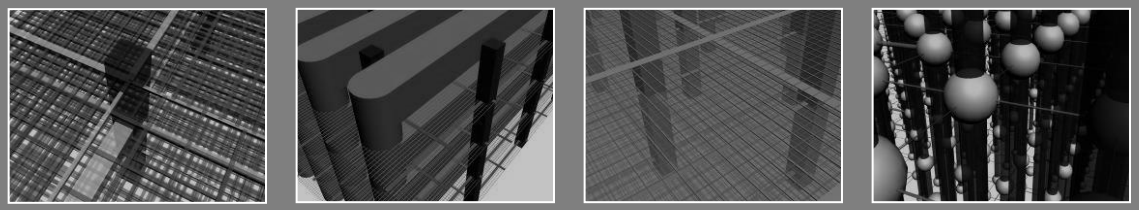
Ph.D. Thesis Defense

CAD Tools and Methodologies for Reliable 3D IC Design, Analysis and Optimization



Yarui Peng

Nov. 4, 2016

Committee:

Prof. Sung Kyu Lim (Advisor)

Prof. Saibal Mukhopadhyay

Prof. Arijit Raychowdhury

Prof. Madhavan Swaminathan

Prof. Hyesoon Kim

Overview of Thesis

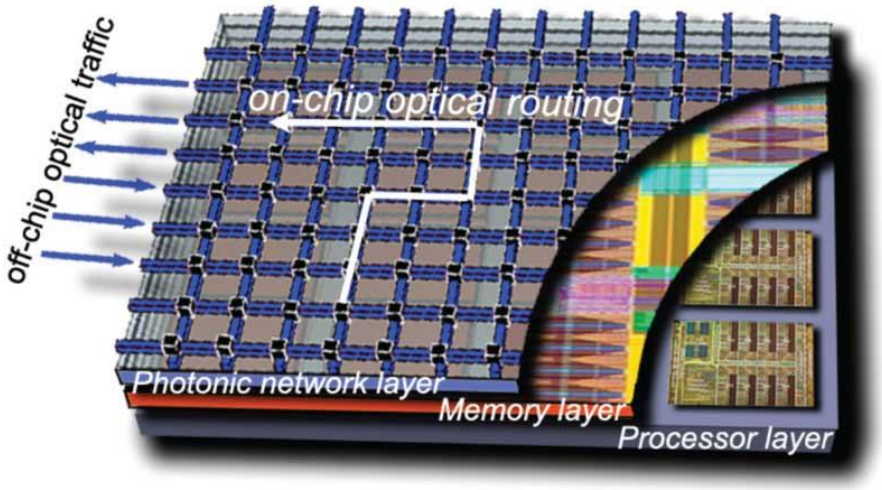
- Introduction
- Power integrity analysis and optimization for 3D DRAM [DAC14]
- Parasitic extraction and optimization in face-to-back 3D IC
  - TSV-to-TSV coupling extraction [ICCAD13, TCAD14]
  - TSV-to-wire coupling extraction [DAC14, TCAD15]

preliminary

completed

- Inter-die coupling extraction of face-to-face 3D IC [ICCAD15]
  - Holistic extraction of face-to-face boned 3D ICs
  - In-context extraction of face-to-face bonded 3D ICs
  - Inter-die coupling aware physical design optimization
- Parasitics extraction of heterogeneous 3D IC [TCPMT under review]
  - Inter-die coupling impacts in advanced technology nodes
  - Extraction for mixed technology nodes
  - Extraction for logic-memory integration

Introduction

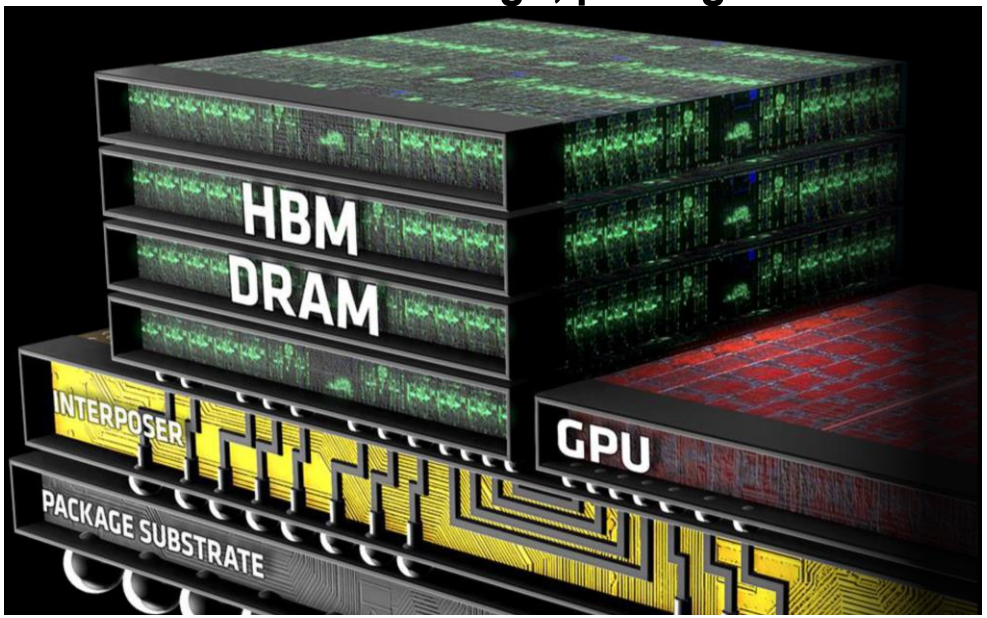


*Source: IBM

- **3D IC enables vertical integration of system components:**
  - **Face-to-back bonding with Through Silicon Via (TSV)**
  - **Face-to-face bonding with Face-to-Face Via (F2FV)**
  - **Monolithic 3D IC with Monolithic Inter-Tier Via (MIV)**
- **3D IC benefits:**
  - **Smaller footprint, higher I/O count & bandwidth, heterogeneous integration**
  - **Shorter interconnect, faster timing, lower power consumption**

Power Integrity Challenges of 3D ICs

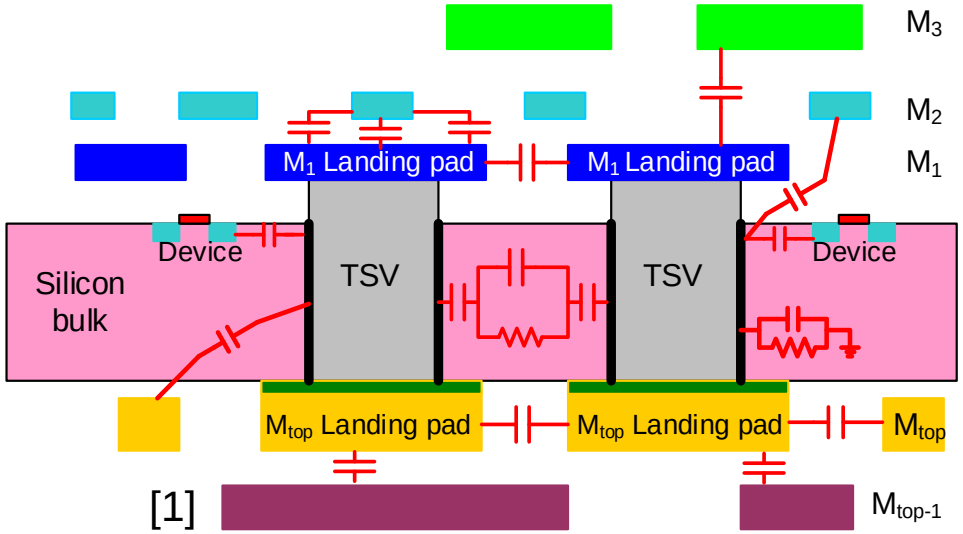
- One challenge in 3D ICs is unreliable power delivery
  - More devices need current while fewer bumps can fit into the footprint
- To solve this, we need to:
  - Assess special IR-drop issues in 3D IC system
  - Co-optimize PDNs in both memory cube and application processor (T2 chip)
  - Build the most efficient PDN design, package and architecture altogether

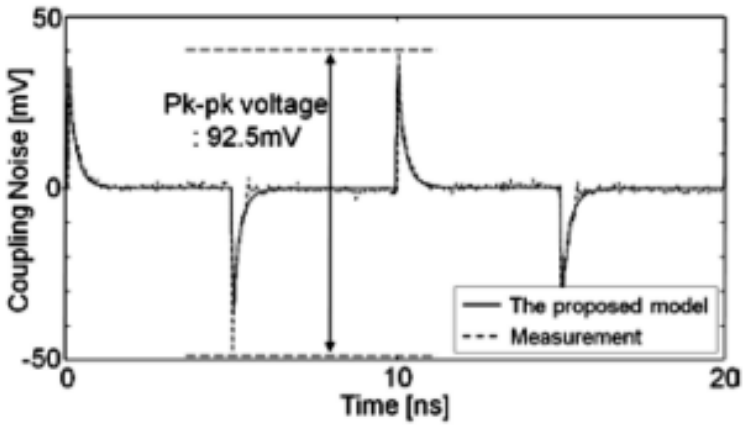


*Source: AMD

Signal Integrity Challenges (TSV-based)

- **New parasitics exist in Face-to-Back (F2B) bonded 3D IC**
- **Previous studies[1,2] showed that TSV nets have SI issues**
  - **Large delay, power consumption and noise compared with regular wires**
- **Need accurate extraction for TSV-to-TSV and TSV-to-wire coupling**
- **Need full-chip optimization methods for TSV noise reduction**



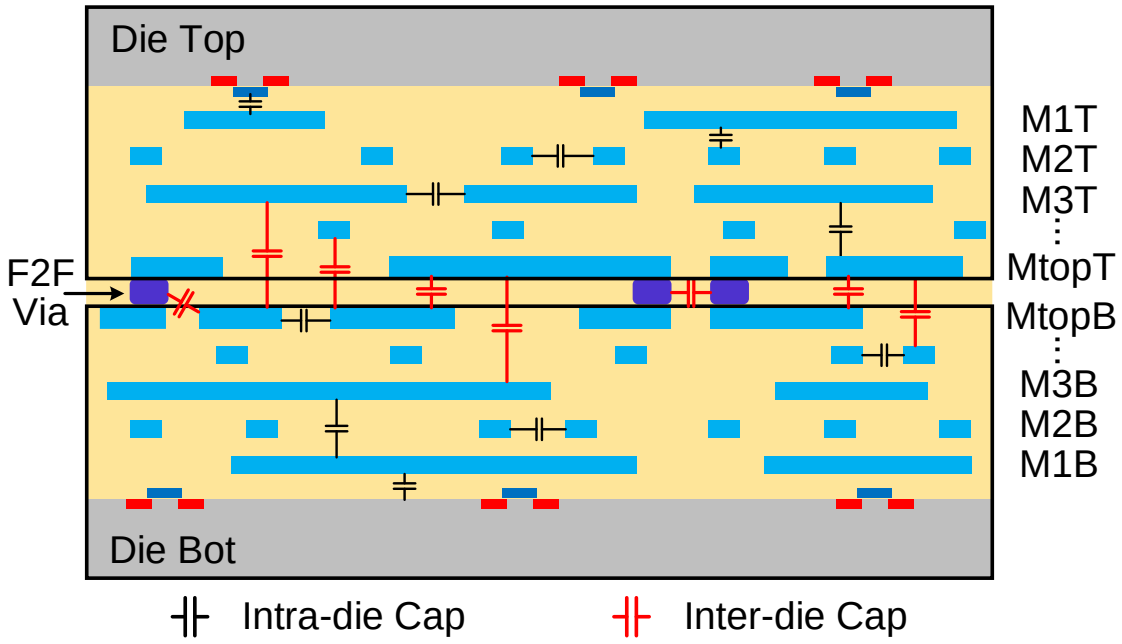


Measured TSV coupling noise [2]

[1] C. Liu et al., "Full-chip TSV-to-TSV coupling analysis and optimization in 3D IC," DAC11
[2] J. Cho et al, "Modeling and Analysis of Through-Silicon Via (TSV) Noise Coupling and Suppression Using a Guard Ring," *CPMT 2011*

Signal Integrity Challenges (F2F-based)

- **Inter-die capacitance becomes important when die-to-die distance is small, especially for face-to-face (F2F) bonded structures with direct copper bonding**



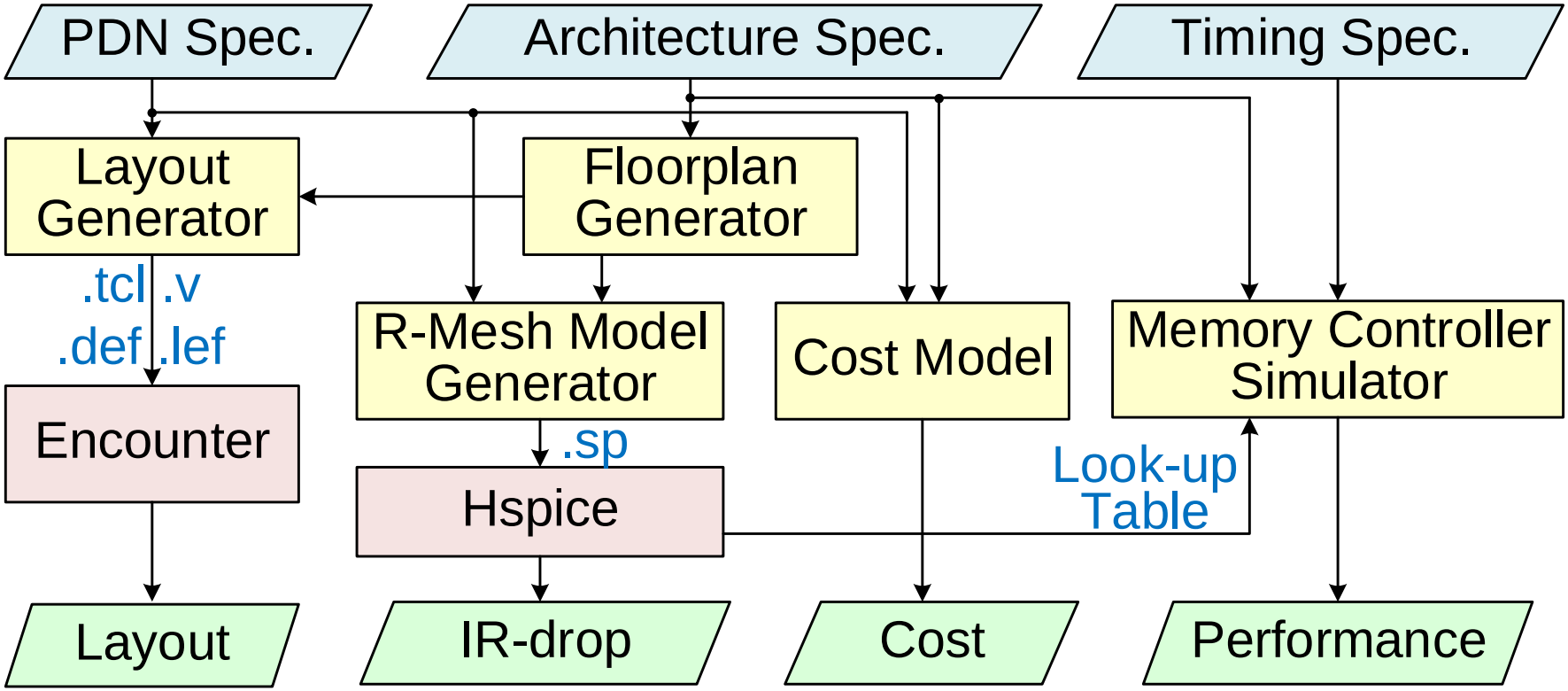
F2F-bonded 3D IC structure with interconnect parasitics

Thesis Statement

The objective of this research is to quantify power and signal integrity issues in 3D ICs, and develop CAD tools and methodologies to enable reliable 3D IC designs, as well as enhance physical design quality.

