Micro-bumping vs. Hybrid Bonding: Pros/Cons in Terms of Power Delivery

Prof. Sung Kyu Lim

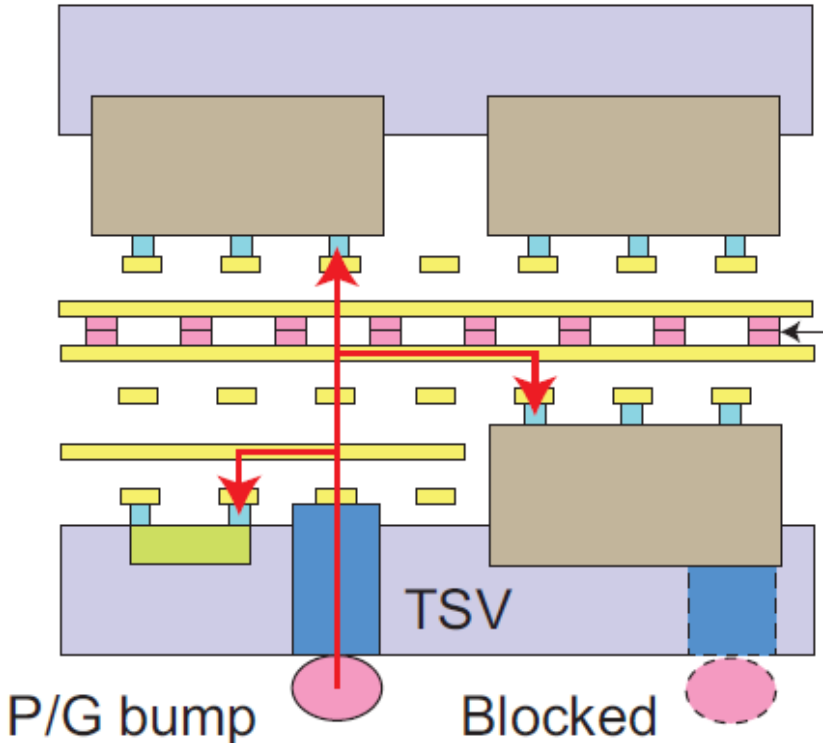
Georgia Institute of Technology

Back-End Design Session, 7/10/2023

Outline

- Motivation
- Executive Summary
- Design and Simulation Setting
- Power Integrity Analysis Details

Power Delivery Challenges in 3D ICs

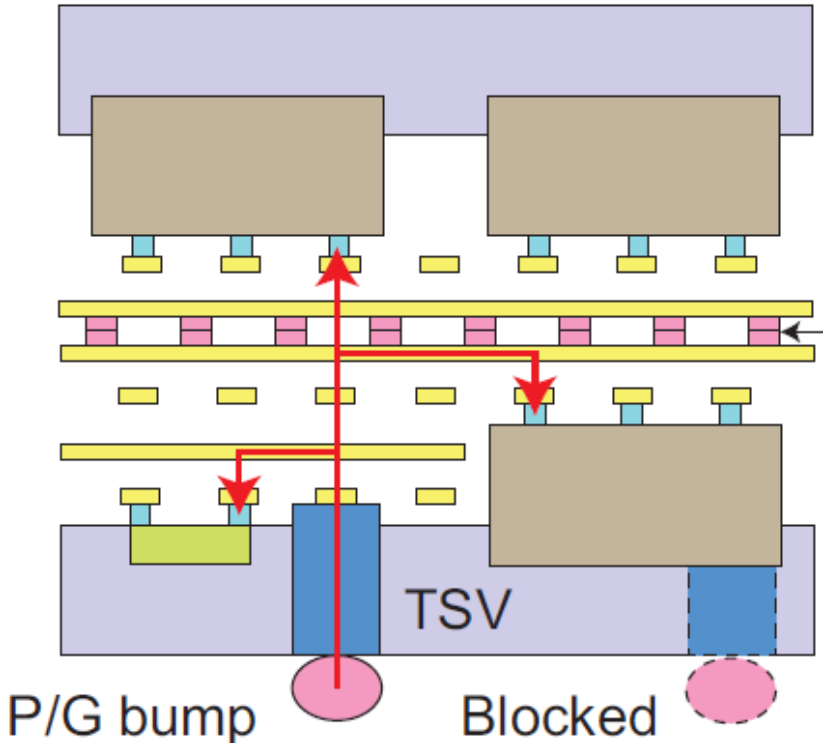


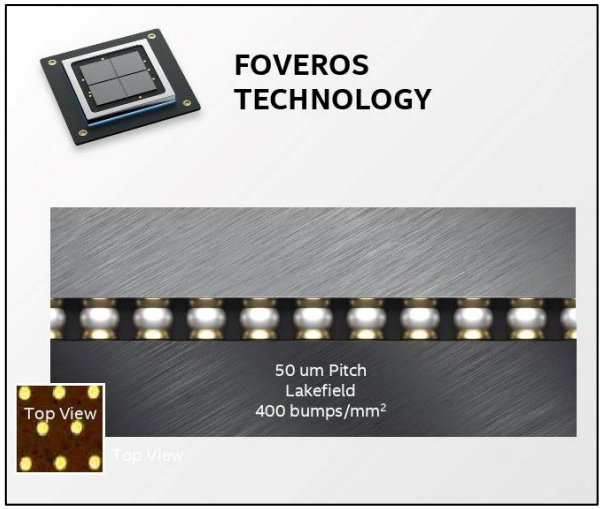
Higher power density

Fewer power bumps/pads

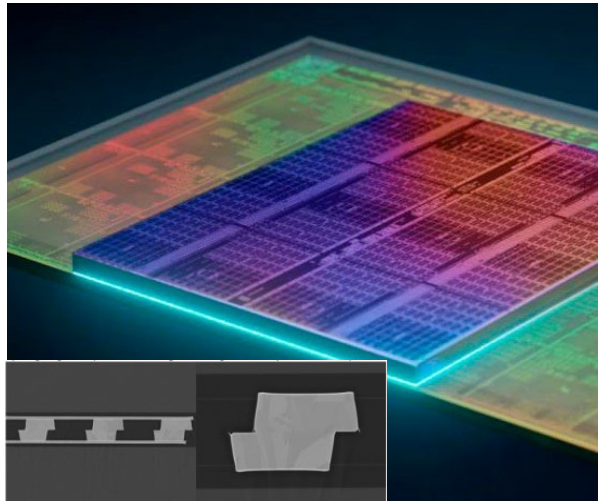
More resistive power paths

Die Bonding: Two Leading Contenders





Micro-Bumping, Intel
50um pitch demonstrated



Hybrid Bonding, TSMC
9um pitch demonstrated

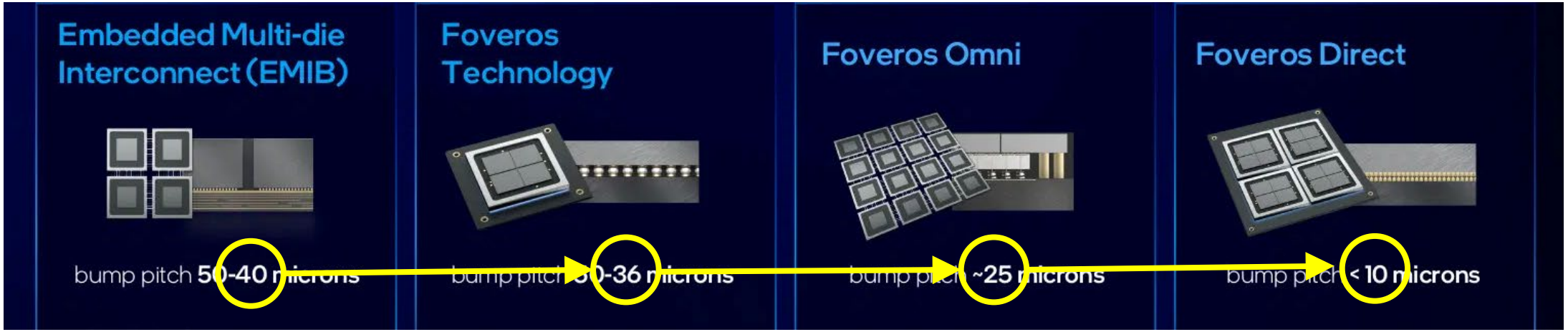
3D Bonding Roadmap

Numbers from public domain

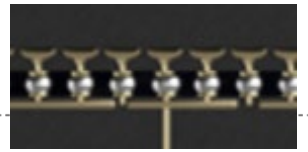
TSMC SoIC Scaling Roadmap					
Generation	SoIC1	SoIC2	SoIC3	SoIC4	SoIC5
Bond Pitch	9-micron	6-micron	4.5-micron	3-micron	2-micron
Bandwidth Density[(1)]	1x	2x	3.4x	6.05x	11.54x
Energy Efficiency	1x	1x	1.17x	1.33x	1.44x

OK, bond pitch will reduce. Great.

