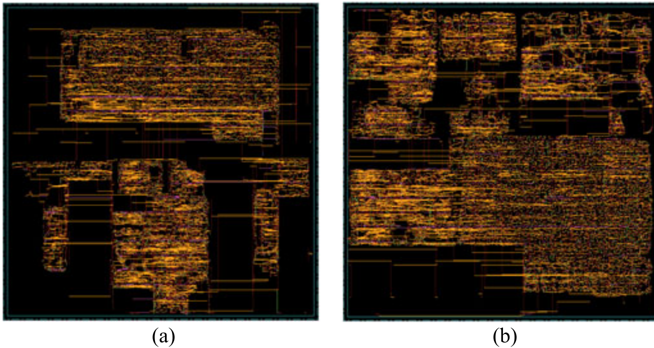


Fig. 8. Clock tree of T2. (a) bottom die, (b) top die.

tree. Fig. 8 shows a clock tree network of a 28 nm T2 design. As most of these clock routes are above M4, they are more sensitive to inter-die coupling. If die-by-die extraction is used, the clock delay, skew, and transition time are underestimated. Since any delay changes on the clock net affect all connected timing paths, it is critical to analyze the impact of inter-die coupling on clock networks and obtain the most accurate clock propagation delay.

To illustrate the impact on clock nets, we use a 28 nm T2 design, which has many memory macros with a significant number of flip-flops and requires many clock wires. We route clock trees in both block- and gate-level designs (Fig. 4) with a target clock period of 1.5 ns. Because no commercial CAD tool currently provides a 3D clock tree synthesis, we use only one clock TSV for the clock tree, and perform 2D clock tree synthesis using Encounter. This results in a slightly larger clock skew across dies. The full-chip timing and power analysis results are shown in Table X. As these results indicate, if die-by-die extraction is used on the clock tree, the max delay and clock transition time are significantly underestimated. Note that the impact of inter-die coupling capacitance on signal net timing is relatively small because of large pin capacitance. However, these small delay increases can accumulate on a clock tree with more than five levels of clock buffers and clock gates. Also, the signal skew increases up to 16.4%, due to the clock net delay changes. Because clock nets are much more sensitive to inter-die coupling, clock tree synthesis for 3D IC needs a detailed inter-die coupling-aware parasitic extraction.

TABLE X
IMPACT OF DIE-BY-DIE (DBD) VS. HOLISTIC EXTRACTION ON VARIOUS FULL-CHIP METRICS FOR T2 DESIGNS SHOWN IN FIG. 4

	Block-level partition			Gate-level partition		
	DBD	Holi	Δ%	DBD	Holi	Δ%
Clock delay (ns)	1.02	1.16	**13.7%**	1.08	1.21	**12.0%**
Clock transition (ns)	0.83	0.96	**15.7%**	1.06	1.25	**17.9%**
Clock skew (ns)	0.54	0.59	**9.3%**	0.55	0.64	**16.4%**
Switching power (W)	0.17	0.17	0.4%	0.17	0.17	0.2%
Total power (W)	0.33	0.33	0.2%	0.34	0.34	0.1%
Worst-case noise (V)	0.48	0.47	−2.1%	0.51	0.53	3.9%
WNS (ns)	−0.07	−0.05		−0.06	−0.10	

Another trend, shown in Table VIII, is that the clock network has the smallest total coupling capacitance, compared to signal nets and power nets. However, its inter-die coupling capacitance portion is the largest. Both power and clock networks are routed in the top level. With the same PDN for both dies, all power wires on the top metal layer overlap with wires of the same net resulting in the smallest inter-die coupling. However, unlike power wires, clock routes in both dies can be significantly different. Therefore, clock routes are likely to interact with other nets, leading to stronger inter-die coupling.

IV. LOGIC-MEMORY EXTRACTION

A. *Context Creation Methodology*

Though both holistic and in-context extraction accurately handle F2F designs during the sign-off verification stage, they require an LVS-clean design to generate interface layers with their electrical connections annotated on layout structures. Without a clean netlist or connectivity information, wires can only be treated as floating or ground, which lowers the extraction accuracy. However, with heterogeneous 3D ICs, the bottom die and top die of the same chip may come from different vendors, and may be fabricated by different foundries. To reduce time-to-market, all dies in a 3D IC may be designed in parallel, making it is difficult to exchange detailed interface layouts before the sign-off stage.