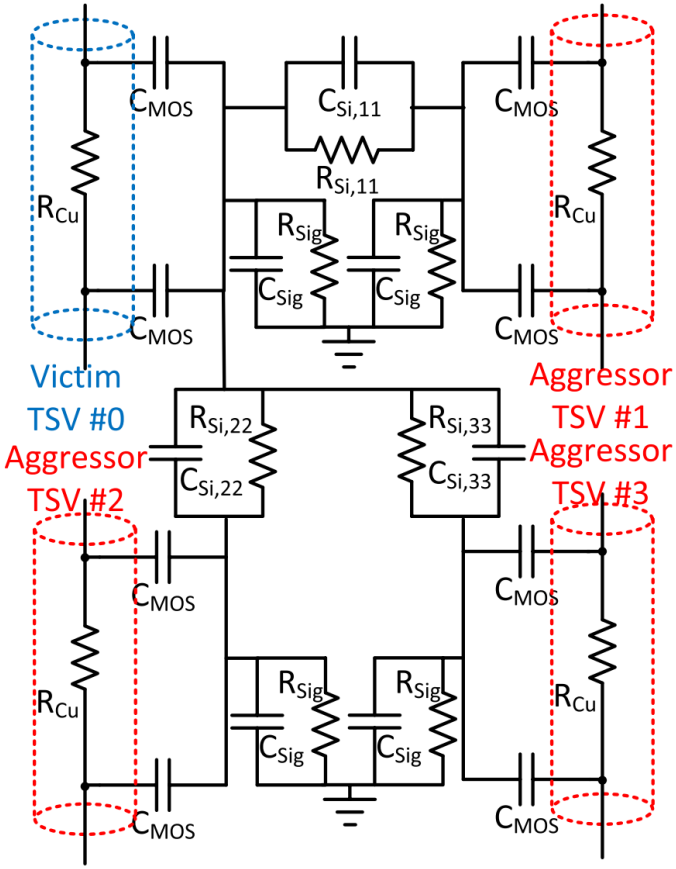
Power Integrity Analysis and Optimization

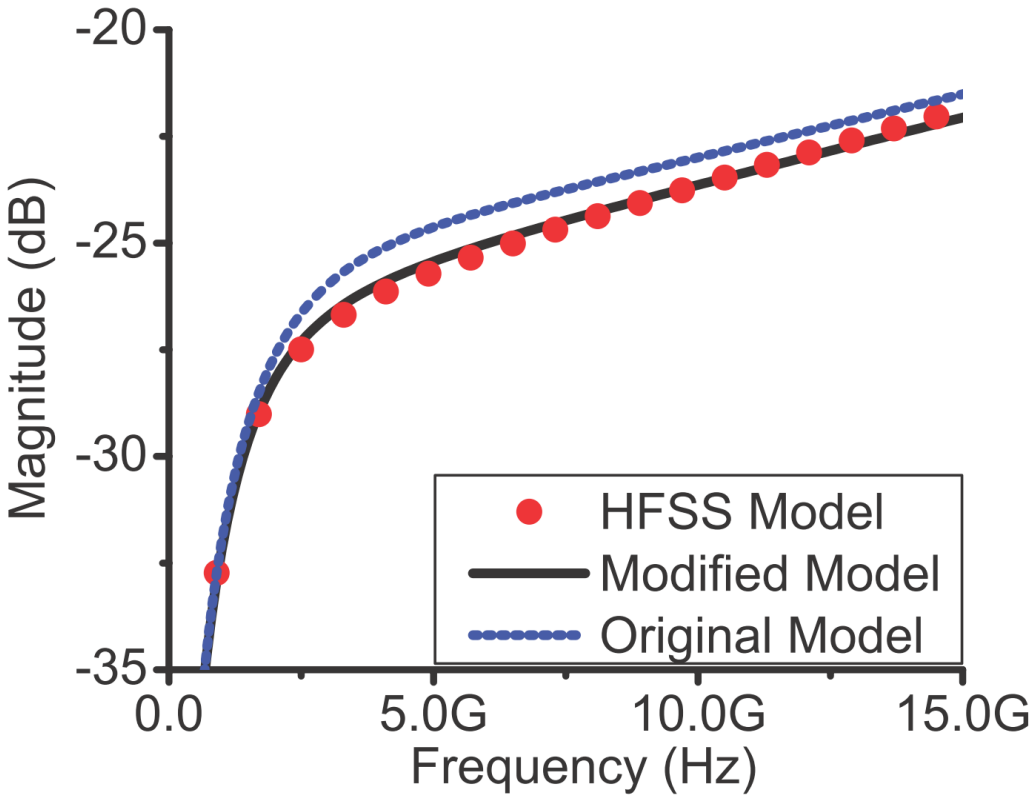
- **Our study combines a floorplanner, a PDN generator, an R-Mesh model, a memory controller simulator, and a cost model altogether**



TSV-to-TSV Coupling Extraction

- **Silicon Effect-aware Full-chip Extraction and Mitigation of TSV-to-TSV Coupling**

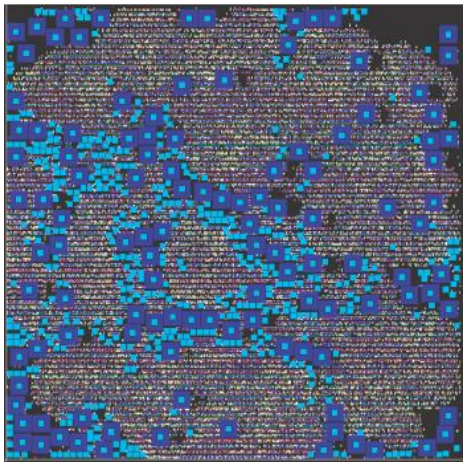




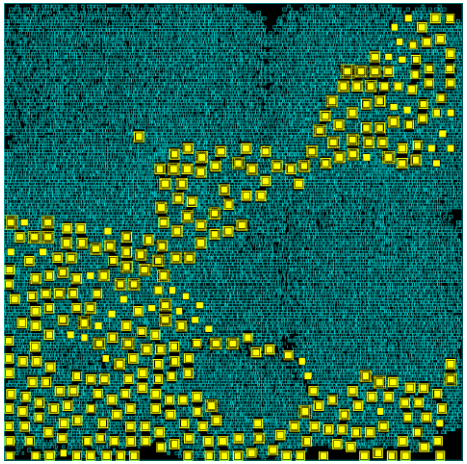
TSV Noise Reduction Techniques

- **We propose guard ring and differential TSV for noise reduction**

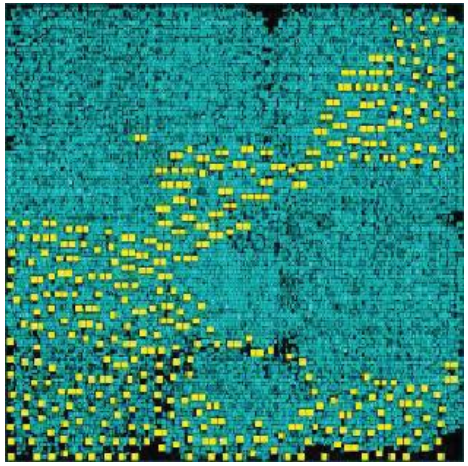
TSV shielding in [1]



Guard Ring Protection



Differential TSV



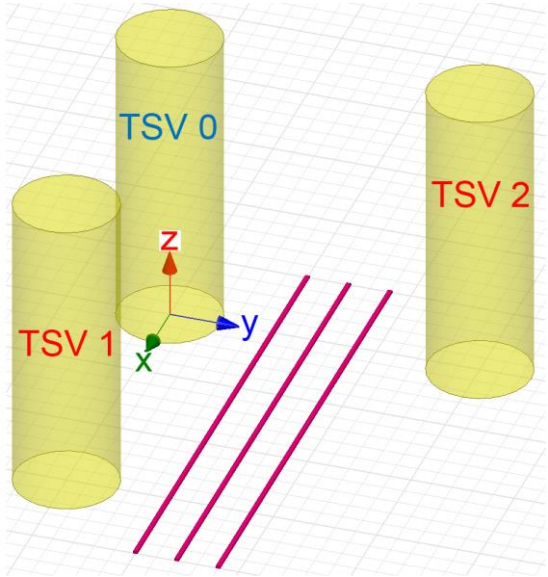
	TSV shielding	Guard ring	Differential TSV
Protected TSV	118/548	298/330	110/330
Area of original TSV (μm^2)	49	49	49
Area of optimized TSV (μm^2)	361	68.9 ~ 121	105
Total TSV noise reduction	42.04%	27.3%	42.2%
Total area overhead	26.4%	7.65%	3.9%

[1] C. Liu et al., "Full-chip TSV-to-TSV coupling analysis and optimization in 3D IC," DAC11

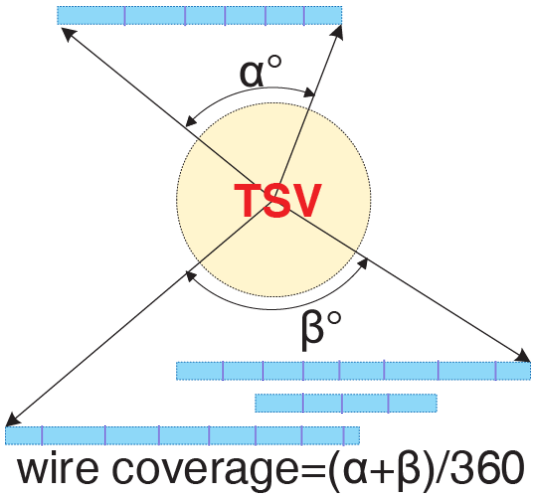
TSV-to-Wire Coupling Extraction

- **Uses pattern matching algorithm with multi-TSV and multi-wire awareness**
- **Highly accurate and fast**

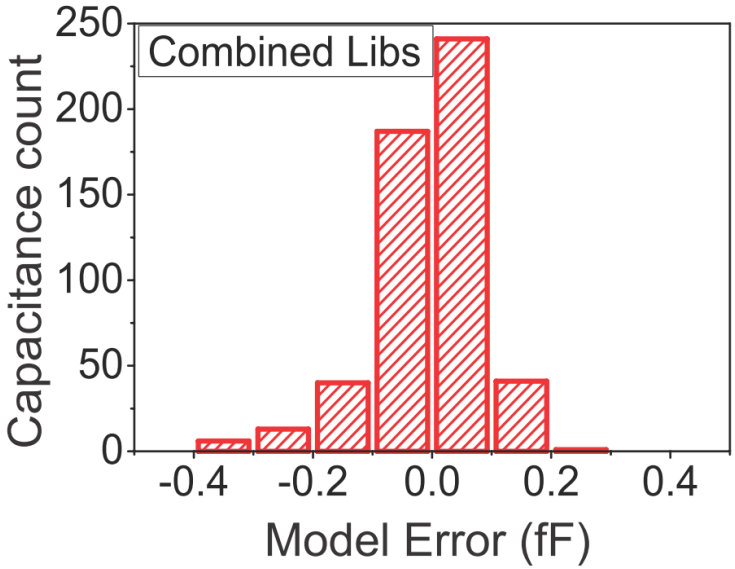
Library Calibration



Pattern matching algorithm



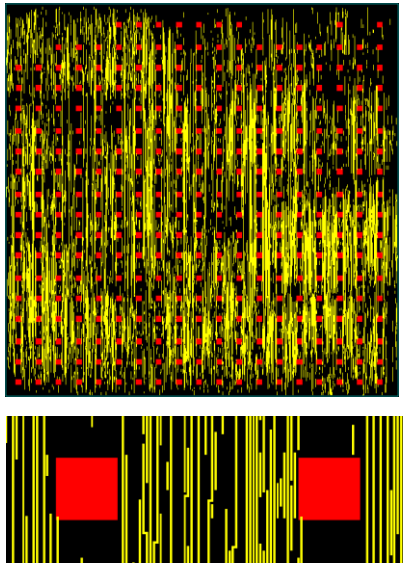
Fast (3.6s) and highly accurate against field solver