What is the impact on power delivery?



Our Choices for Pitch



micro-bumping

present

25um

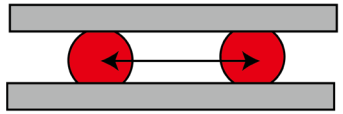
10mm

10mm

160,000
bumps

R: 40 mO
C: 34.2 fF

future



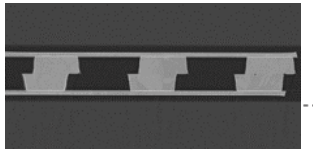
10um

10mm

10mm

1,000,000
bumps

R: 99 mO
C: 3.3 fF



hybrid bonding

present

5um

10mm

10mm

4,000,000
pad

R: 18 mO
C: 0.1 fF

future

1um

10mm

10mm

100,000,000
pad

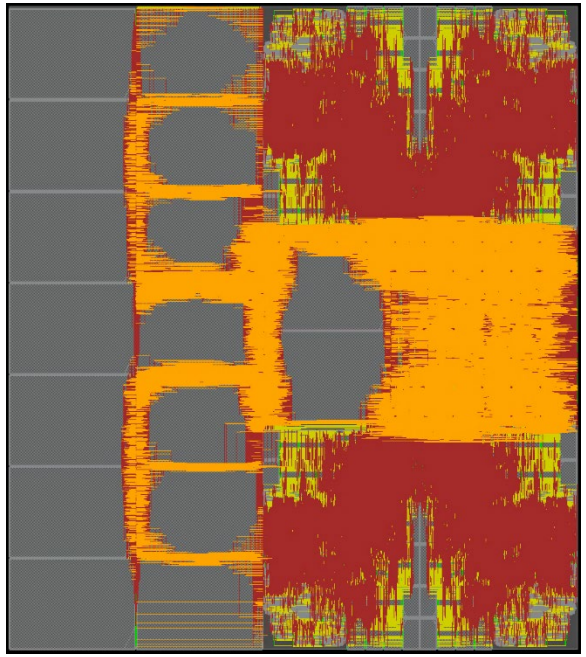
R: 100 mO
C: 0.07 fF

Executive Summary

Sanity check: this is our results, not someone else's.

We Do Full-Chip Designs

Architecture	Technology	Quality	Design Flow	3D Config
ARM A53 Quad Core	16nm	GDS + Sign-off	Macro-3D	L2-on-Logic

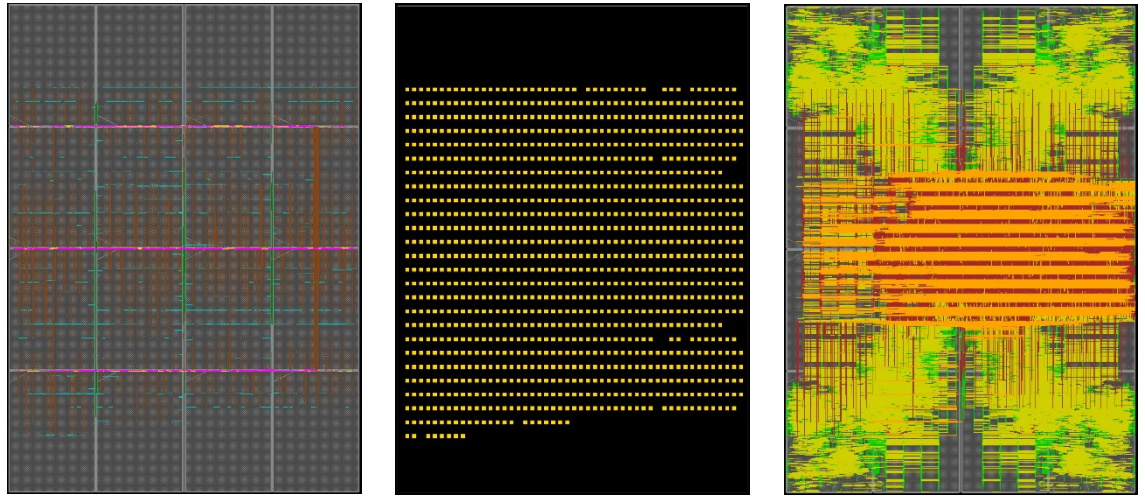


2D IC GDS (6 metal)

memoryc

signal bump (1,188)

logic



3D IC GDS (7M + 6L metals)

Sanity check: this is our results, not someone else's.

2D vs. 3D Iso-Performance Comparison

- Iso-performance (low power apps)
  - Target freq. is fixed to 0.6GHz for all designs

← tier partitioning is identical →

NPPAT	Innovus 2D	U-bump 25um	U-bump 10um	H-bond 5um	H-bond 1um
Static IR-drop (mV)	12.78	41.48	42.91	27.54	24.43
Power (mW)	537.3	536.4	537.1	466.3	450.6
Fmax (GHz)	0.600	0.600	0.600	0.600	0.600
Footprint (mm)	2.0 x 2.2	1.2 x 1.7	1.2 x 1.7	1.2 x 1.7	1.2 x 1.7
Tmax (°C)	51.60	59.52	59.54	57.63	57.20

Sanity check: this is our results, not someone else's.

2D vs. 3D Max-Performance Comparison

- Max-performance (high performance apps)
  - Fmax is pushed to the upper limit for each design

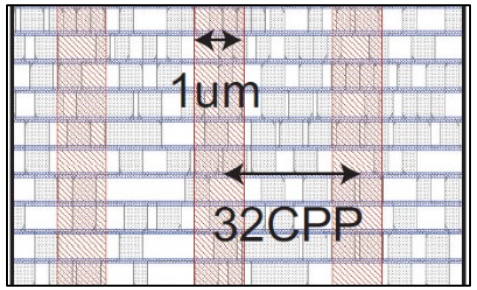
← tier partitioning is identical →

NPPAT

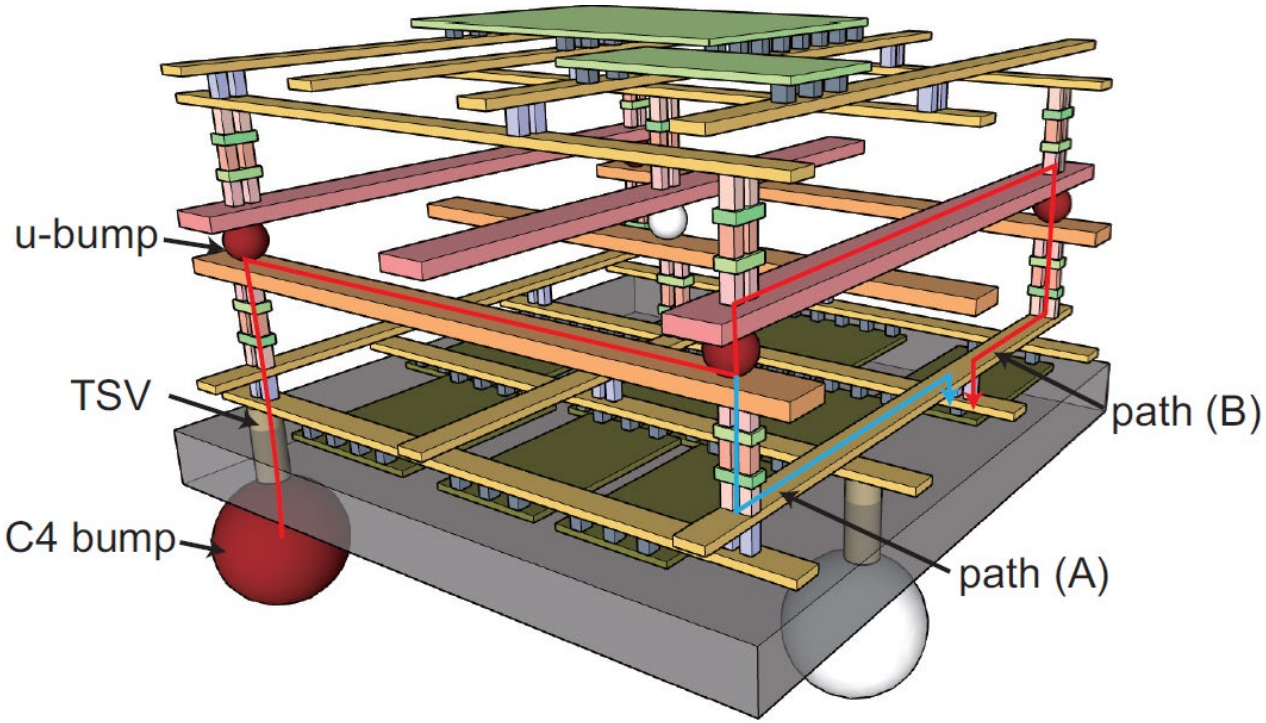
	Innovus 2D	U-bump 25um	U-bump 10um	H-bond 5um	H-bond 1um
Static IR-drop (mV)	14.25	46.63	68.10	74.22	68.23
Power (mW)	598.6	651.1	888.9	1267.1	1214.5
Fmax (GHz)	0.630	0.694	0.977	1.172	1.218
Footprint (mm)	2.0 x 2.2	1.2 x 1.7	1.2 x 1.7	1.2 x 1.7	1.2 x 1.7
Tmax (°C)	52.34	62.63	69.07	79.31	77.89