During the initial design stage, for procedures such as floorplanning, placement, and routing, designers may not have LVS-clean interface layers from the neighboring die for

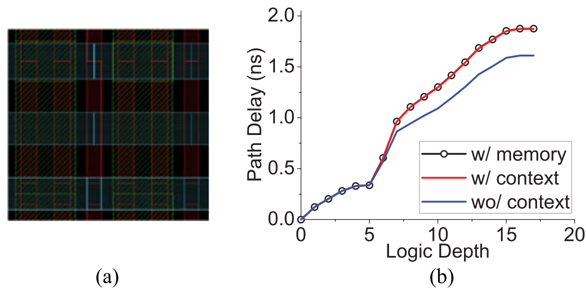


Fig. 9. (a) M2-M4 routing of a memory block. (b) Longest path delay calculation comparison.

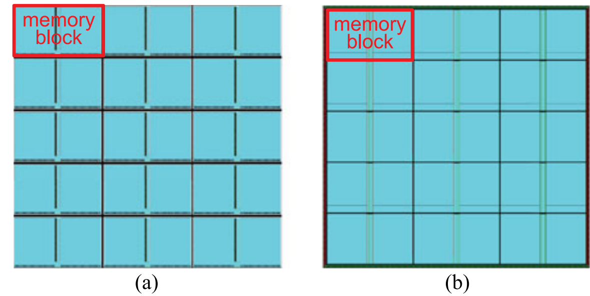


Fig. 10. Memory die layout comparison. (a) Memory die in GDS format. (b) Auto-generated context die in Cadence Encounter.

extraction. But if dies are designed unaware of each other, inaccurate parasitics lead to miscalculated timing, power, and noise. This increases the risks of having to redesign the whole chip after the two dies are bonded. Traditionally, designers of individual dies have to leave large design margins and assume worst-case. This approach requires inserting lots of buffers for the IO interface, which increases area overhead and power consumption. Even if all F2F via nets are buffered, inter-die coupling still affects single die performance, since 2D nets routed on the top metal layer are also affected by the neighboring die. Therefore, E-field sharing from the neighboring die must be considered even during early stage designs.

As discussed in Section II-C, accurate extraction can be achieved by creating an extraction context for a single die. To handle early stage designs, we propose an effective way of creating the extraction context by taking advantage of the regularity of the top layer metal geometry. If the top layers of the neighboring die follow certain layout patterns, only a small amount of information is needed to rebuild the extraction environment. This is a very common situation, since logic chips usually have their top layers covered by PDNs in a regular fashion, while memory chips usually have regular layouts for both signal and power nets.

B. Extraction of Logic-Memory Design

We demonstrate our Context Creation method with a heterogeneous logic-cache-partitioned 3D IC design routed up to M4, where the bottom die is a 45 nm signal processor unit and the top die is a 28 nm L2 cache die. As shown in Fig. 9(a), the memory die has highly regular layouts in layers from the M2 to M4 layers, while top two layers are used mostly for PDNs. Therefore, we only need the memory floorplan, metal pitch and spacing of each layer to rebuild the extraction context. These parameters can be determined even before the memory die design stage.

To demonstrate this, we build a floorplan generator which takes this information and automatically rebuilds the memory floorplan with all blocks by using power and ground wires. Since the M1 of the memory die contains mostly non-Manhattan routing, the floorplan does not contain M1 layer geometries. However, this does not degrade in-context extraction accuracy since the impact from M1 to the bottom die is small. This auto-generated layout is formatted in LEF/DEF so it can be further processed by standard layout tools such as Encounter. As shown in Fig. 10, the auto generated layout accurately mimics the original design, which is in GDS format.

TABLE XI
PARASITIC EXTRACTION COMPARISON OF THE 45 NM LOGIC + 28 NM MEMORY DESIGN

Layer	M1B	M2B	M3B	M4B	Total	Err%
Logic die + memory GDS						
GCap	18.2	126.6	221.5	122.8	489.1	–
CCap	1.23	28.6	71.4	92.7	193.9	–
Logic die only						
GCap	18.2	126.6	218.5	113.7	477.1	−2.46%
CCap	1.24	28.9	72.7	95.9	198.8	2.51%
Logic die + context die						
GCap	18.2	126.7	220.9	125.2	491.0	0.39%
CCap	1.23	28.6	71.3	92.0	193.1	−0.41%