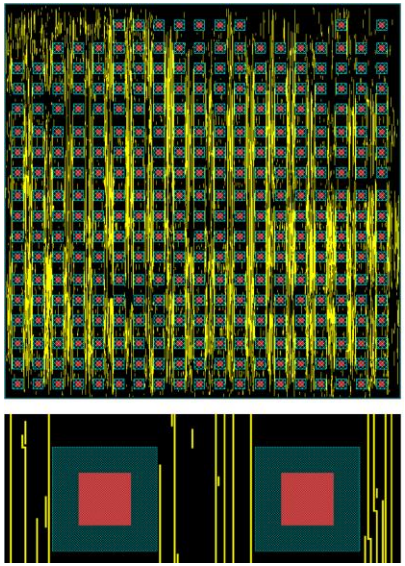
TSV-to-Wire Coupling Reduction

Guard ring width (μm)	0	0	0.5	1.5
Routing KOZ size (μm)	0	2.5	2.5	2.5
Longest path delay (ns)	4.83	4.77 (-1.2%)	4.84 (+0.21%)	4.90 (+1.45%)
Total TSV net power (mW)	0.335	0.326 (-2.7%)	0.340 (+1.49%)	0.349 (+4.18%)
Total net switching power (mW)	2.46	2.45 (-0.4%)	2.46 (+0%)	2.47 (+0.41%)
Total noise on TSV net (V)	61.2	54.4 (-11.1%)	48.1 (-21.4%)	45.5 (-25.6%)

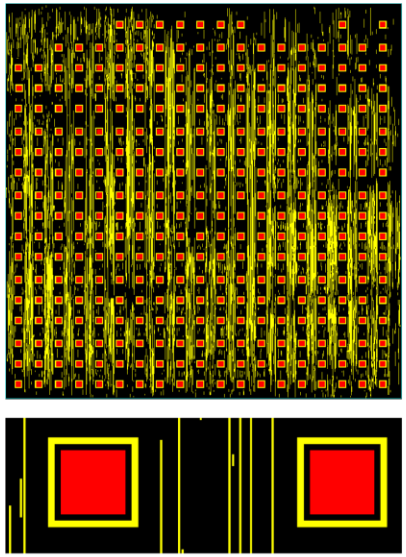
Original Design



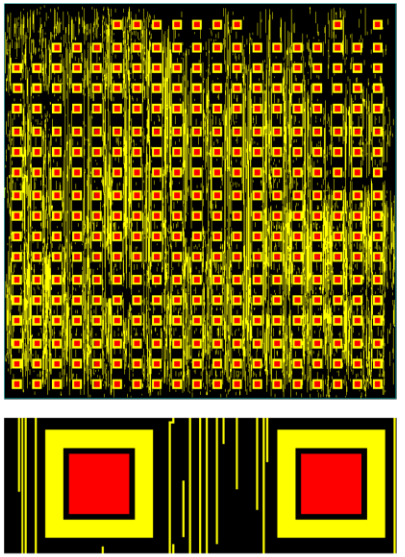
2.5μm KOZ only



+ 0.5μm guard ring

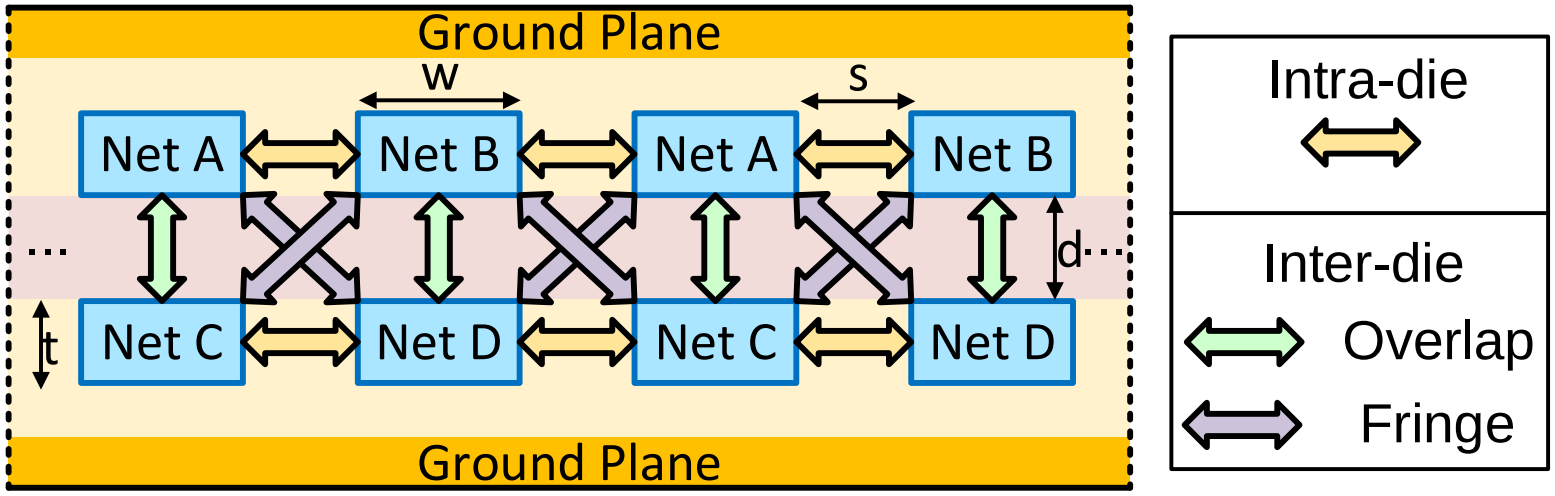


+ 1.5μm guard ring



Coupling Elements in F2F Structure

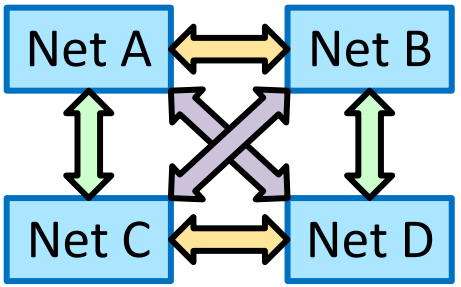
- **To analyze the trend in F2F structure, we build a test structure in Raphael with repeated pattern**
  - **Wire dimensions are based on M6 dimensions in a 45nm technology**
  - **Intra-die coupling: AB cap and CD cap**
  - **Inter-die overlap coupling: AC cap and BD cap**
  - **Inter-die fringe coupling: AD cap and BC cap**

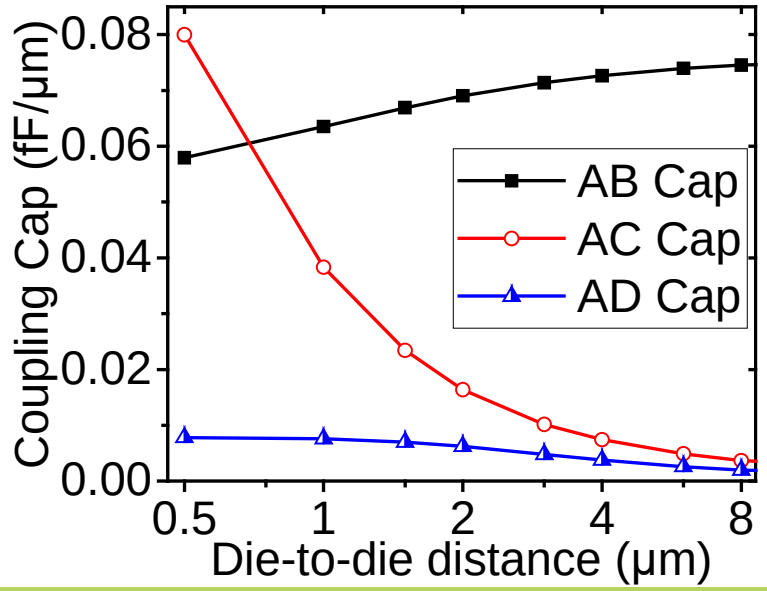


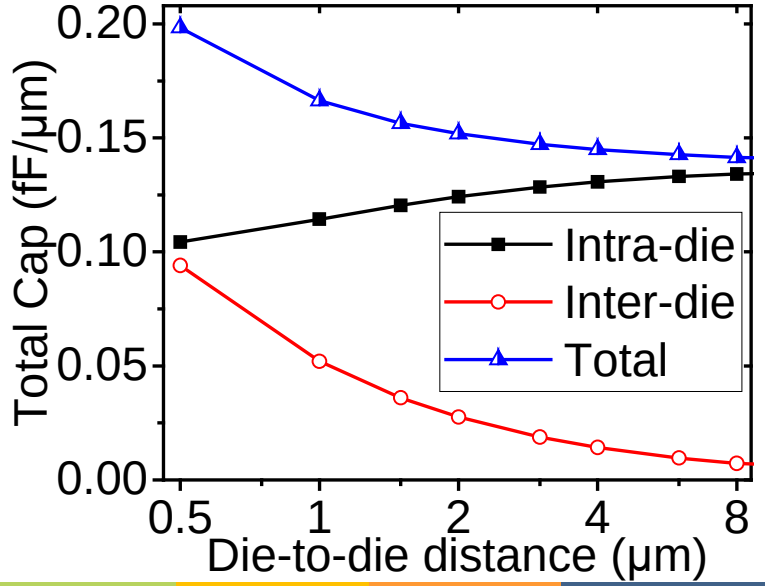
Raphael simulation structure

Die-to-die Distance Impact

- **With a closer die-to-die distance:**
  - **Intra-die cap (AB Cap) decreases due to stronger E-field sharing**
  - **Inter-die cap (AC+AD Cap) increases significantly**
  - **Inter-die overlap cap (AC cap) increases much more than fringe cap (AD cap)**

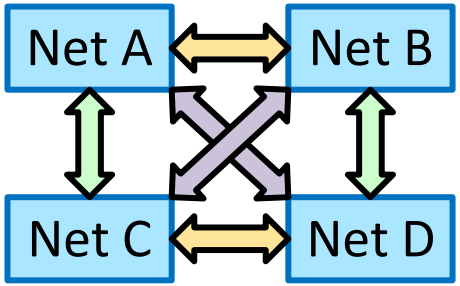


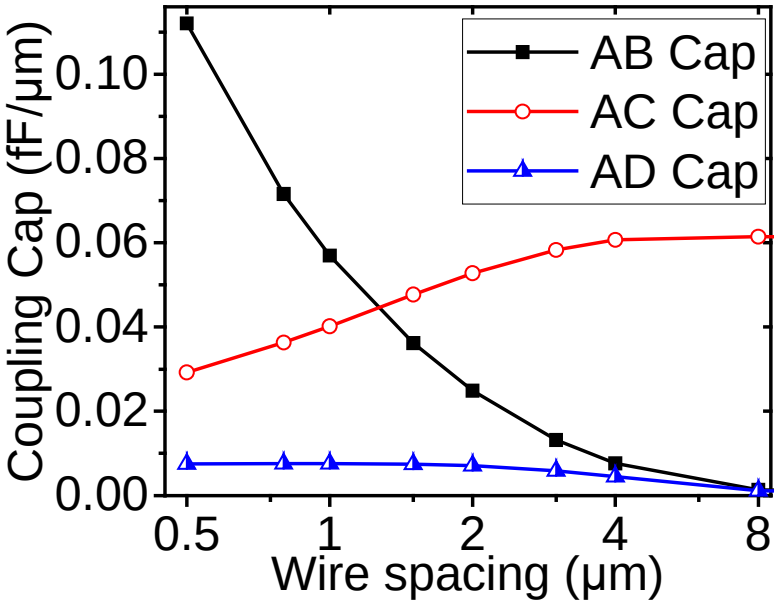


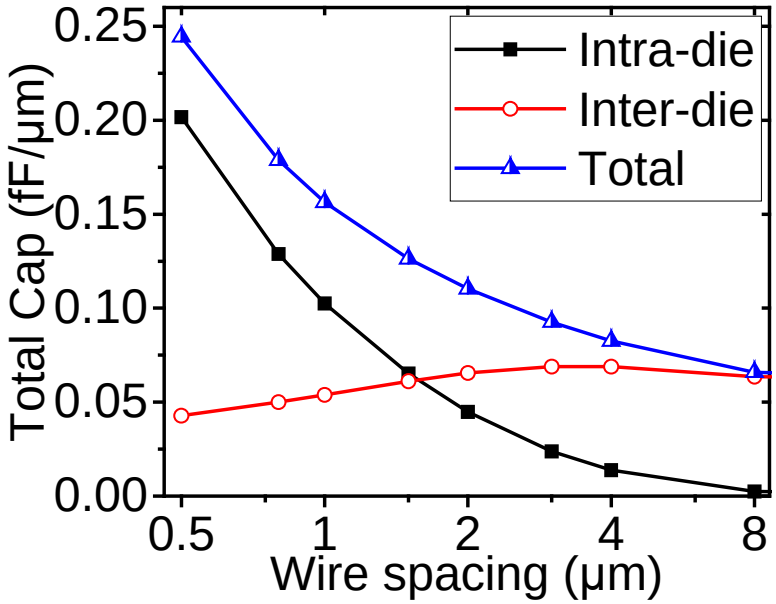


Wire Spacing Impact

- **With a larger wire-to-wire distance**
  - **Both intra-die coupling (AB cap) and total cap reduces**
  - **Inter-die first increases with larger overlap cap (AC cap) due to weaker E-field sharing then slightly decrease due to smaller fridge cap (AD cap)**