Design and Simulation Setting

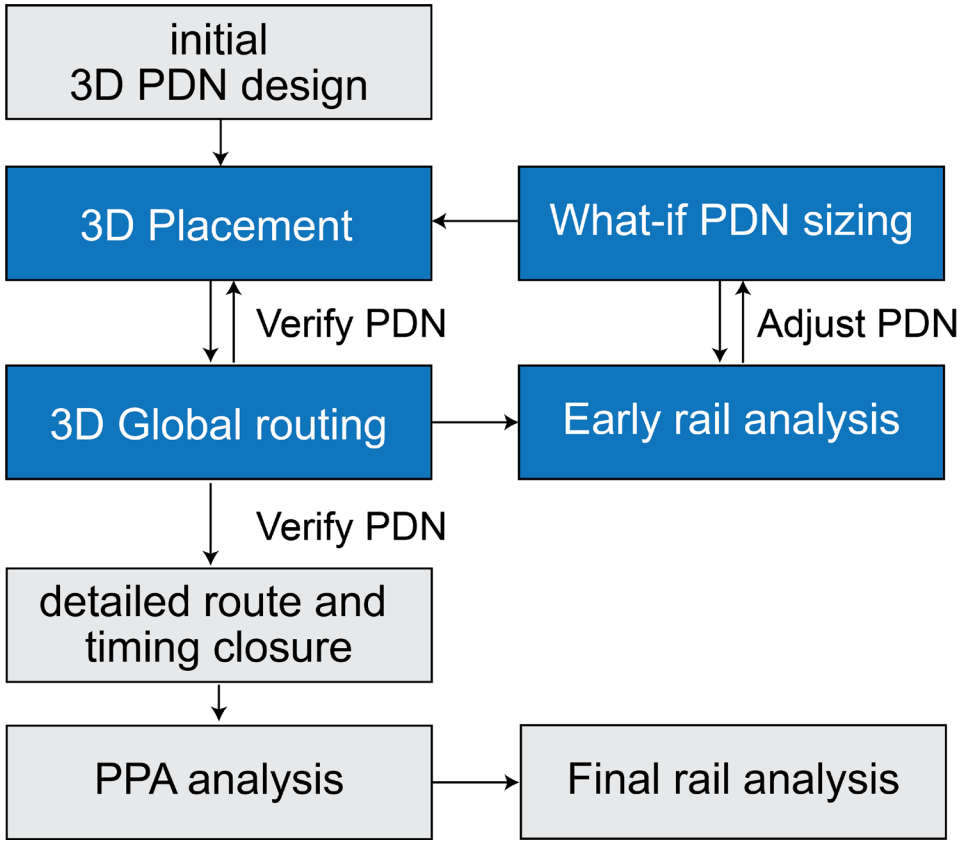
Assembly Design Kit (ADK)

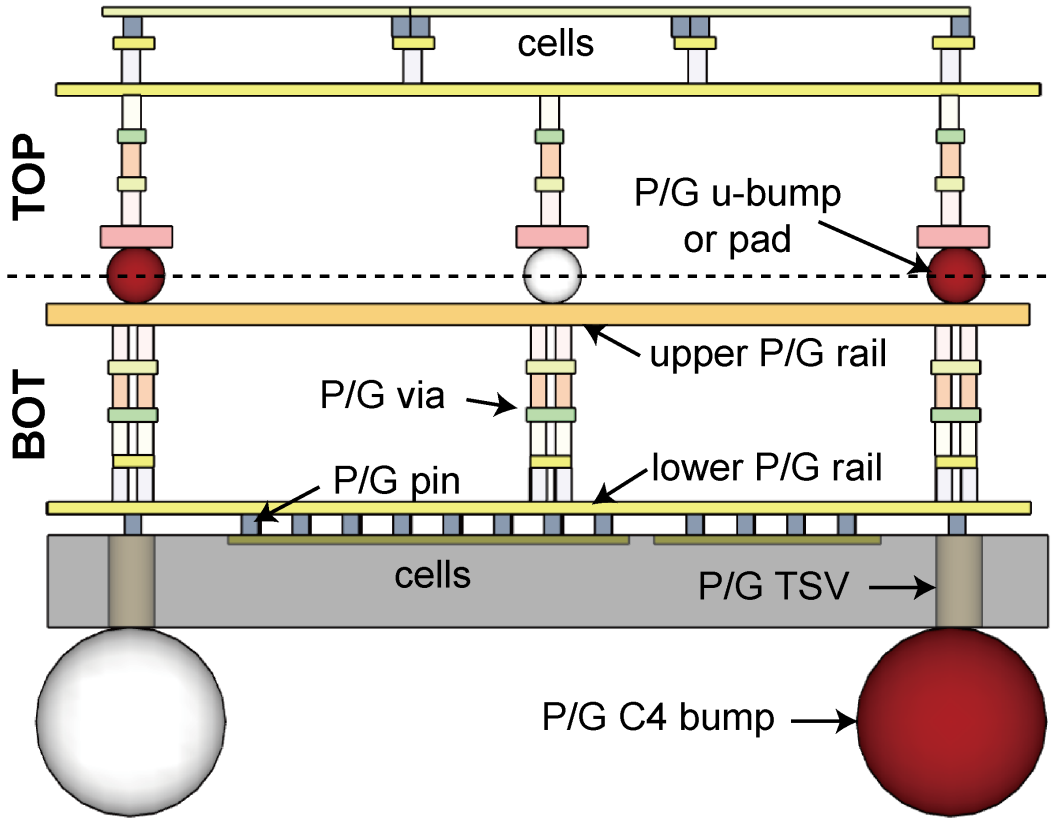


3D Technology Setup	
Metal layer #	7M
Vdd	800mV
PDN layer	M1, M5, M6, M7
Rail pitch / width (um)	2.8 / 1
C4 pitch (um)	100
C4 RC (mΩ / fF)	18 / 38
TSV pitch (um)	10
TSV RCL (mΩ / fF / pH)	47 / 86 / 48