Units are in pF

Using the auto-generated context die with M2 to M4 routing, we apply the in-context extraction on the logic die, assuming the top die metals are floating. We also use holistic extraction as reference, where the top die GDS is created with a memory compiler. We compare all three methods for accuracy: Single-die extraction, in-context extraction, and holistic extraction. The results are shown in Table XI. Without the context, extraction of the logic die is inaccurate. The ground capacitance is underestimated by 2.46%, since inter-die coupling between M4 and the memory PDN is ignored. The coupling capacitance is overestimated by 2.51%, since the E-field sharing of the top die is ignored. With our Context Creation method, the extraction errors are significantly reduced, to less than 0.39% and 0.41% for ground and coupling capacitance, respectively. The Context Creation method is highly accurate by taking advantage of regular top layer routing. Designers can use the Content Creation method to perform accurate static timing analysis, even in early stages, and improve physical design quality.

Since inter-die coupling mostly affects wires on top metal layers, only parts of signal nets are affected. However, as we observed in Section III-F, the delay calculation error propagates along the path. Even if only one node has incorrect load capacitance, timing calculation becomes inaccurate for all following nodes. This is because the delay and power calculations depend not only on the load capacitance of a node itself, but also on the input transition time and signal arrival time. If one node has underestimated capacitance load, both the delay and output

TABLE XII
FULL-CHIP TIMING AND POWER COMPARISON.

Design	w/ GDS	wo/ context	Err%	w/ context	Err%
LPD (ns)	1.875	1.611	−14.1%	1.872	−0.16%
Net power	135.6	128.8	−5.01%	137.8	1.62%
Cell power	798.0	797.2	−0.10%	798.4	0.05%
Leakage	6.85	6.85	0%	6.85	0%
Total power	940.5	932.9	−0.81%	943.0	0.27%

Power Units are in mW

transition time are reduced, and the error propagates through the timing path. This results in a faster input transition time at the next logic level, and delays of all following fan-out nets are further underestimated even if their load capacitance is correct. As a result, even though only a part of nets have routing wires on the top layer, the delay miscalculation propagates through the whole chip, and amplifies along the timing path.

We perform Primetime timing and power analysis, and compare the critical path delay in Fig. 9(b). Without the extraction context, the longest path delay is underestimated by 14.1%, and results clearly show delay error propagation after a logic depth of 5, even though not all nets have incorrect load capacitance. With the auto-generated neighboring die, timing error is reduced significantly to only 0.13%. In terms of power, inter-die coupling shows much smaller impacts. As power is generally dominated by the pin capacitance and the cell internal power, inter-die coupling impacts are relatively small, but still noticeable. As results show in Table XII, using the auto-created context die, the error in net switching power is reduced from 6.76% to 1.35%.

V. TECHNOLOGY SCALING IMPACT

A. Logic-Logic Design

In this section, we discuss the impact of future technology scaling on inter-die coupling and full-chip metrics. We design LDPC and T2 cores in all three nodes: 28 nm, 14 nm, and 7 nm to provide a comprehensive analysis. All designs are routed up to M6 without dedicated F2F via layers. As we do not have memory compilers for FinFET nodes, memory macros are scaled from 28 nm accordingly. A comparison of the T2 core layout is shown in Fig. 11.

If dies are fabricated with the same technology, one impact we observe from the previous discussion is that the average distance between intra-die wires decreases while the average inter-die wire distance remains about the same. This significantly reduces the inter-die coupling cap portion in the advanced technology node. For example, a comparison of LDPC in 14 nm vs 7 nm is shown in Table XIII. With much smaller wire dimensions, the inter-die coupling capacitance decreases in 7 nm with a D2D distance of 0.5 μm, resulting in a smaller impact when using die-by-die extraction vs. holistic extraction. Also, a general trend with the advanced technology nodes is that more metal layers are needed to complete routing. Therefore, the intra-die portion is likely to further increase because more coupling capacitors are formed within each die.