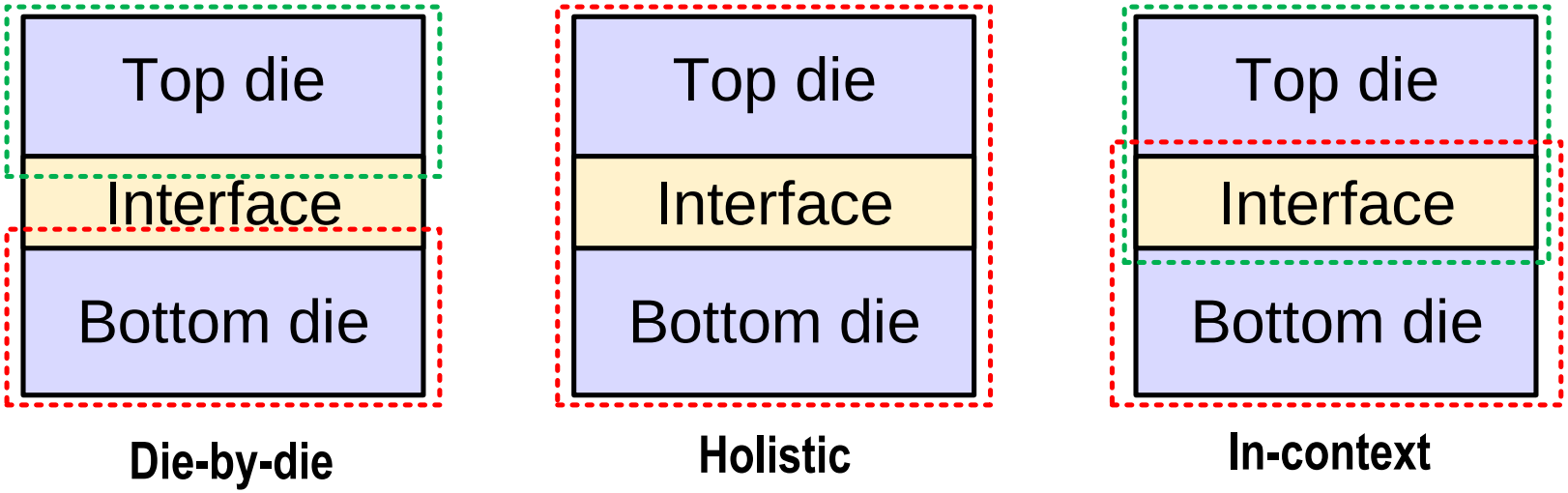






Three Ways of Full-chip F2F Extraction

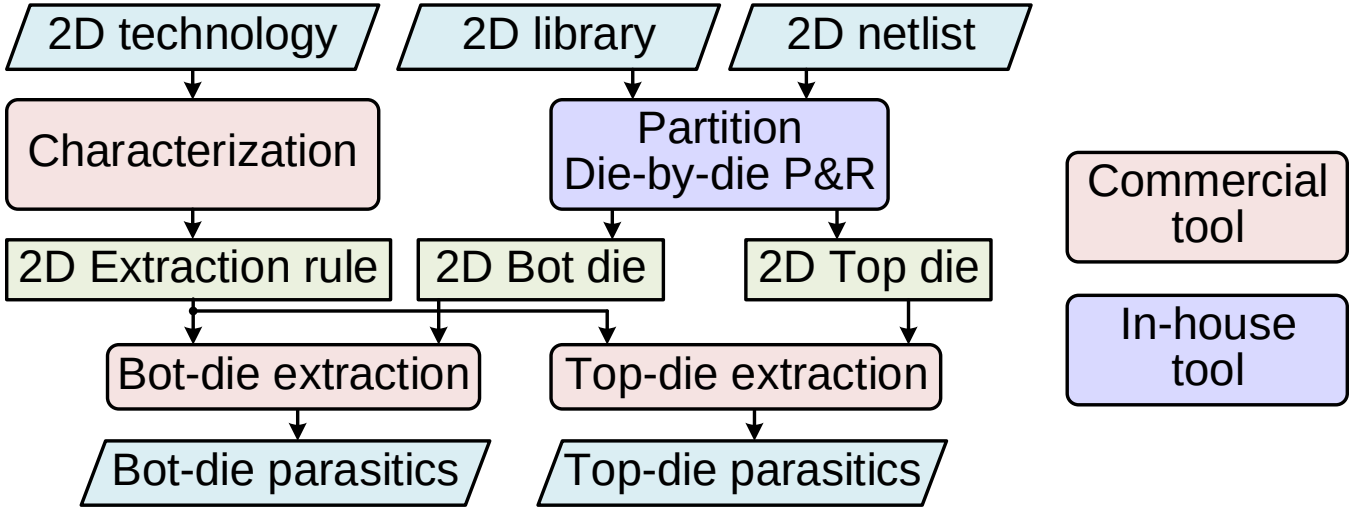
- **Die-by-die extraction**
  - **Extract dies separately**
- **Holistic extraction[1]**
  - **Extract all layers simultaneously**
- **In-context extraction**
  - **Extract each die separately but aware of a few neighboring die layers**



[1] Taigon Song and Sung Kyu Lim, "Die-to-Die Parasitic Extraction Targeting Face-to-Face Bonded 3D ICs," Journal of Information and Communication Convergence Engineering, Vol. 13, No. 3, pp. 172-179, 2015

Die-by-die Extraction Flow

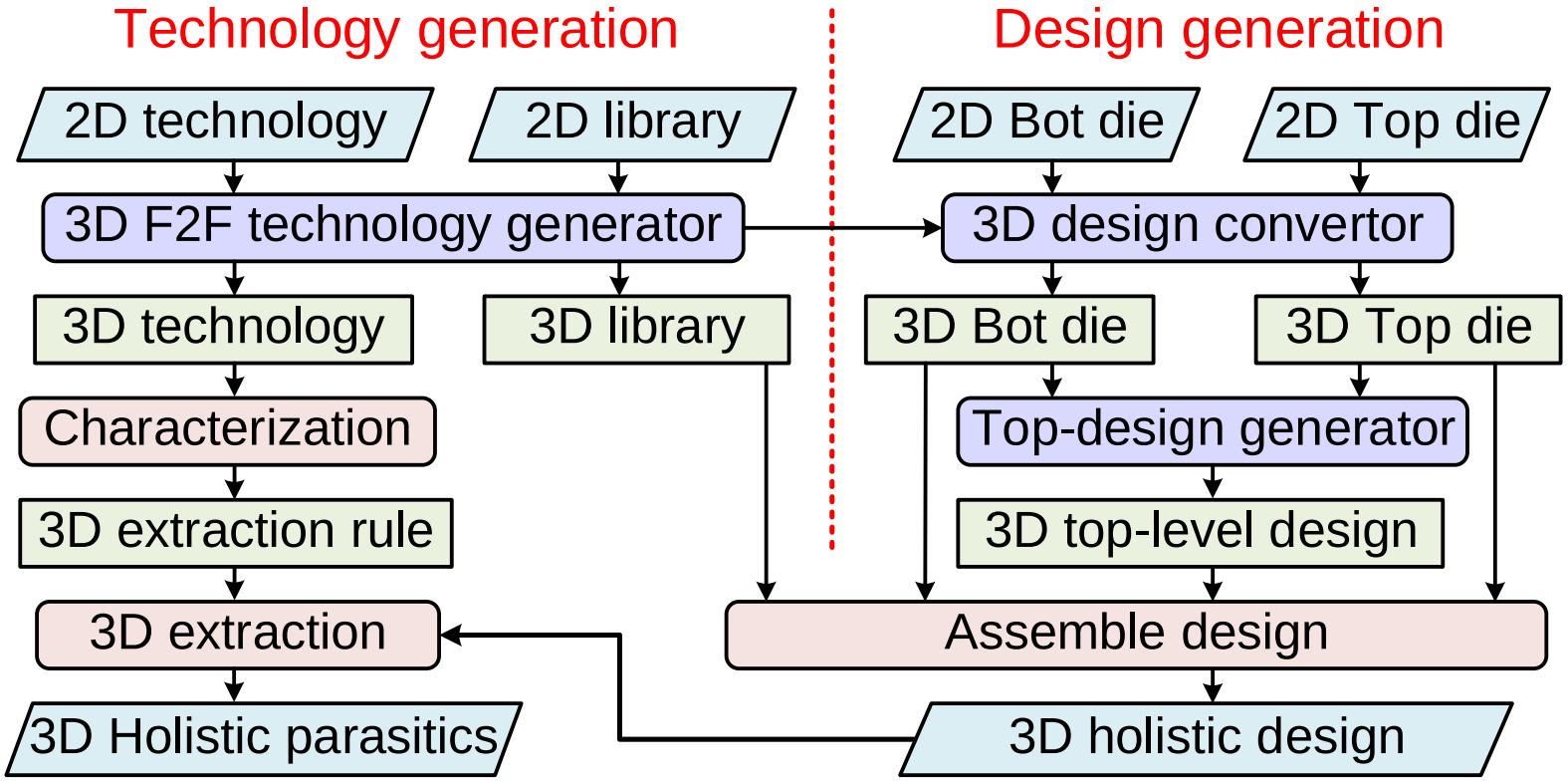
- **Die-by-die extraction is the straight-forward flow currently enabled by many CAD tools**
  - **Assumes each die can be extracted separately**
  - **Ignores all parasitic between dies**
  - **Accurate when dies are separated far or have a ground layer in between**



Die-by-die extraction flow

Holistic Extraction Flow

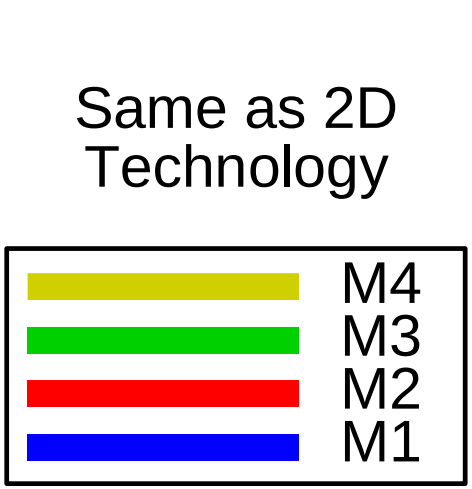
- **Holistic extraction takes all layers into consideration and it introduces more CAD and LVS complexity**



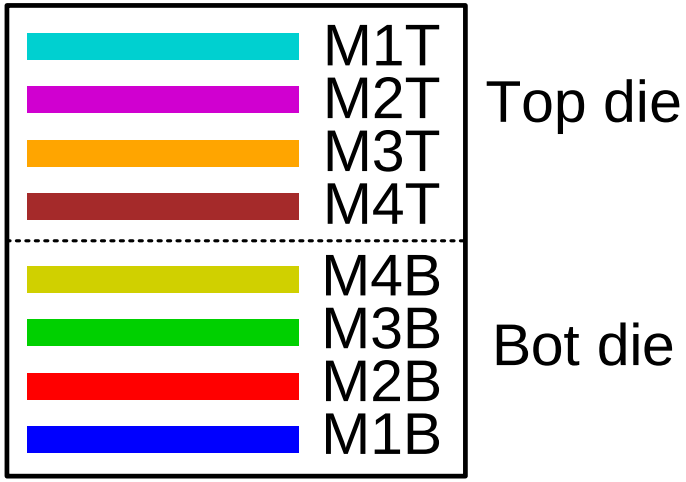
Holistic extraction flow

Die-by-die vs. Holistic Extraction

- **Die-by-die uses the same metal stack as 2D technology**
  - **Enables reuse of existing DRC, LVS and PEX rule decks**
- **Holistic extraction needs to rebuild rule decks**
  - **All original and derived layers and device renamed and remapped**
  - **Need technology recalibration**