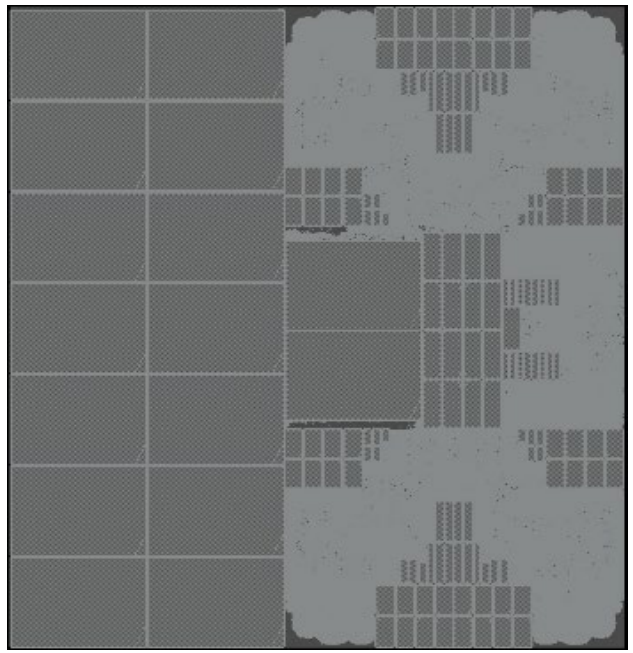
PDN Design Flow



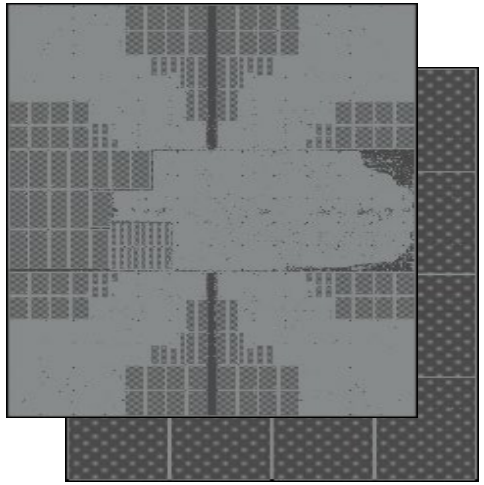


Full-chip 3D Design Flow

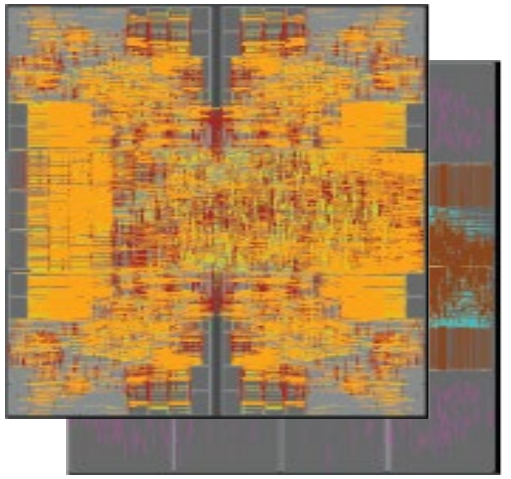
- Leverage commercial 2D IC tools (as much as possible)
- Then add missing pieces



2D IC placement
[existing 2D IC tool]



Gate-level tier partitioning
[innovation]

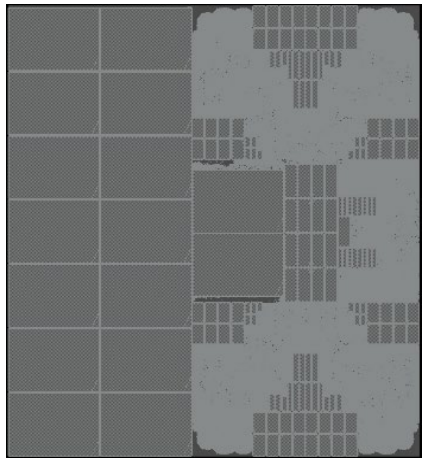
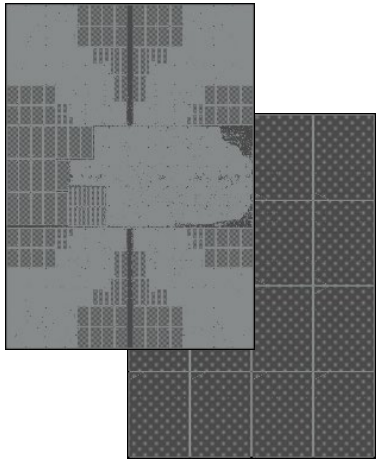
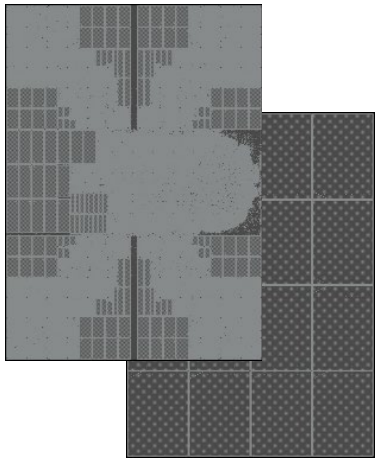
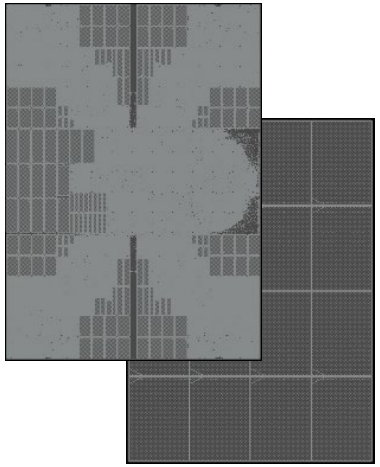
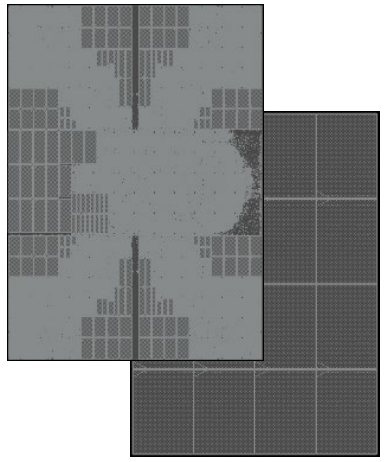


3D IC routing and timing closure
[innovation]

Power Integrity Analysis Details

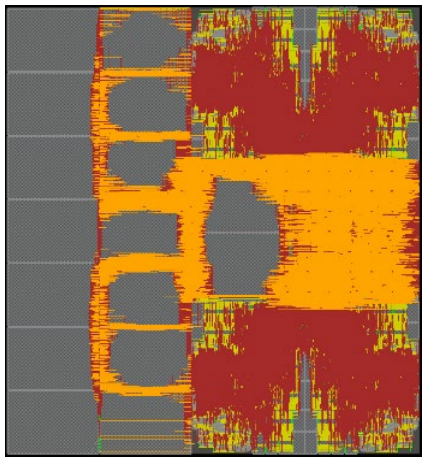
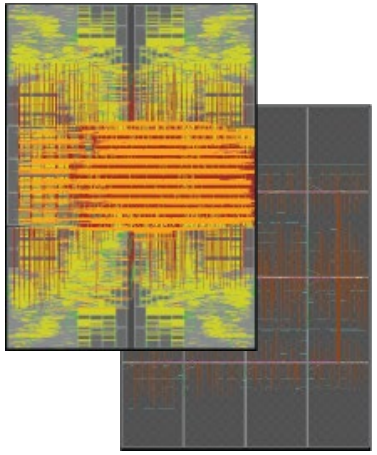
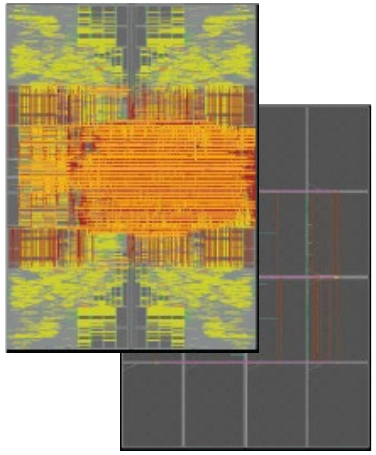
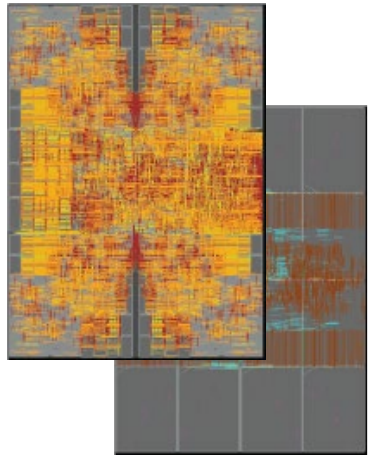
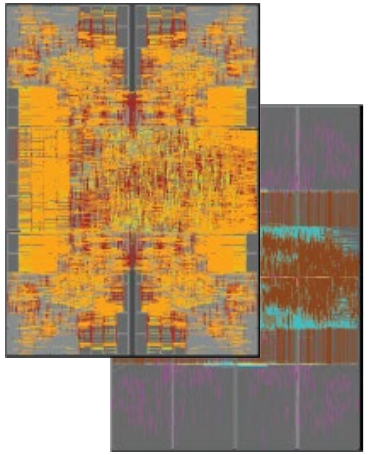
Five Full-Chip GDS Layouts