Another impact of advanced technology comes from bonding scaling. Without D2D distance and F2F via dimension scaling, it will be difficult to design a complicated 3D chip with most

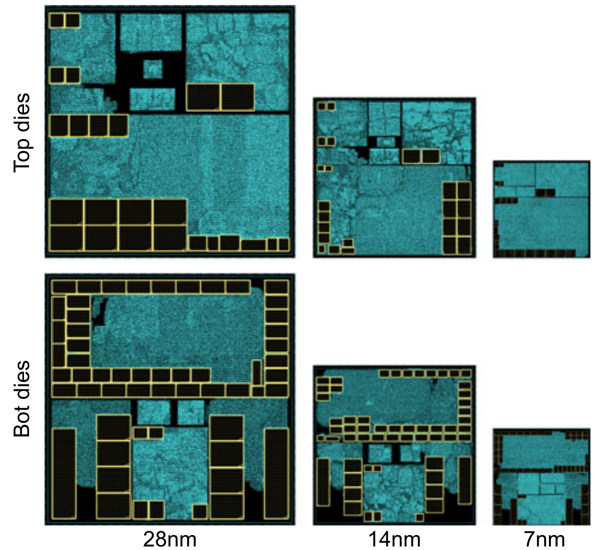


Fig. 11. Block-level T2 layouts. The dimension of 28 nm, 14 nm, and 7 nm designs are 880 μm, 560 μm, and 340 μm square, respectively.

TABLE XIII
TECHNOLOGY TRENDS OF INTER-DIE COUPLING WITH VALUES IN pF

Node	Die gap	Layer	M4B	M5B	M5T	M4T	All	%
28 nm	1.0 μm	intra-die	22.2	20.6	18.42	21.49	208.3	96.3%
		inter-die	0.24	3.75	3.72	0.25	8.11	3.74%
	0.7 μm	intra-die	22.2	20.2	18.03	21.42	207.3	95.0%
		inter-die	0.28	5.13	5.10	0.30	10.97	5.03%
14 nm	0.7 μm	intra-die	11.9	12.6	2.01	8.13	59.5	97.7%
		inter-die	0.07	0.65	0.61	0.07	1.42	2.34%
	0.5 μm	intra-die	11.9	12.5	8.97	8.10	59.3	96.8%
		inter-die	0.08	0.91	0.87	0.09	1.99	3.25%
7 nm	0.5 μm	intra-die	5.09	4.31	3.69	4.18	37.6	97.4%
		inter-die	0.05	0.45	0.45	0.05	1.00	2.58%
	0.35 μm	intra-die	5.06	4.20	3.66	4.17	37.4	96.3%
		inter-die	0.06	0.66	0.66	0.06	1.45	3.73%

The Specifications are Shown in Table I

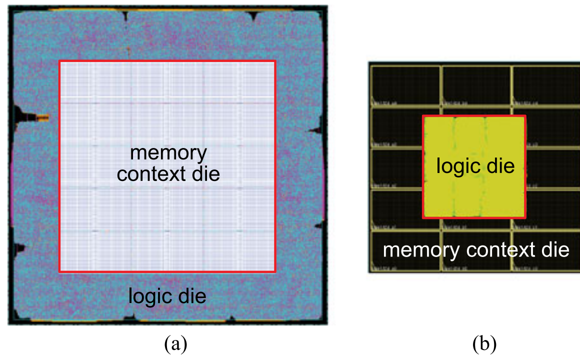
of the top metal layer fully occupied by the F2F pads. Due to technology node scaling and D2D distance shrinking, the inter-die coupling capacitance increases. For example, when we compare the LDPC in 7 nm, we observe that inter-die coupling significantly increases by 45% with a 0.7x closer D2D distance. Also, intra-die coupling capacitance decreases slightly as a result of E-field sharing. If the D2D distance shrinks further with future technologies (e.g. monolithic 3D ICs), inter-die coupling will play a more important role since the D2D distance shrinks to less than 100 nm. A summary of T2 and LDPC designs is listed in Table XIV. As results show, if the D2D distance is kept the same, the inter-die coupling portion declines. With both technology and bonding distance scaling, a similar portion of inter-die coupling remains. Therefore, we conclude that the impact of inter-die coupling still needs to be carefully extracted and analyzed in future technologies with a high metal density.

B. Logic-Memory Design

To verify our Context Creation method across technology scaling, we also implement the logic-memory design in

TABLE XIV
TECHNOLOGY TREND SUMMARY

	28 nm	14 nm	7 nm
Die-to-die distance (μm)	1.00	0.50	0.35
LDPC inter-die coupling (pF)	208.3	59.3	37.4
LDPC intra-die coupling (pF)	8.10	1.99	1.45
LDPC intra-die coupling %	3.74%	3.25%	3.73%
T2 inter-die coupling (pF)	621.2	256.7	191.0
T2 intra-die coupling (pF)	18.9	14.9	5.55
T2 intra-die coupling %	2.95%	5.49%	2.82%