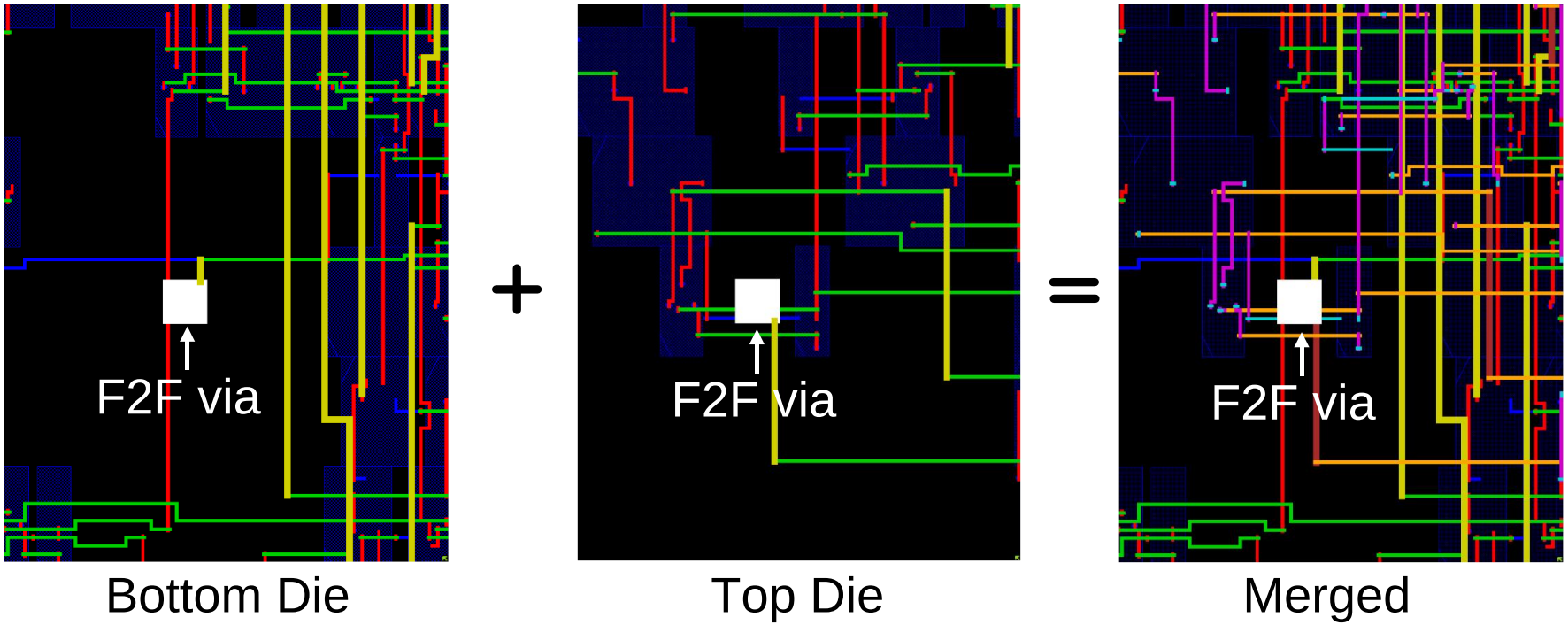
Die-by-die layer structure



Holistic layer structure

Holistic Design Example

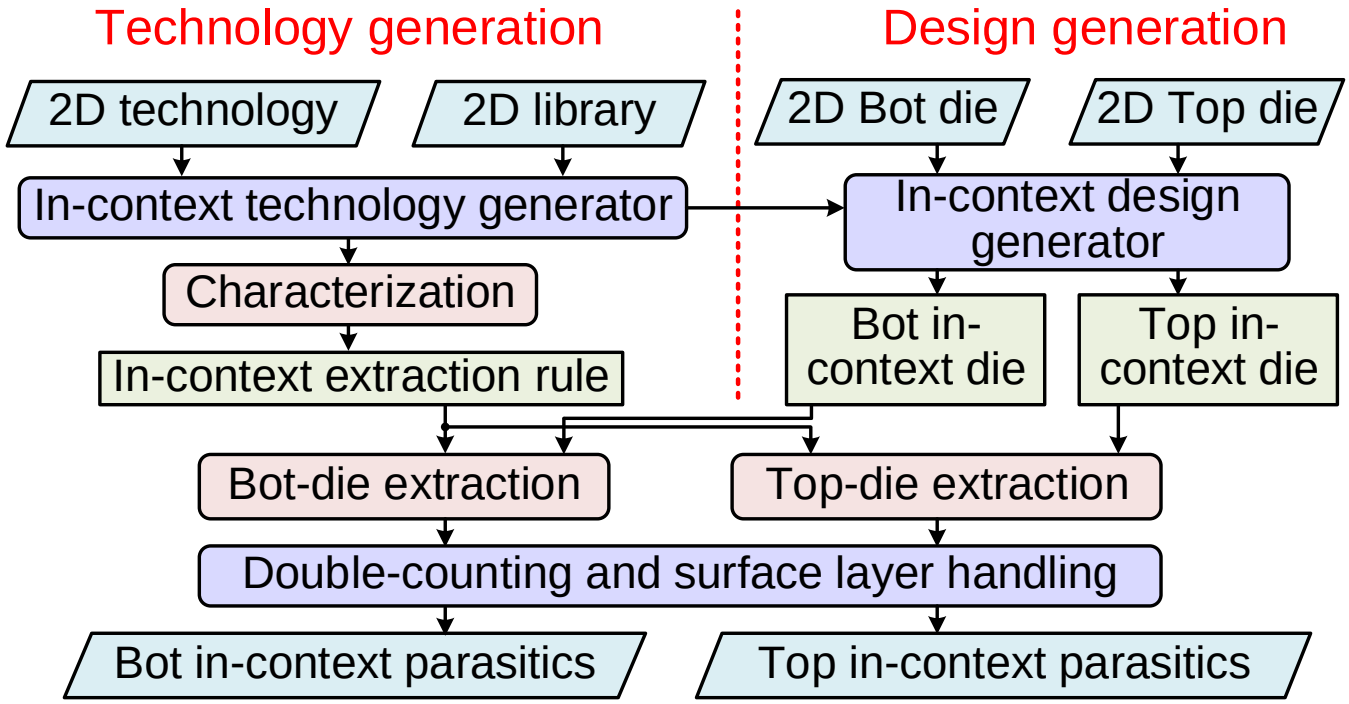
- **By assembling of individual dies, we are able to create a holistic design which contains all metal layers**



Design assemble process

In-context Extraction Flow

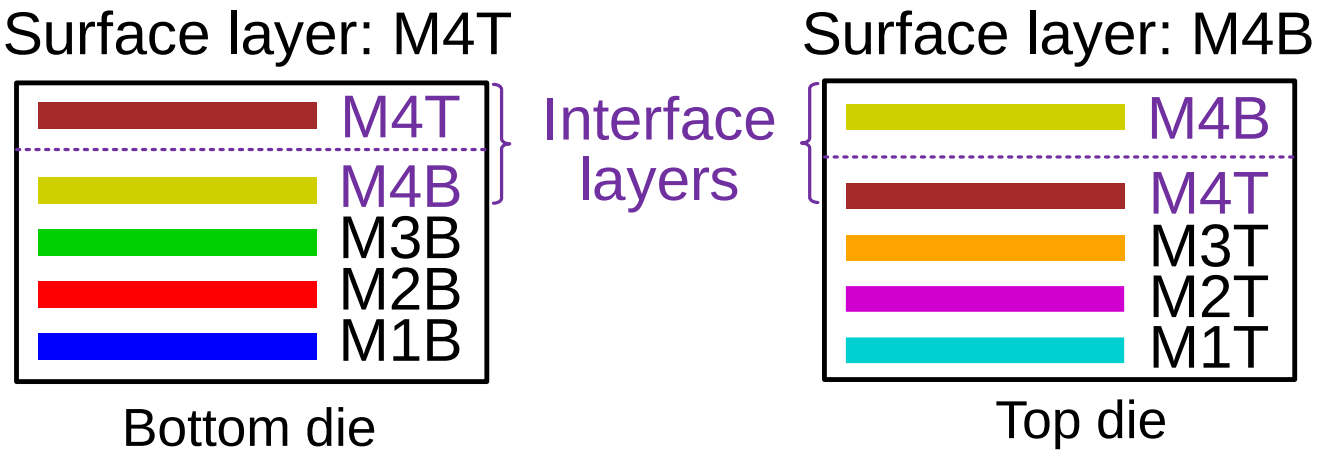
- **In-context extraction takes in a few metal layers from the neighboring die as interface layers**
  - **Keeps most of inter-die coupling and remains accurate**
  - **Reduces CAD complexity and compatible with current tool flow**



In-context extraction flow

In-context Design Example

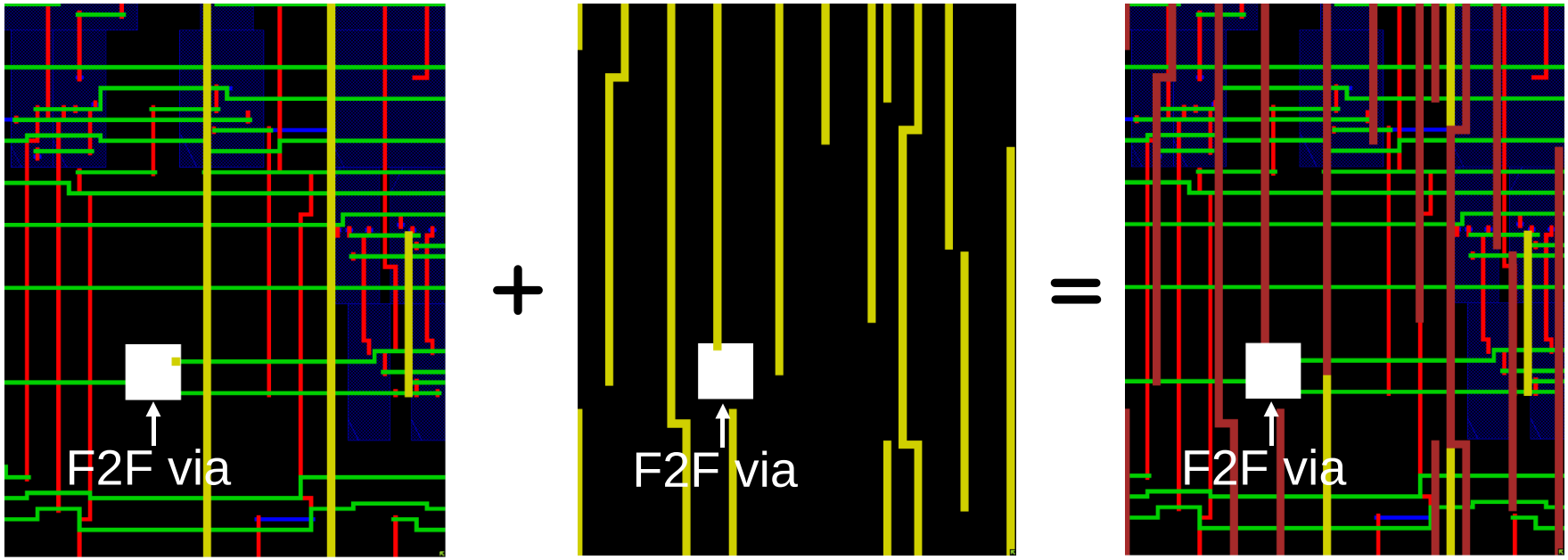
- **In-context technology can be calibrated incrementally**
  - **Base layer calibration results can be derived from existing rule decks**
  - **The surface layer in the in-context extraction is defined as the layer furthest from the substrate**



In-context layer structure
(with one interface layer from each die)

In-context Design Example

- **In-context design only needs additional routing information from the neighbor die**
  - **Enables much simpler rule deck generation**

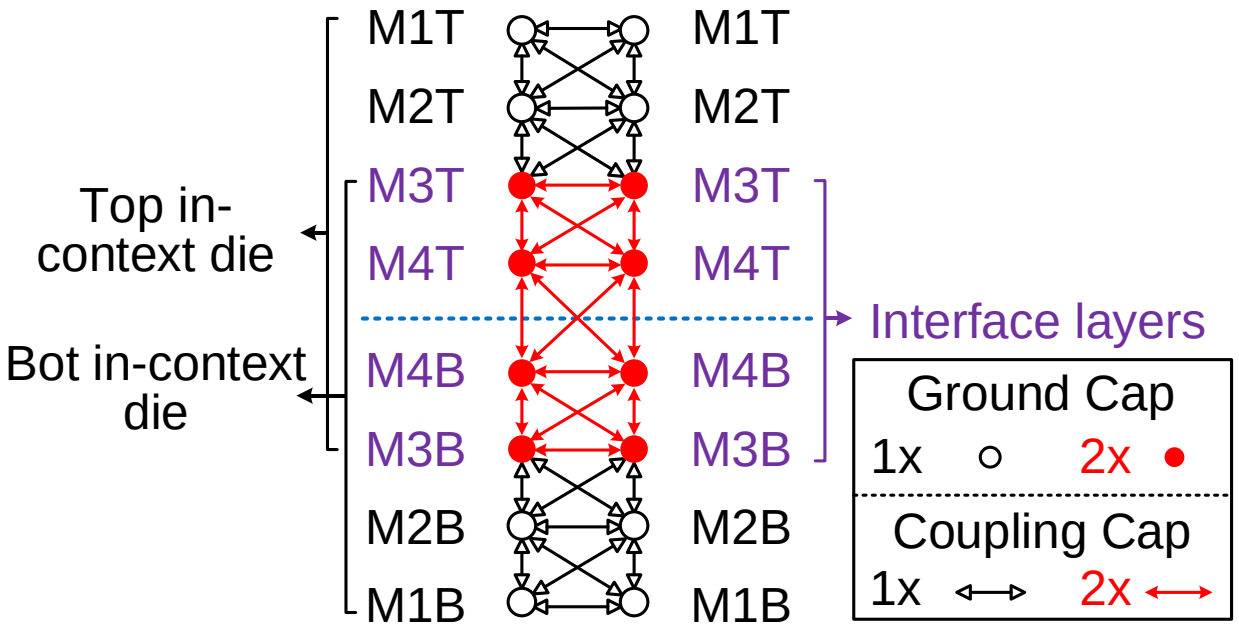


Bottom Die Top Interface layer Bottom In-context die

In-context design generation

Double Counting Correction

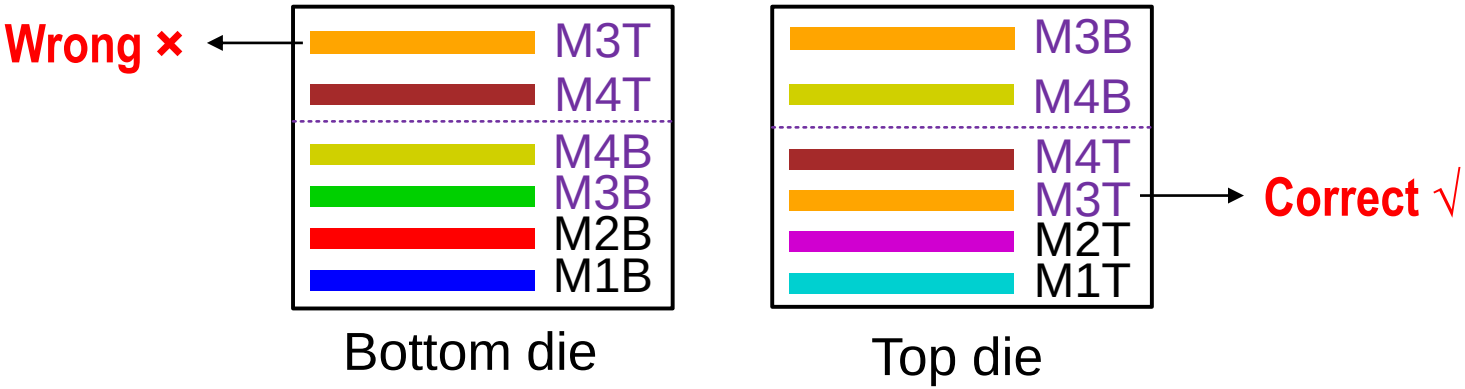
- **With in-context extraction, capacitance on interface layers are double-counted**
  - **A simple solution is to halve all caps from interface layers in SPEF files**



Double-counted capacitance with in-context extraction

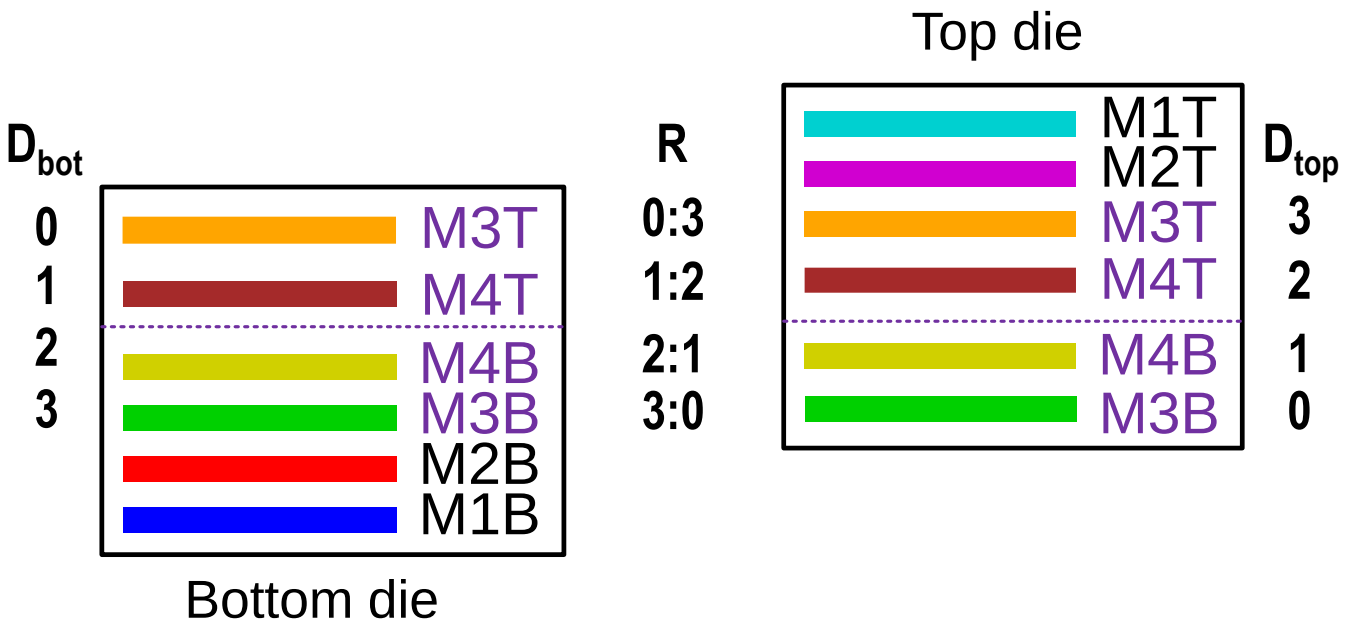
Surface Layer Correction

- Surface layer only sees one neighboring layer
  - Introduce large error with less E-field sharing
- Note each layer is not the surface layer in both in-context dies
  - E.g., M3T is the surface layer in bottom die but not in top die
- Surface layer correction based on weighted average
  - Use a weighted average for caps on interface layers
  - Larger weights for layers farther from the surface