- Placement
  - Identical tier partitioning, but different placement

				
2D	U-bump 25um	U-bump 10um	H-bond 5um	H-bond 1um

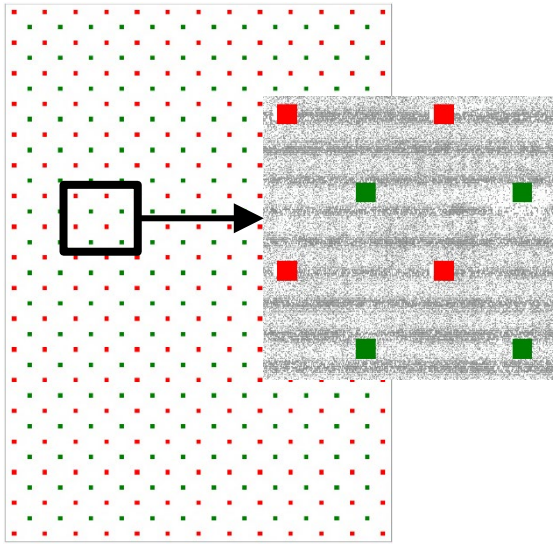
Five Full-Chip GDS Layouts

- Routing
  - Identical tier partitioning, but different routing

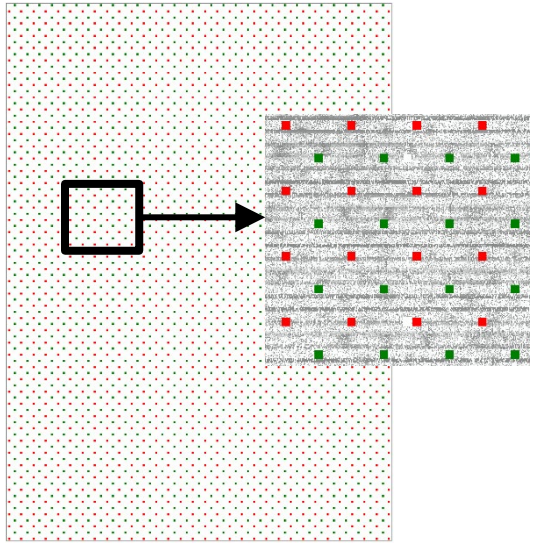
				
2D	U-bump 25um	U-bump 10um	H-bond 5um	H-bond 1um

P/G Bump/Pad Placement

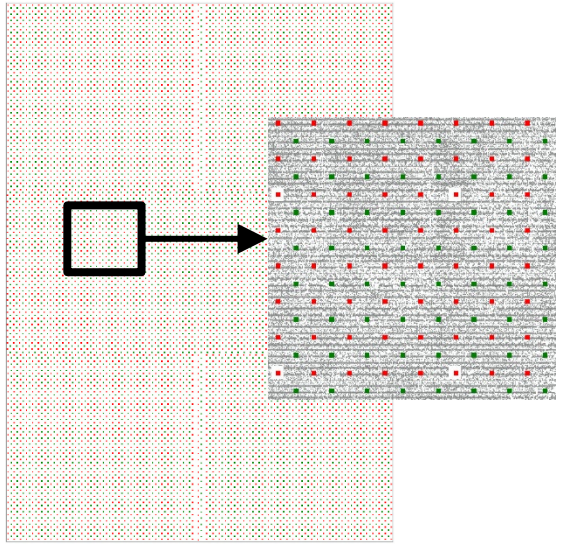
Red: power, Green: ground



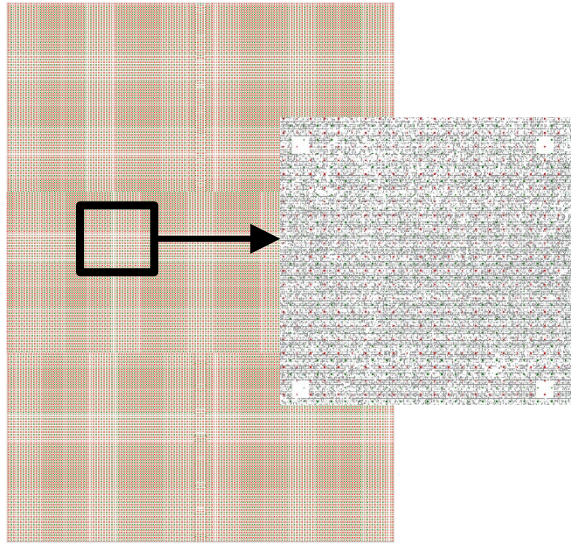
U-bump 25um



U-bump 10um



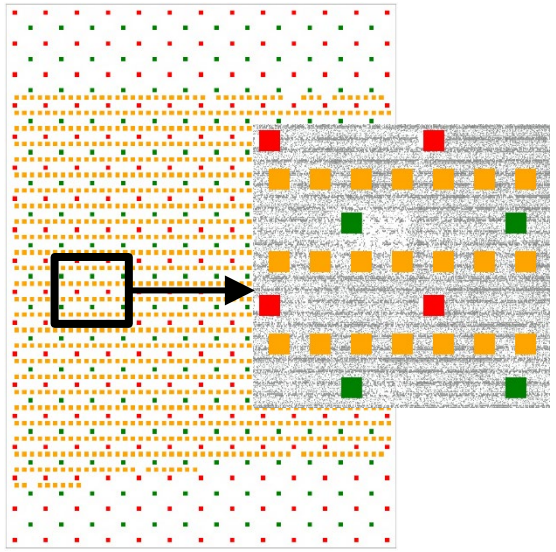
H-bond 5um



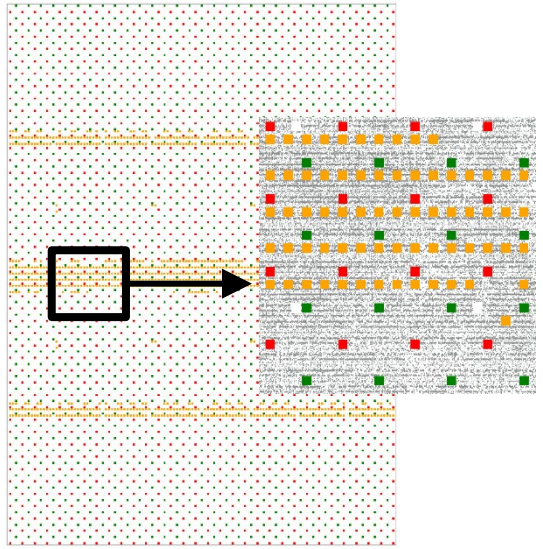
H-bond 1um

P/G/S Bump/Pad Placement

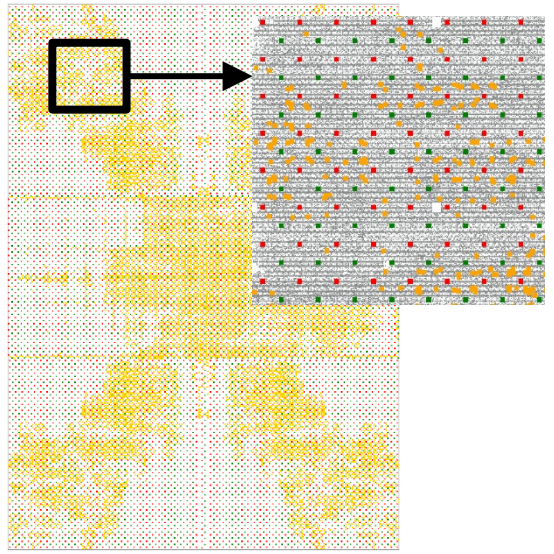
Red: power, Green: ground, Yellow: signal