(a) (b)

Fig. 12. Logic-memory design with a 28 nm memory die. (a) logic die in 45 nm, (b) logic die in 14 nm.

advanced nodes. In this new design, the logic die is shrunk to a 14 nm FinFET node, which results in a more than 2x performance increase. The layouts of our logic-memory designs are shown in Fig. 12. Although the wire dimension shrinks in the advanced node, compared with the logic die in 45 nm, the inter-die coupling impact increases, and results in a larger error for single die extraction without the context. This occurs because: First, D2D distances shrinks from 45 nm to 28 nm node, as the bonding distance is determined by the older node of the die pair. Second, the logic die dimension shrinks from a square of 1.4 mm to 0.5 mm. In the 45–28 nm nodes, the memory die only covers only 50% of the logic die in the center, while in 14 nm node, the memory die covers the whole logic die. This is different from previous designs in Section V-A with both die scaling. Therefore, the inter-die coupling impact area increases. As shown in Table XV, our Context Creation method is still highly effective for reducing extraction error in advanced nodes.

TABLE XV
PARASITIC EXTRACTION COMPARISON OF THE 14 NM LOGIC + 28 NM MEMORY DESIGN

Logic die + memory GDS						
Layer	M1B	M2B	M3B	M4B	Total	Err%
GCap	0.75	60.4	94.9	54.9	210.9	–
CCap	0.00	8.59	35.0	37.2	80.8	–
Logic die only						
GCap	0.75	60.6	93.0	50.1	204.4	−3.08%
CCap	0.00	8.67	35.4	39.4	83.5	3.38%
Logic die + context die						
GCap	0.75	60.5	94.3	53.6	209.1	−0.87%
CCap	0.00	8.61	35.0	37.1	80.7	−0.13%

Units are in pF

VI. CONCLUSION

In this paper, we analyze inter-die coupling impact on full-chip 3D F2F designs from the perspectives of extraction methodology, physical design, and future technology scaling. The impact of inter-die coupling significantly affects full-chip performance and noise. To reduce LVS complexity and improve IP protection, in-context extraction can be applied with high accuracy. By rebuilding the neighboring die floorplan with PDNs, our Context Creation method reduces extraction error to 0.41% and timing error to 0.16% for early stage designs. Physical design choices determine the inter-die coupling, and both the PDN and the clock network are significantly affected. Moreover, with advanced technology, the inter-die coupling portion decreases with thinner and denser wires. However, inter-die coupling still remains in a similar level and cannot be ignored.

To alleviate inter-die coupling and improve the quality of the physical design, hierarchy-aware floorplanning and partitioning reduce the total wirelength by 28.1% and inter-die coupling by 27.5%. Reducing the F2F via and top metal wirelength is critical to reducing inter-die coupling. Depending on the technology generation, using orthogonal routing on top metal layers reduces coupling of the neighbor layer at the cost of increasing coupling of the non-neighbor layer. For maximum inter-die coupling reduction, denser top metal layer PDN and a dedicated layer for F2F via pads can be used.