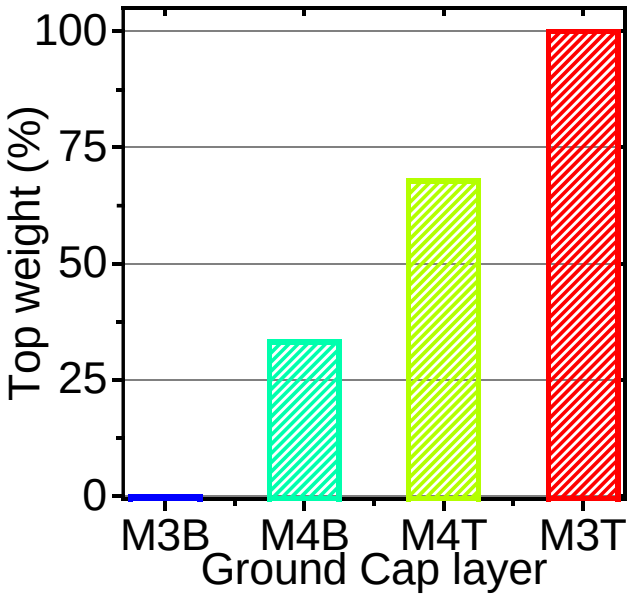
Surface Layer Correction

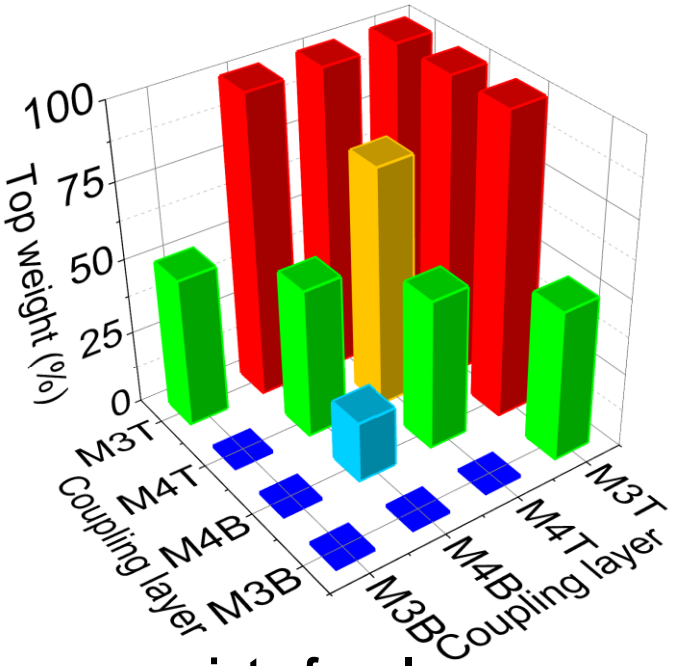
- **For each layer, we define:**
  - **D: distance to the surface**
  - **R: ratio between D values in the bottom and top in-context die**
- **Example (two interface layers per die)**

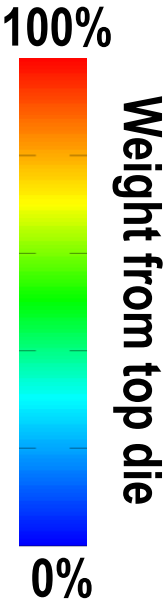


Surface Layer Correction

- **The surface correction weight of a capacitor is the product of R ratios of both its landing layers, normalized to 100%**
  - **The R ratio of ground layer is defined as 1:1**
  - **A ground cap on M4T: $R_{M4T} \times R_{gnd}$ = 1:2 × 1:1 = 33% (bot) : 67% (top)**
  - **A cap between M4T and M4B: $R_{M4T} \times R_{M4B}$ = 1:2 × 2:1 = 50% (bot) : 50% (top)**





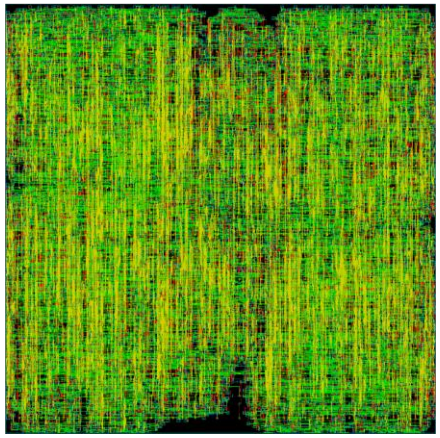


Weighted average for caps on interface layers

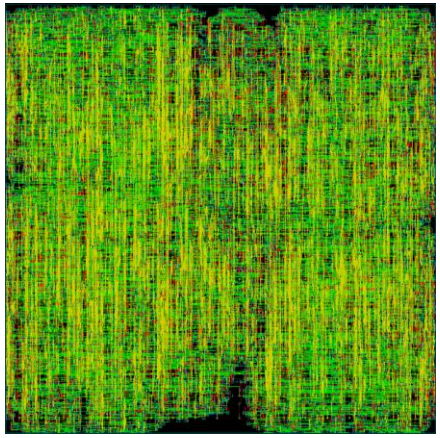
Sample FFT Design in F2F

- **A 64-point FFT with 38K gates and 330 F2F vias is implemented**

Die-by-die

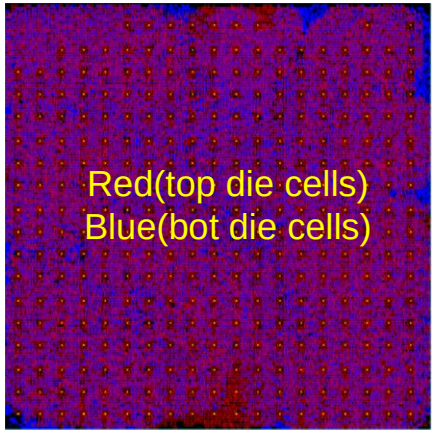


Bottom Die

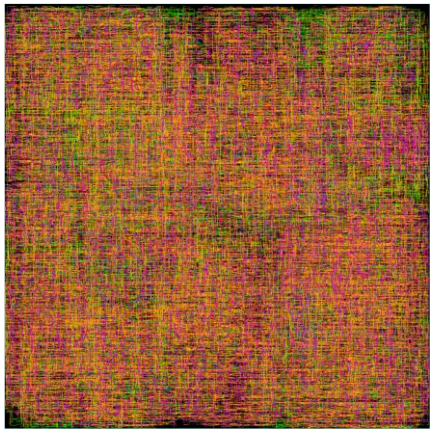


Top Die

Holistic



Placement



Routing

Breakdown of Holistic Extraction

- **Inter-die coupling occupies a large portion of total coupling cap**
  - **Especially when dies are close and few metal layers are used**

Layer	M1B	M2B	M3B	M4B	M4T	M3T	M2T	M1T
M1B	5.76	3.03	17.1	0.13	0.03	0.14	0.00	0.00
M2B	3.03	381	147	396	18.6	0.69	2.58	0.01
M3B	17.1	147	1261	231	9.9	140	0.72	0.28
M4B	0.13	396	231	1826	1184	18.6	46.7	0.12
M4T	0.03	18.6	9.9	1184	1311	196	369	0.28
M3T	0.14	0.69	140	18.6	196	1226	148	25.3
M2T	0.00	2.58	0.72	46.7	369	148	442	4.63
M1T	0.00	0.01	0.28	0.12	0.28	25.3	4.63	7.54

Breakdown of holistic extraction

*Cap units are in fF

Comparison of Three Methodologies

- With a 1um F2F via height, die-by-die extraction underestimates coupling capacitance significantly
  - Especially for layers close to the other die
- More interface layers in in-context extraction help improve accuracy
  - Two interface layers per die provide a good tradeoff

Layer	Holi	D-D	In-C:1	In-C:2	In-C:3
Total	12381	8032	12018	12287	12292
Err%	-	-35.00%	-2.93%	-0.76%	-0.72%

Total Coupling Extraction Comparison

*Cap units are in fF

Interface Layer Impact

- **Simply dividing all double counted capacitance by half is not accurate enough**
- **Our weighted methods improves in-context extraction accuracy**

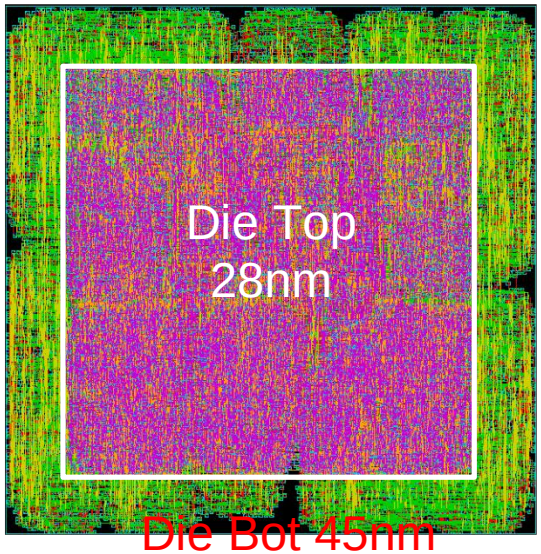
Layer	M3B	M4B	M4T	M3T	Total	ERR%
Holi	1808	3703	3089	1755	10354	-
Original	3069	6779	5781	3522	19151	+84.9%
Halved	1618	3611	3082	1849	10159	-1.89%
Weighted	1803	3679	3058	1734	10272	-0.79%

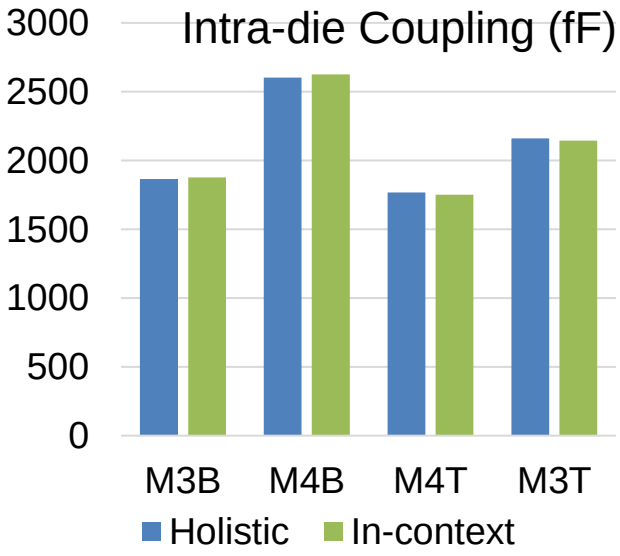
*Cap units are in fF

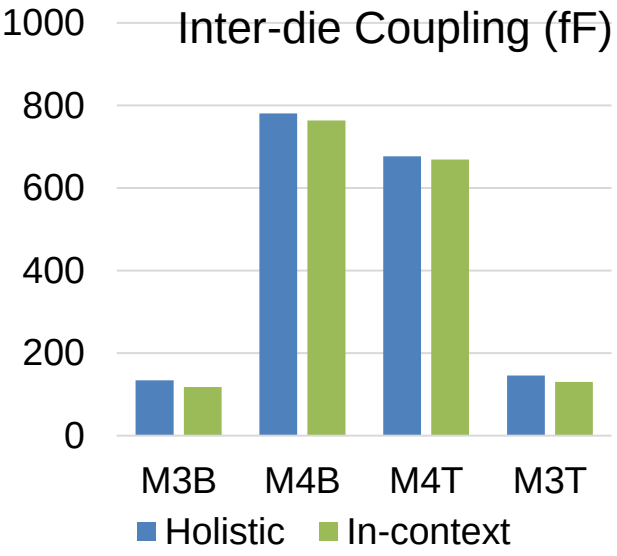
Heterogeneous In-Context Extraction

- **Holistic extraction has problems handling non-overlapping areas**
- **Our in-context extraction is also highly accurate on heterogeneous F2F designs with mixed technology nodes**