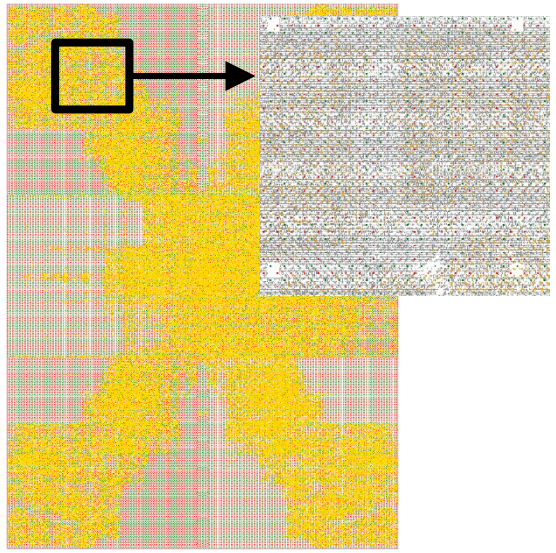
U-bump 25um



U-bump 10um



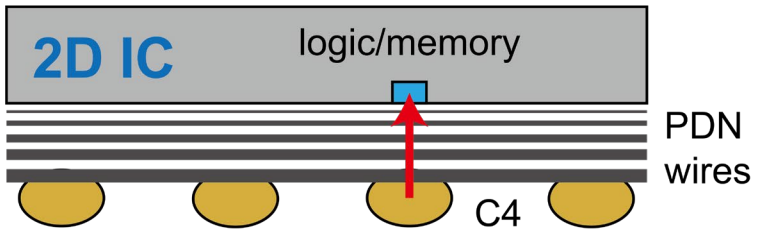
H-bond 5um

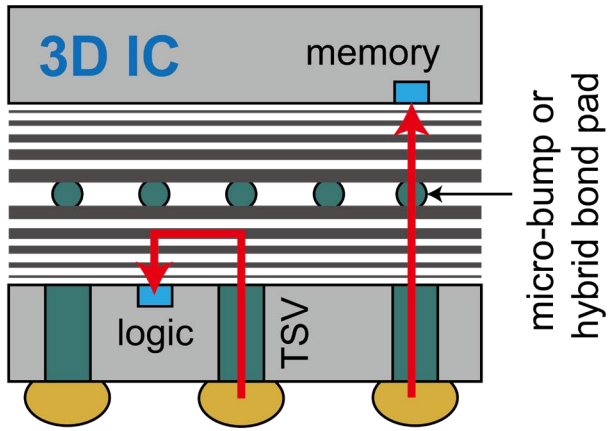


H-bond 1um

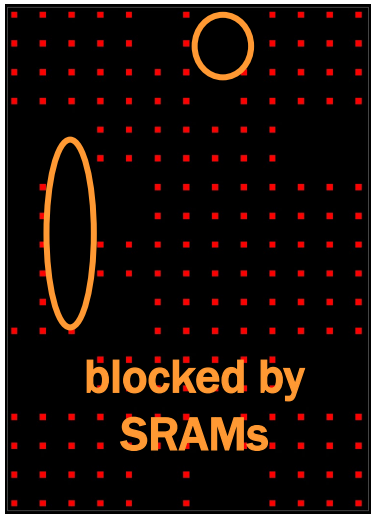
PDN Usage Summary

	Innovus 2D	U-bump 25um	U-bump 10um	H-bond 5um	H-bond 1um
# P/G C4	460	234	234	234	234
PDN wire area (mm^2)	3.26	3.33	3.34	3.33	3.34
# P/G bump/pad used	-	438	3K	11K	139K

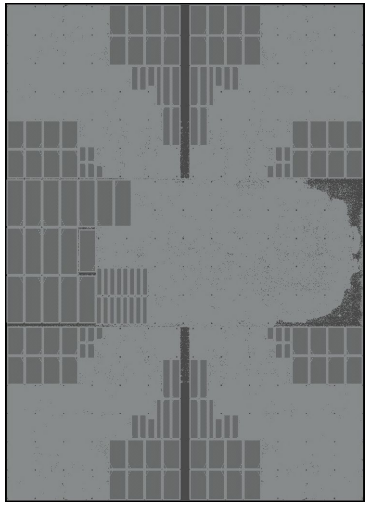




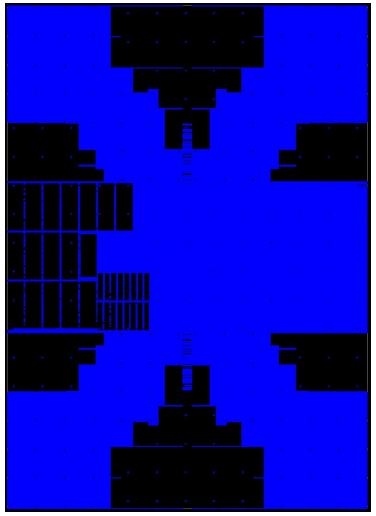
PDN Routing: Logic Tier



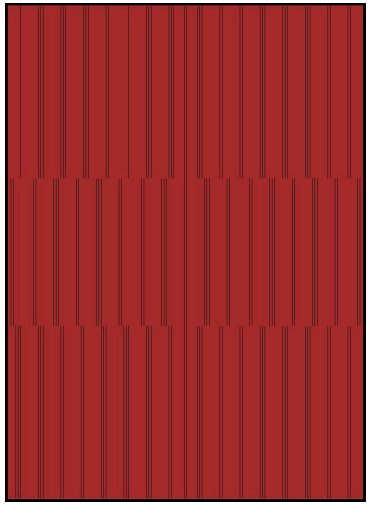
C4 (pitch = 100um)



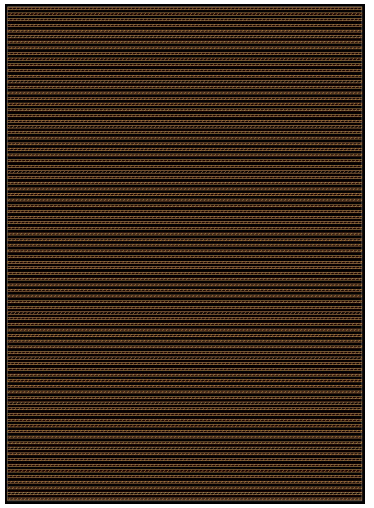
placement



M1 P/G (31.7%)

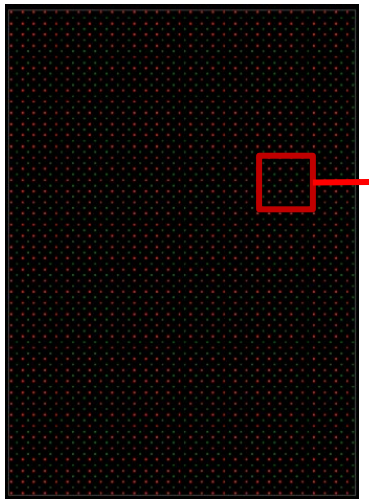


M5 P/G (70.0%)

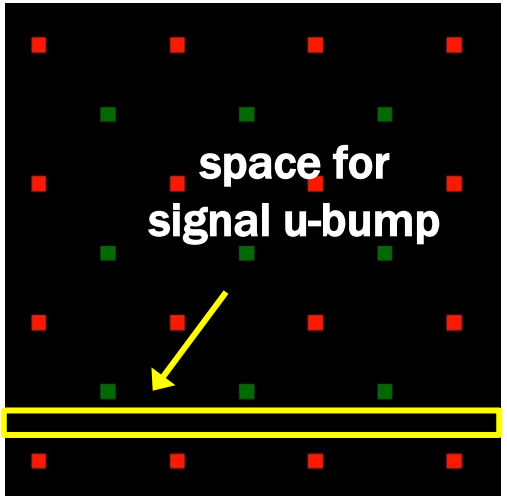


M6 P/G (71.3%)