REFERENCES

[1] M. Jung *et al.*, "On enhancing power benefits in 3D ICs: Block folding and bonding styles perspective," in *Proc. ACM Des. Autom. Conf.*, Jun. 2014, pp. 1–6.
[2] Z. Li, Y. Li, and J. Xie, "Design and package technology development of face-to-face die stacking as a low cost alternative for 3D IC integration," in *Proc. IEEE Electron. Compon. Technol. Conf.*, May 2014, pp. 338–341.
[3] "International Technology Roadmap for Semiconductors," 2015. [Online]. Available: http://www.itrs.net/
[4] T. Ohba, "Production-worthy WOW 3D integration technology using bumpless interconnects and ultra-thinning processes," in *Proc. Symp. VLSI Technol.*, Jun. 2016, pp. 1–2.
[5] L. Peng *et al.*, "Ultrafine pitch (6 μm) of recessed and bonded Cu-Cu interconnects by three-dimensional wafer stacking," *IEEE Trans. Electron Devices*, vol. 33, no. 12, pp. 1747–1749, Dec. 2012.
[6] C. S. Tan, L. Peng, H. Y. Li, D. F. Lim, and S. Gao, "Wafer-on-wafer stacking by bumpless Cu-Cu bonding and its electrical characteristics," *IEEE Trans. Electron Devices*, vol. 32, no. 7, pp. 943–945, Jul. 2011.
[7] L. Peng *et al.*, "High density bump-less Cu-Cu bonding with enhanced quality achieved by pre-bonding temporary passivation for 3D wafer stacking," in *Proc. IEEE Int. Symp. VLSI Technol. Syst. Appl.*, Apr. 2011, pp. 1–2.
[8] B. Charlet *et al.*, "Chip-to-chip interconnections based on the wireless capacitive coupling for 3D integration," *Microelectron. Eng.*, vol. 83, nos. 11/12, pp. 2195–2199, 2006.
[9] S. L. Chua, A. Razzaq, K. H. Wee, K. H. Li, H. Yu, and C. S. Tan, "3D CMOS-MEMS stacking with TSV-less and face-to-face direct metal bonding," in *Proc. Symp. VLSI Technol.*, Jun. 2014, pp. 1–2.
[10] Y. S. Kim *et al.*, "Ultra thinning down to 4-μm using 300-mm wafer proven by 40-nm node 2 Gb DRAM for 3D multi-stack WOW applications," in *Proc. Symp. VLSI Technol.*, Jun. 2014, pp. 1–2.
[11] A. K. Panigrahi *et al.*, "High quality fine-pitch Cu-Cu wafer-on-wafer bonding with optimized Ti passivation at 160 °C," in *Proc. IEEE Electron. Compon. Technol. Conf.*, May 2016, pp. 1791–1796.

[12] D. Milojevic *et al.*, "Design issues in heterogeneous 3D/2.5D integration," in *Proc. Asia South Pacific Des. Autom. Conf.*, Jan. 2013, pp. 403–410.

[13] T. T. Wu *et al.*, "Low-cost and TSV-free monolithic 3D-IC with heterogeneous integration of logic, memory and sensor analogy circuitry for Internet of Things," in *Proc. IEEE Int. Electron Devices Meeting*, Dec. 2015, pp. 25.4.1–25.4.4.

[14] W. S. Liao *et al.*, "3D IC heterogeneous integration of GPS RF receiver, baseband, and DRAM on CoWoS with system BIST solution," Jun. 2013, pp. C18–C19.

[15] L. Yu *et al.*, "Electrical characterization of RF TSV for 3D multi-core and heterogeneous ICs," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, Nov. 2010, pp. 686–693.

[16] J. Knechtel, I. L. Markov, and J. Lienig, "Assembling 2-D blocks into 3-D chips," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 31, no. 2, pp. 228–241, Feb. 2012.

[17] D. Kim and S. Mukhopadhyay, "Partitioning methods for interface circuit of heterogeneous 3-D-ICs under process variation," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 24, no. 5, pp. 1626–1635, May 2016.

[18] C. Erdmann *et al.*, "A heterogeneous 3D-IC consisting of two 28 nm FPGA die and 32 reconfigurable high-performance data converters," *IEEE J. Solid State Circuits*, vol. 50, no. 1, pp. 258–269, Jan. 2015.

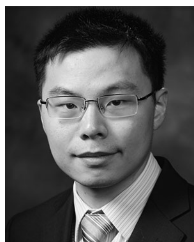
[19] Y. Peng, T. Song, D. Petranovic, and S. K. Lim, "Full-chip inter-die parasitic extraction in face-to-face-bonded 3D ICs," in *Proc. IEEE Int. Conf. Comput.-Aided Des.*, 2015, pp. 649–655.

[20] M. Martins *et al.*, "Open cell library in 15 nm FreePDK technology," in *Proc. Int. Symp. Phys. Des.*, 2015, pp. 171–178.

[21] K. Nabors and J. White, "FastCap: A multipole accelerated 3-D capacitance extraction program," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 10, no. 11, pp. 1447–1459, Nov. 1991.

[22] W. Yu *et al.*,"Utilizing macromodels in floating random walk based capacitance extraction," in *Proc. Des. Autom. Test Eur.*, 2016, pp. 1225–1230.

[23] Y. Zhou, Z. Li, and W. Shi, "Fast capacitance extraction in multilayer, conformal and embedded dielectric using hybrid boundary element method," in *Proc. ACM Des. Autom. Conf.*, Jun. 2007, pp. 835–840.



Yarui Peng (M'17) received the B.S. degree from Tsinghua University, Beijing, China, in 2012, and the M.S. and Ph.D. degrees from the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2014 and 2016, respectively.