Mixed-node design

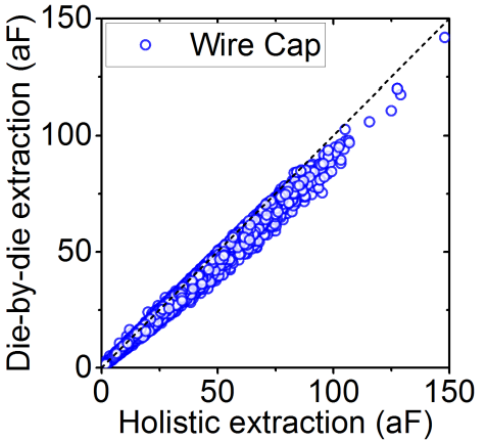


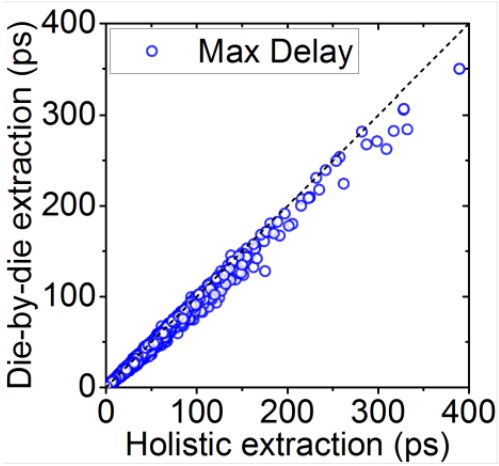


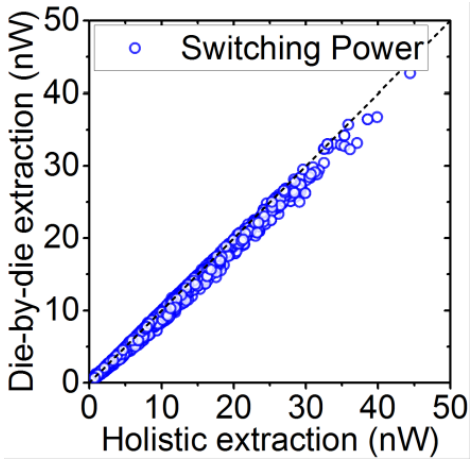


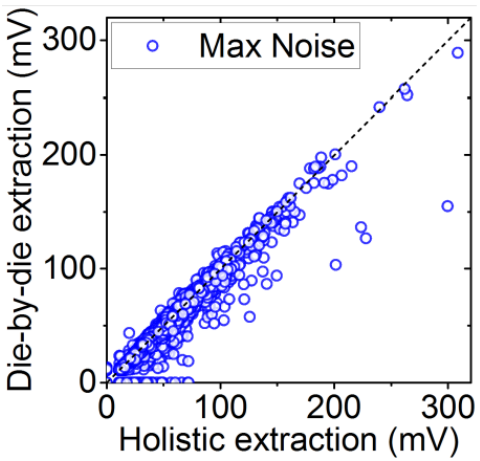
Die-by-die Extraction Error in Full-chip

- **With die-by-die extraction, underestimated inter-die coupling results in underestimated timing, power and noise**



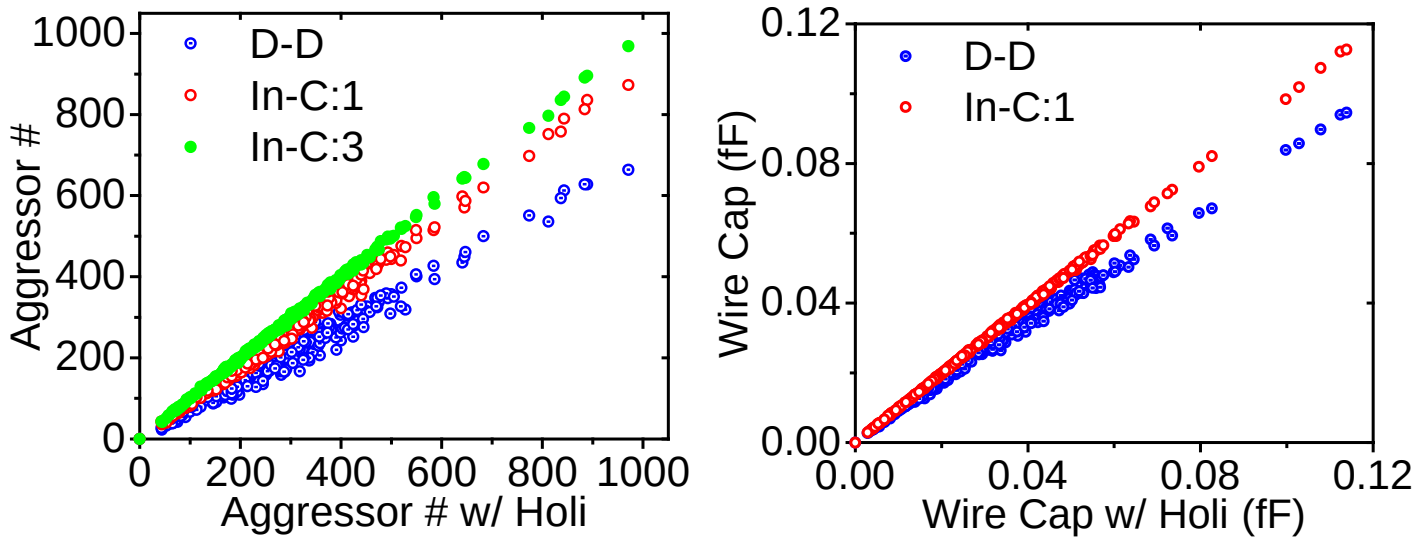






In-context Extraction Accuracy

- **In-context extraction captures inter-die aggressors, provides better accuracy in full-chip analysis**
  - **Especially for 3D nets which communicate across dies**



In-context extraction of 3D nets

Full-chip Analysis Summary

- **Full-chip analysis also shows non-negligible impact from inter-die capacitance, especially on noise results and 3D nets**
  - **Die-by-die extraction underestimates delay, power and noise**
  - **In-context extraction gives much more accurate results**

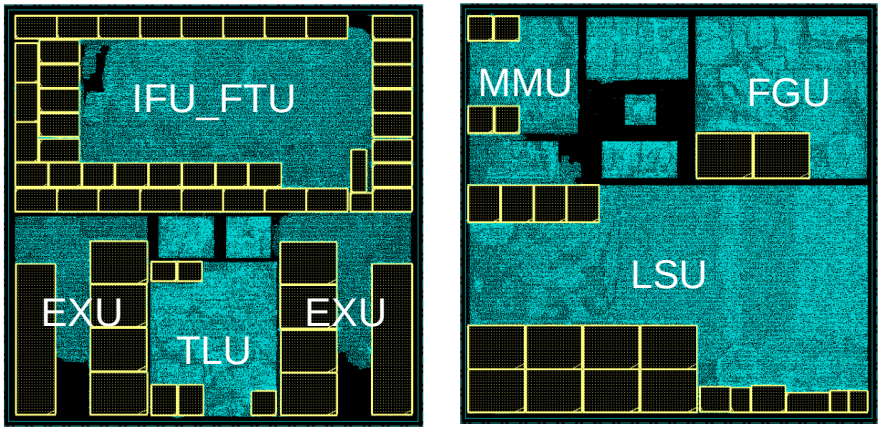
Primetime measurement	Holi	D-D	Err%	In-C	Err%
Longest path delay (ns)	3.90	3.66	-6.2%	3.81	-2.3%
Total switching power (mW)	12.1	11.9	-1.7%	12.0	-0.7%
Impacts on 3D Nets					
Switching power (mW)	1.05	1.01	-3.5%	1.04	-0.5%
Total coupling cap (fF)	4.37	2.96	-32%	4.19	-4.1%
Total wire cap (fF)	10.8	9.35	-13%	10.6	-1.8%
Avg aggressor #	285	200	-30%	277	-2.8%
Max noise (mV)	41.3	30.40	-26%	38.8	-6.1%

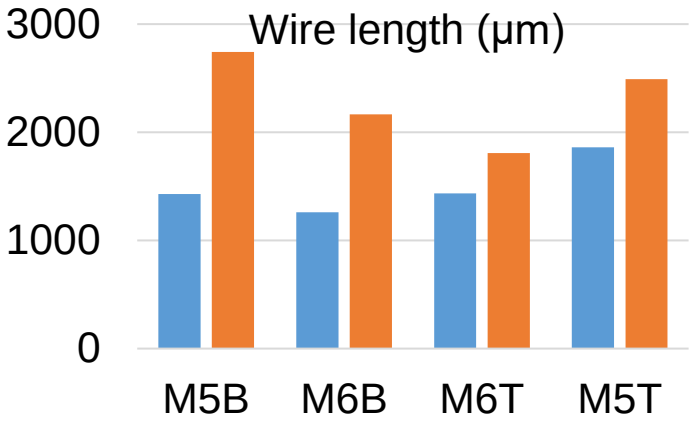
*single interface layer is used

Partition Scheme Impact

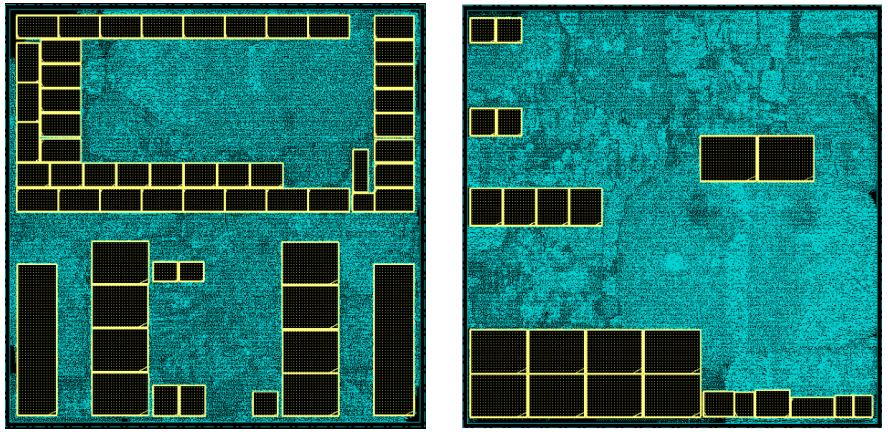
- **Partition affects F2F via count and routing in top metal layer**

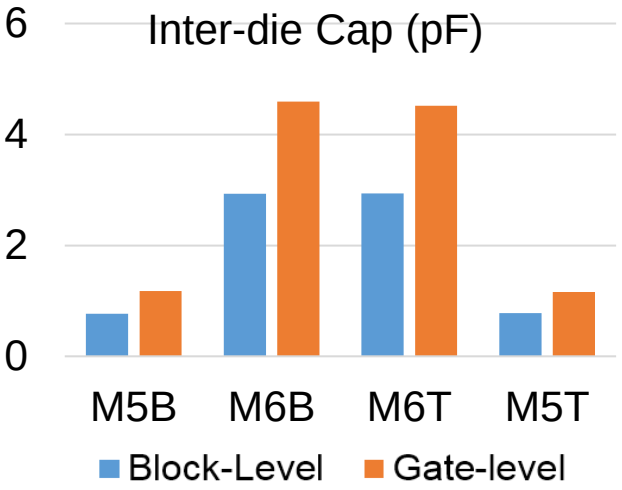
Block-level 2967 F2F Vias





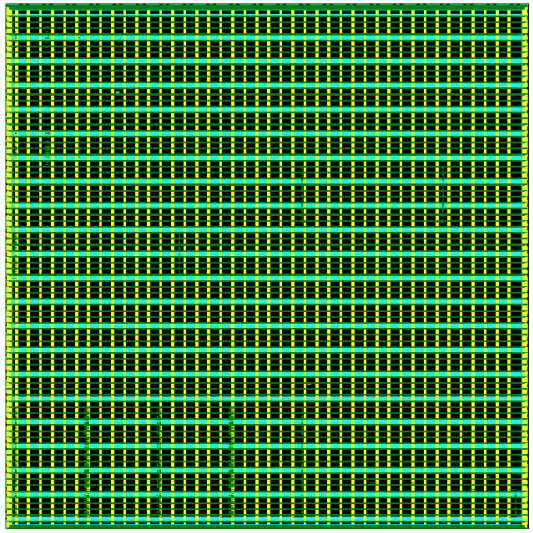
Gate-level 11201 F2F Vias



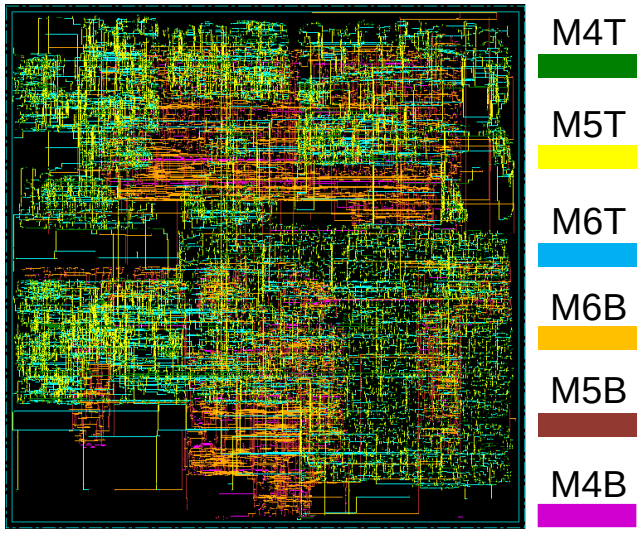


Inter-die Impacts on Special Nets

- **A large portion of inter-die coupling is on PDNs and Clock Tree**



PDN



Clock Tree

Cap (pF)	Signal		Clock		Power	
Layer	intra-die	inter-die	intra-die	inter-die	intra-die	inter-die
Total	232.5	15.18	18.01	1.7	17.84	1.07

Inter-die Impacts on Special Nets

- **Ignoring inter-die coupling underestimates clock tree delay and introduces large error in clock slew calculation**

Design	Block-level			Gate-level		
Extraction	D-D	Holi	Δ%	D-D	Holi	Δ%
WNS (ns)	-0.07	-0.05		-0.06	-0.10	
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%