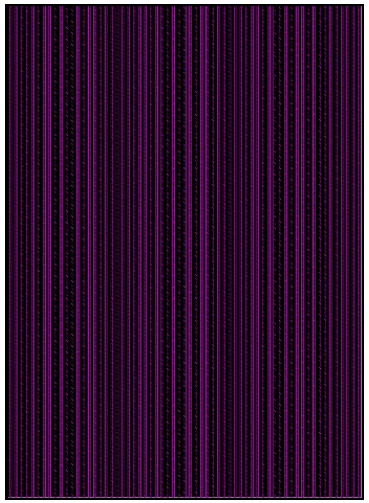
PDN Routing: Memory Tier



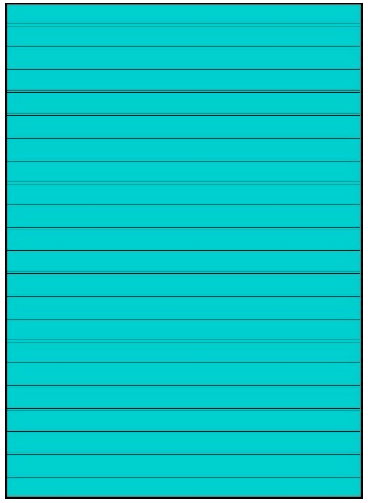
P/G u-bump



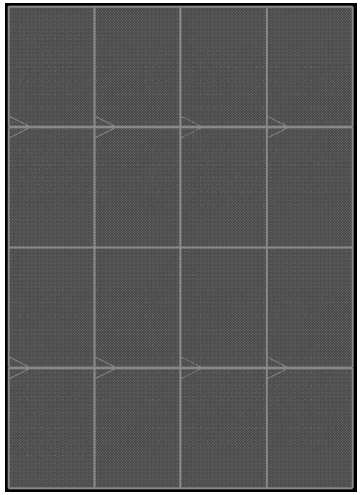
red = P, green = G



M7 P/G (71.1%)



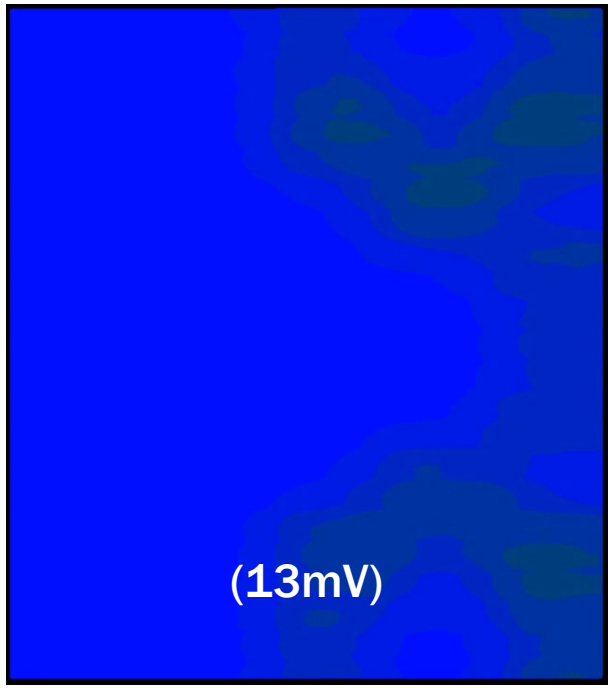
M6 P/G (71.0%)