He joined the Department of Computer Science and Computer Engineering at the University of Arkansas as an Assistant Professor in 2017. His research interests include the areas of computer-aided design, analysis, and optimization for emerging technologies and systems, such as wafer-level-packaging and 3-D ICs, as well as high-efficiency VLSI and memory systems. He is also interested in design automation of high bandgap power electronics and mobile electrified systems. He received the best-in-session award in SRC TECHCON'14 and the best student paper award in ICPT'16.

Dusan Petranovic (M'92) received the B.S. degree from the University of Belgrade, Belgrade, Serbia, the M.S. degree from the Worcester Polytechnic Institute, Worcester, MA, USA, and the Ph.D. degree from the University of Montenegro, Podgorica, Montenegro. He is an Interconnect Modeling Technologist with Design to Silicon group at Mentor Graphics working on all aspects of parasitic extraction. He was also employed as a Professor at the University of Montenegro and served as the Chairman of the Electrical Engineering department. He spent six years teaching at Harvey Mudd College before joining LSI Logic Advanced Development Laboratory as a member of technical staff. He also has worked as a consultant for NASA and NOVA Management Inc. He holds 15 U.S. patents and has published numerous journal and conference papers.

Kambiz Samadi (S'04–M'12) received the M.Sc. and Ph.D. degrees from the University of California, San Diego, CA, USA, in 2007 and 2010, respectively.

He joined Qualcomm Research, San Diego, in 2011, where he is currently a Staff Research Engineer, focusing on 3-D IC EDA solutions and 3-D IC architecture-level design space explorations. He has authored more than 25 publications in refereed journals and conferences. His current research interests include on-chip interconnection modeling and optimization for system-level design, 3-D IC modeling and optimization, and very-large-scale integration design manufacturing interface.

Pratyush Kamal received the B.Tech. degree in electrical engineering from the Indian Institute of Delhi, New Delhi, India, in 1999. He is currently a Senior Staff Engineer/Manager at Qualcomm Technologies Inc., San Diego, CA, USA. He is currently working on the development of 3-D design flows. Prior to joining Qualcomm, he held various engineering positions ranging from RTL design to CAD development at NXP, Eindhoven, The Netherlands, and Sagantec, The Netherlands and California. He holds 17 granted/filed patents related to IP design and 3-D flows.

Yang Du (M'96) received the Ph.D. degree from Columbia University, New York, NY, USA, in 1994.

He is currently the Director of Engineering with Qualcomm Research, San Diego, CA, USA, where he leads a team in advanced nanotechnology and semiconductor research. He has held various engineering positions in Analog Devices, Norwood, MA, USA, AMD, Sunnyvale, CA, USA, Motorola, Chicago, IL, USA, and Qualcomm. He has authored/co-authored more than 50 patents/patent publications and numerous conference/journal papers in very-large-scale integration technology, SPICE modeling, IC design, test, and design automation. His current research interests include emerging semiconductor devices, predictive device, circuit modeling, novel VLSI circuits and architecture, next generation 3-D IC technology and design, emerging 3-D VLSI circuit, architecture and system integration, design automation, advanced thermal modeling, and thermal aware design methodologies.

Sung Kyu Lim (SM'05) received the B.S., M.S., and Ph.D. degrees from the University of California, Los Angeles, CA, USA, in 1994, 1997, and 2000, respectively. He joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2001, where he is currently a Dan Fielder Endowed Chair Professor. His current research interests include modeling, architecture, and electronic design automation for 3-D ICs. His research on 3-D IC reliability is featured as Research Highlight in the Communication of the ACM in 2014. His 3-D IC test chip published in the IEEE International Solid-State Circuits Conference (2012) is generally considered the first multicore 3-D processor ever developed in academia. He has authored the book entitled *Practical Problems in VLSI Physical Design Automation* (Springer, 2008).

He received the National Science Foundation Faculty Early Career Development (CAREER) Award in 2006. He was on the Advisory Board of the ACM Special Interest Group on Design Automation (SIGDA) from 2003 to 2008 and awarded a Distinguished Service Award in 2008. He received the Best Paper Awards from the IEEE Asian Test Symposium (2012) and the IEEE International Interconnect Technology Conference (2014). He was an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS from 2007 to 2009. He has been an Associate Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS since 2013. He has served on the Technical Program Committee of several premier conferences in EDA. He received the Class of 1940 Course Survey Teaching Effectiveness Award from Georgia Institute of Technology (2016).