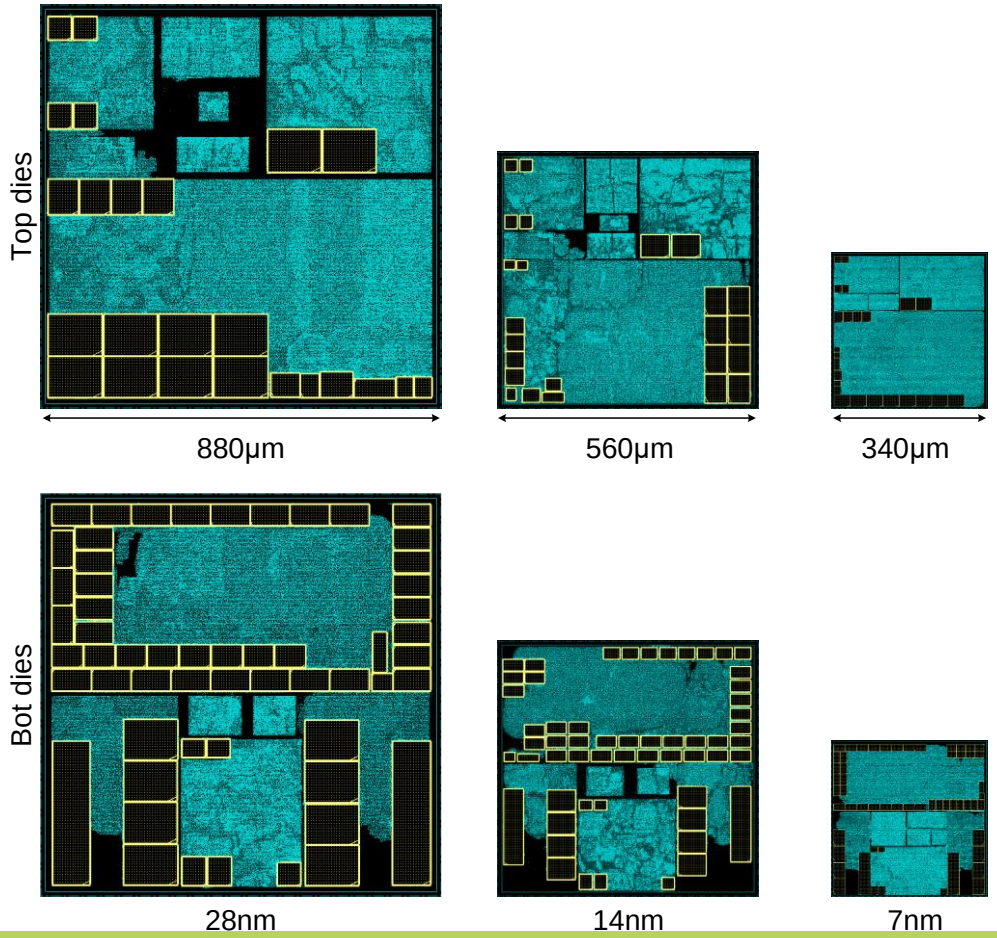
- **Increase die-to-die distance and PDN usage can be used to reduce inter-die coupling because of E-field sharing**

Top layer	M6	M7			
PDN layer	M4-M6	M4-M6	Δ%	M4-M7*	Δ%
Total	18.87	8.2	-56.5%	6.5	-65.6%

*20% PDN usage on M7

Technology Scaling Impact

- **We predict that inter-die coupling remains important with advanced technology nodes because of die-to-die distance scaling**

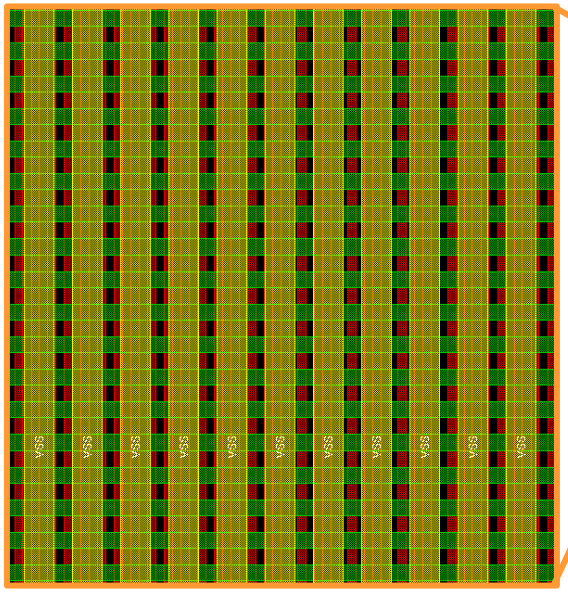


Tech Node	Die gap	Intra-die	Inter-die	Inter-die%
28nm	1.0μm	208.3	8.11	3.75%
	0.7μm	207.3	11.0	5.03%
14nm	0.7μm	59.5	1.42	2.33%
	0.5μm	59.3	1.99	3.25%
7nm	0.5μm	37.6	1.00	2.59%
	0.35μm	37.4	1.45	3.73%

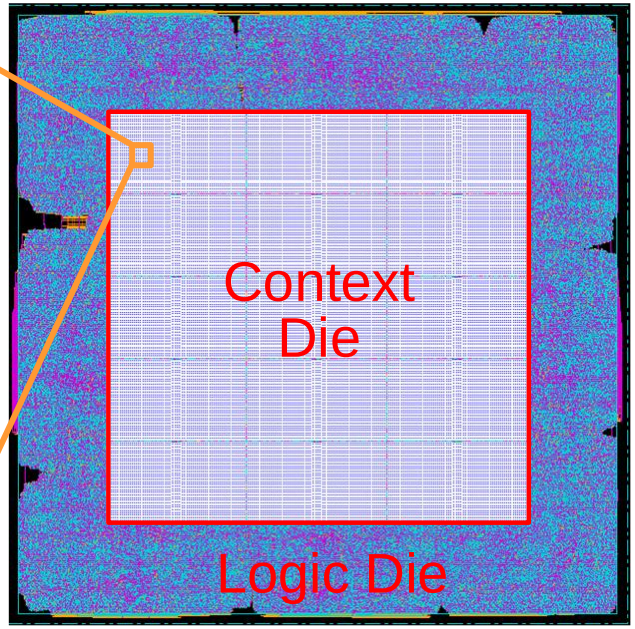
*Cap units are in pF

Context Creation in Early Design Stage

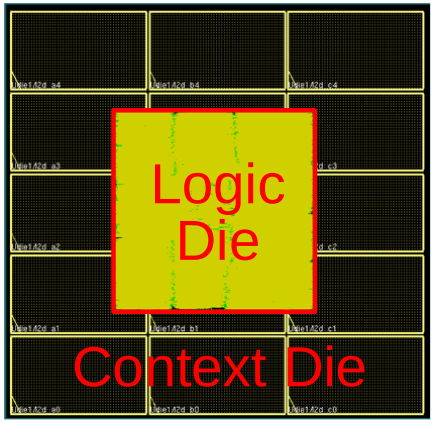
- **Previous extraction requires LVS-clean layouts from both dies, which are not available in early design stage**
- **Since the memory die has a highly regular layout, it is possible to automatically generate layouts to mimic memory context**



Auto-generated
Memory Context die



45nm logic+28nm SRAM



14nm logic+28nm SRAM

Extraction with Context Memory Die

- **In early design stage, if context is ignored, parasitics on the logic die are underestimated because of incorrect context**
- **With auto-generated memory die, which mimics the memory context, extraction error is significantly reduced**

w/ real memory GDS			wo/ context		w/ context	
Layer	GCap	CCap	GCap	CCap	GCap	CCap
Total	489.1	193.9	477.1	198.8	491	193.1
Err%	-	-	-2.46%	2.51%	0.39%	-0.41%

45nm logic+28nm SRAM

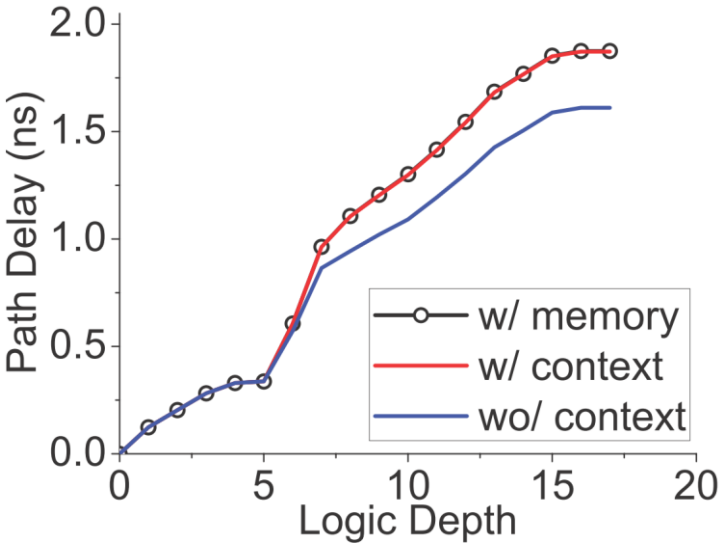
w/ real memory GDS			wo/ context		w/ context	
Layer	GCap	CCap	GCap	CCap	GCap	CCap
Total	210.9	80.8	204.4	83.5	209.1	80.7
Err%	-	-	-3.08%	3.38%	-0.87%	-0.13%

14nm logic+28nm SRAM

*Cap units are in fF

Full-chip Extraction with Context

- **With the context die, timing and power analysis of the logic die is highly accurate compared with real GDS layouts**



Delay propagation of the longest path

Design	w/ GDS	wo/ Context	Err%	w/ Context	Err%
LPD (ns)	1.875	1.611	-14.1%	1.872	-0.16%
Power (mW)					
Net	135.6	128.8	-5.01%	137.8	1.62%
Cell	798.0	797.2	-0.10%	798.4	0.05%
Total	940.5	932.9	-0.81%	943.0	0.27%

Full-chip timing and power analysis

Conclusions and Future Work

- Inter-die coupling cannot be ignored in F2F-bonded 3D ICs, with a large impact on full-chip timing, power and noise
- Three extraction methods are compared with full-chip analysis
  - Die-by-die extraction underestimates total coupling capacitance
  - Holistic extraction is able to capture all inter-die coupling at the cost of high complexity
  - In-context extraction is highly accurate, and captures most E-field interactions across dies. Further, it handles heterogeneous integration
- Physical design style has large impact on inter-die coupling
- Context creation methodology is effective to reduce extraction error in early design stage
- Future work: 2.5D Wafer-level package extraction with both capacitance and inductance extraction

Publications (1/2)

Journal

[1] **Yarui Peng**, Taigon Song, Dusan Petranovic, and Sung Kyu Lim, "Silicon Effect-aware Full-chip Extraction and Mitigation of TSV-to-TSV Coupling," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 33, no. 12, pp.1900-1913, Dec. 2014
[2] Sandeep Samal, **Yarui Peng**, Mohit Pathak, and Sung Kyu Lim, "Ultra-Low Power Circuit Design with Sub/Near-Threshold 3D IC Technologies," IEEE Transactions on Components, Packaging, and Manufacturing Technology, vol.5, no.7, pp.980-990, July 2015
[3] **Yarui Peng**, Dusan Petranovic, and Sung Kyu Lim, "Multi-TSV and E-Field Sharing Aware Full-chip Extraction and Mitigation of TSV-to-wire Coupling," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.
[4] Moongon Jung, Taigon Song, **Yarui Peng**, and Sung Kyu Lim, "Fine-Grained 3D IC Partitioning Study with A Multi-core Processor," IEEE Transactions on Components, Packaging, and Manufacturing Technology, vol.5, no.10, pp.1393-1401, Oct. 2015
[5] Taigon Song, Chang Liu, **Yarui Peng**, and Sung Kyu Lim, "Full-Chip Signal Integrity Analysis and Optimization of 3D ICs", IEEE Transactions on Very Large Scale Integration Systems.
[6] **Yarui Peng**, Taigon Song, Dusan Petranovic, and Sung Kyu Lim, “Parasitic Extraction for Heterogeneous Face-to-Face Bonded 3D ICs”, (submitted) IEEE Transactions on Components, Packaging, and Manufacturing Technology
[7] Moongon Jung, Taigon Song, **Yarui Peng**, and Sung Kyu Lim, “Design Methodologies for Low Power 3D ICs with Advanced Tier Partitioning”, (submitted) IEEE Transactions on Very Large Scale Integration Systems

Conference

[1] Taigon Song, Chang Liu, **Yarui Peng**, and Sung Kyu Lim, "Full-Chip Multiple TSV-to-TSV Coupling Extraction and Optimization in 3D ICs," ACM Design Automation Conference, 2013.
[2] Taigon Song, Chang Liu, **Yarui Peng**, and Sung Kyu Lim, "Full-Chip Multiple TSV-to-TSV Coupling Extraction and Optimization in 3D ICs," SRC TECHCON Conference, 2013.
[3] Sandeep Samal, **Yarui Peng**, Yang Zhang, and Sung Kyu Lim, "Design and Analysis of Ultra Low Power Processors Using Sub/Near-Threshold 3D Stacked ICs," International Symposium on Low Power Electronics and Design, 2013.
[4] **Yarui Peng**, Taigon Song, Dusan Petranovic, and Sung Kyu Lim, "On Accurate Full-Chip Extraction and Optimization of TSV-to-TSV Coupling Elements in 3D ICs," IEEE International Conference on Computer-Aided Design, 2013.

Publications (2/2)

Conference (continued)

[5] Sandeep Samal, Yarui Peng, and Sung Kyu Lim, "Design and Analysis of Ultra Low Power Processors Using Sub/Near-Threshold 3D Stacked ICs," SRC TECHCON Conference, 2014.
[6] Yarui Peng, Dusan Petranovic, and Sung Kyu Lim, "Fast and Accurate Full-chip Extraction and Optimization of TSV-to-Wire Coupling", SRC TECHCON Conference, 2014. ***Best in Session Award.***
[7] Moongon Jung, Taigon Song, Yang Wan, **Yarui Peng**, and Sung Kyu Lim, "On Enhancing Power Benefits in 3D ICs: Block Folding and Bonding Styles Perspective," ACM Design Automation Conference, 2014.
[8] **Yarui Peng**, Dusan Petranovic, and Sung Kyu Lim, "Fast and Accurate Full-chip Extraction and Optimization of TSV-to-Wire Coupling," ACM Design Automation Conference, 2014.
[9] **Yarui Peng**, Bon Woong Ku, Younsik Park, Kwang-Il Park, Seong-Jin Jang, Joo Sun Choi, and Sung Kyu Lim, "Design, Packaging, and Architectural Policy Co-Optimization for DC Power Integrity in 3D DRAM," ACM Design Automation Conference, 2015.
[10] **Yarui Peng**, Moongon Jung, Taigon Song, Yang Wan, and Sung Kyu Lim, "Thermal Impact Study of Block Folding and Face-to-Face Bonding in 3D IC", IEEE International Interconnect Technology Conference, 2015
[11] Taigon Song, Moongon Jung, Yang Wan, **Yarui Peng**, and Sung Kyu Lim, "3D IC Power Benefit Study Under Practical Design Considerations", IEEE International Interconnect Technology Conference, 2015.
[12] **Yarui Peng**, Taigon Song, Dusan Petranovic, and Sung Kyu Lim, "Full-chip Inter-die Parasitic Extraction in Face-to-Face-Bonded 3D ICs," IEEE International Conference on Computer-Aided Design, 2015
[13] Can Rao, **Yarui Peng**, Tongqing Wang, Sung Kyu Lim, and Xinchun Lu, ''Investigation of Post-annealing Stress and Pop-out in TSV Front-side CMP", International conference on planarization/CMP technology, 2016, ***Best student paper award***

Thank You