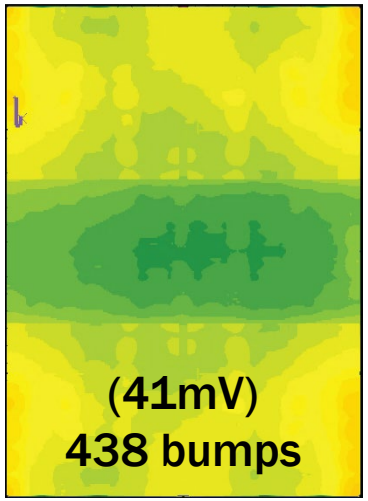
placement

Iso-Performance Design Analysis

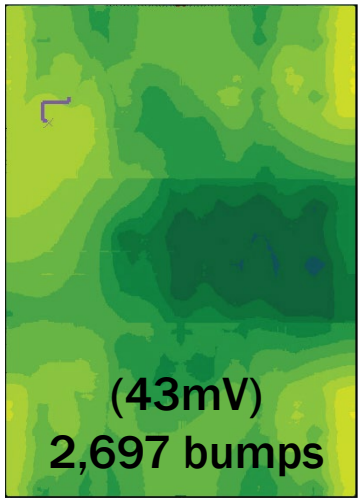
Static IR-drop Maps



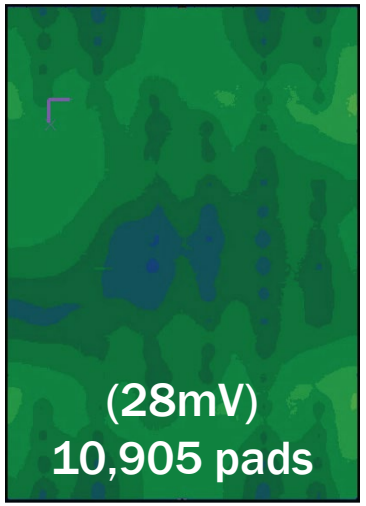
2D



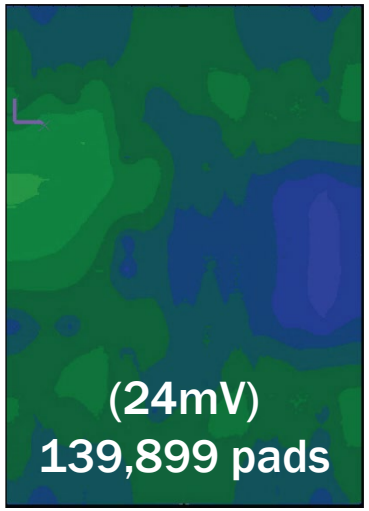
U-bump 25um



U-bump 10um

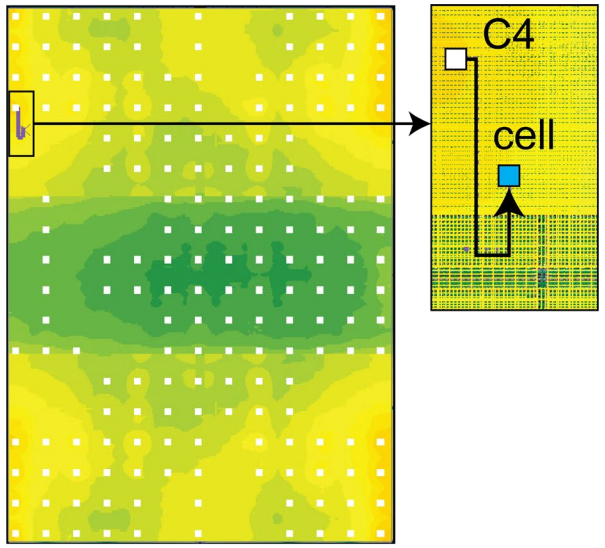


H-bond 5um

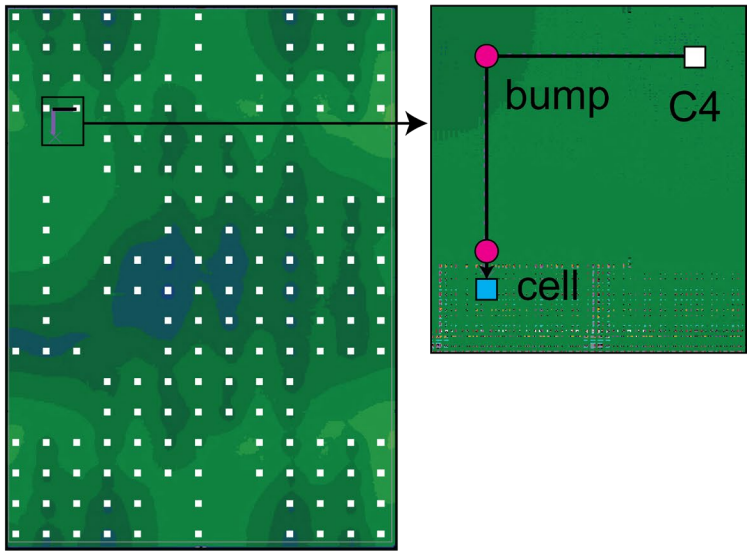


H-bond 1um

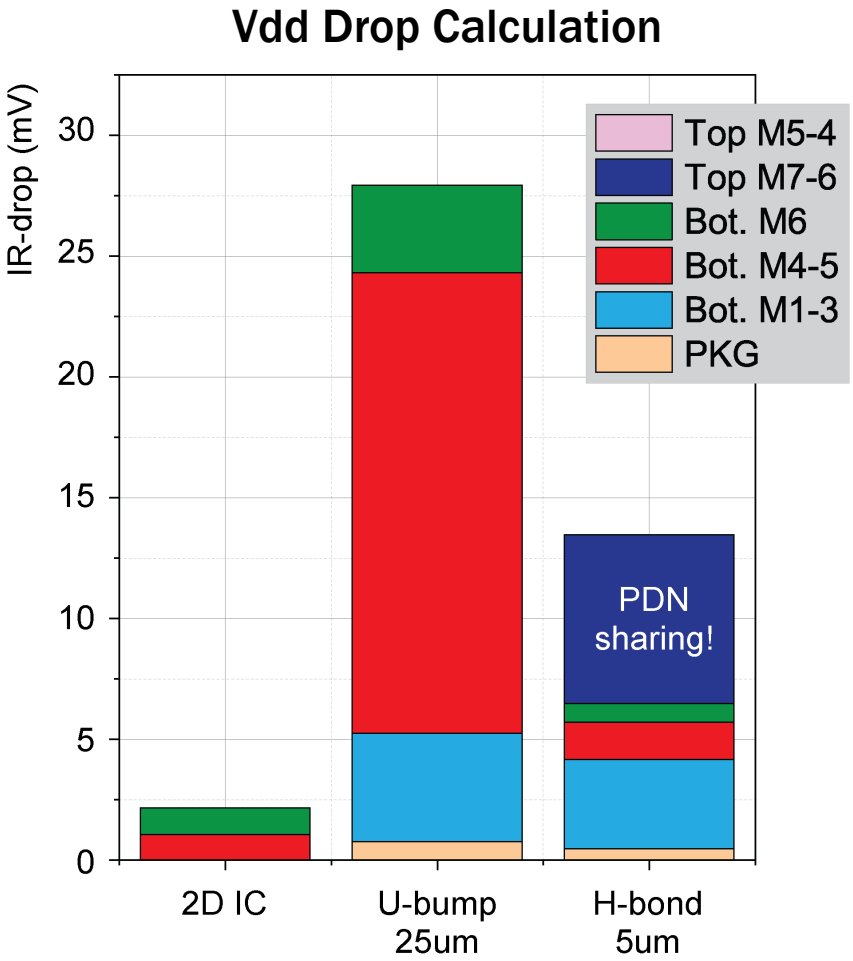
Least Resistive Path



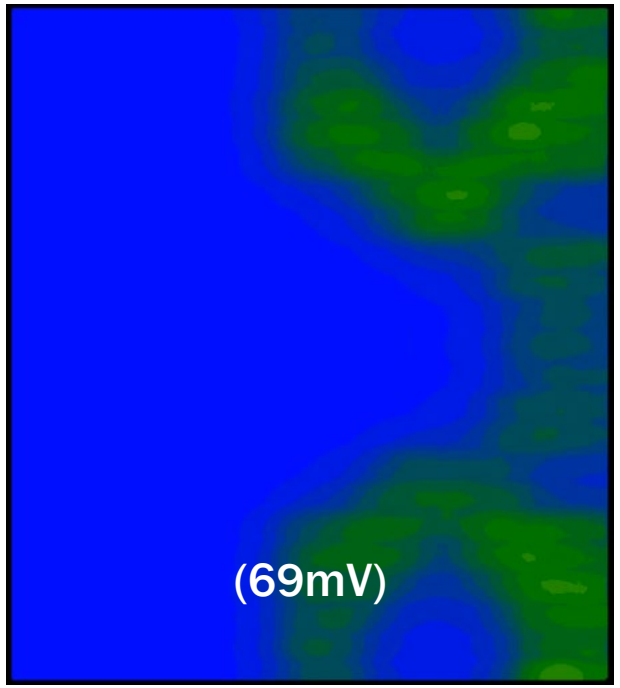
U-bump 25um



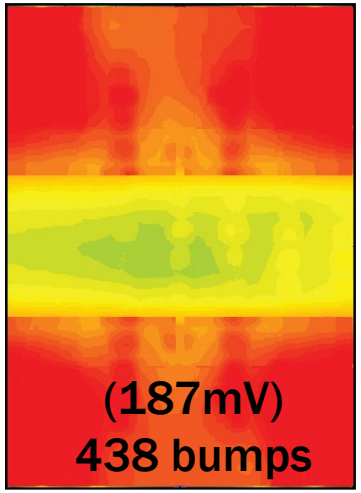
H-bond 5um



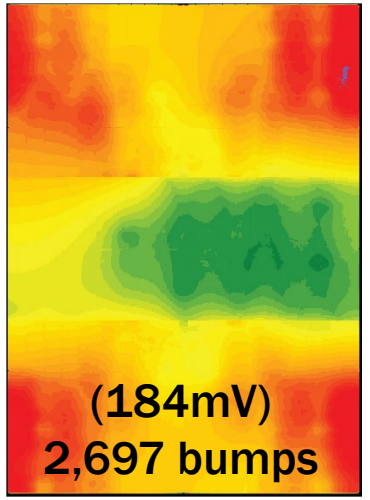
Dynamic IR-drop Maps



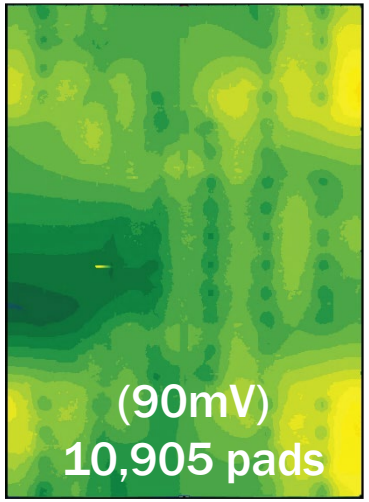
2D



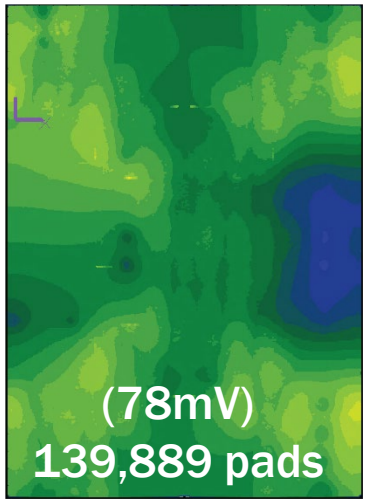
U-bump 25um



U-bump 10um

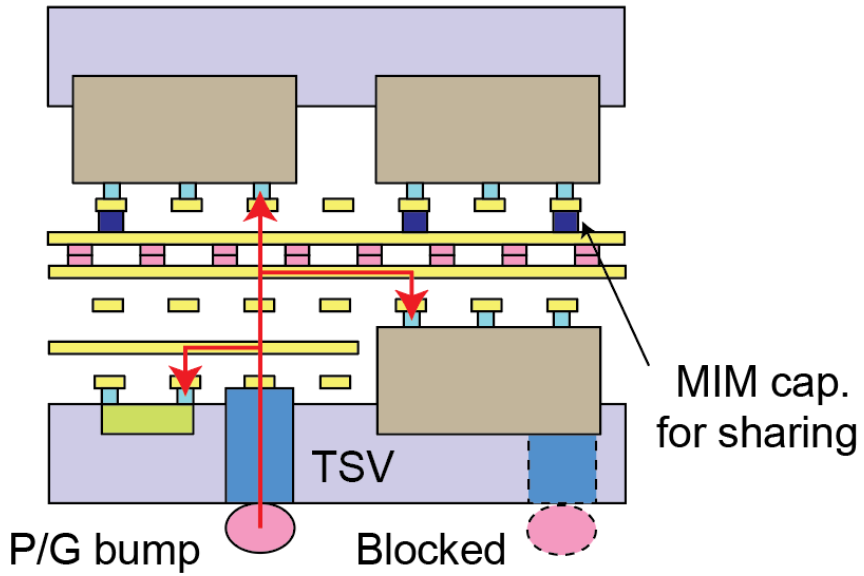


H-bond 5um

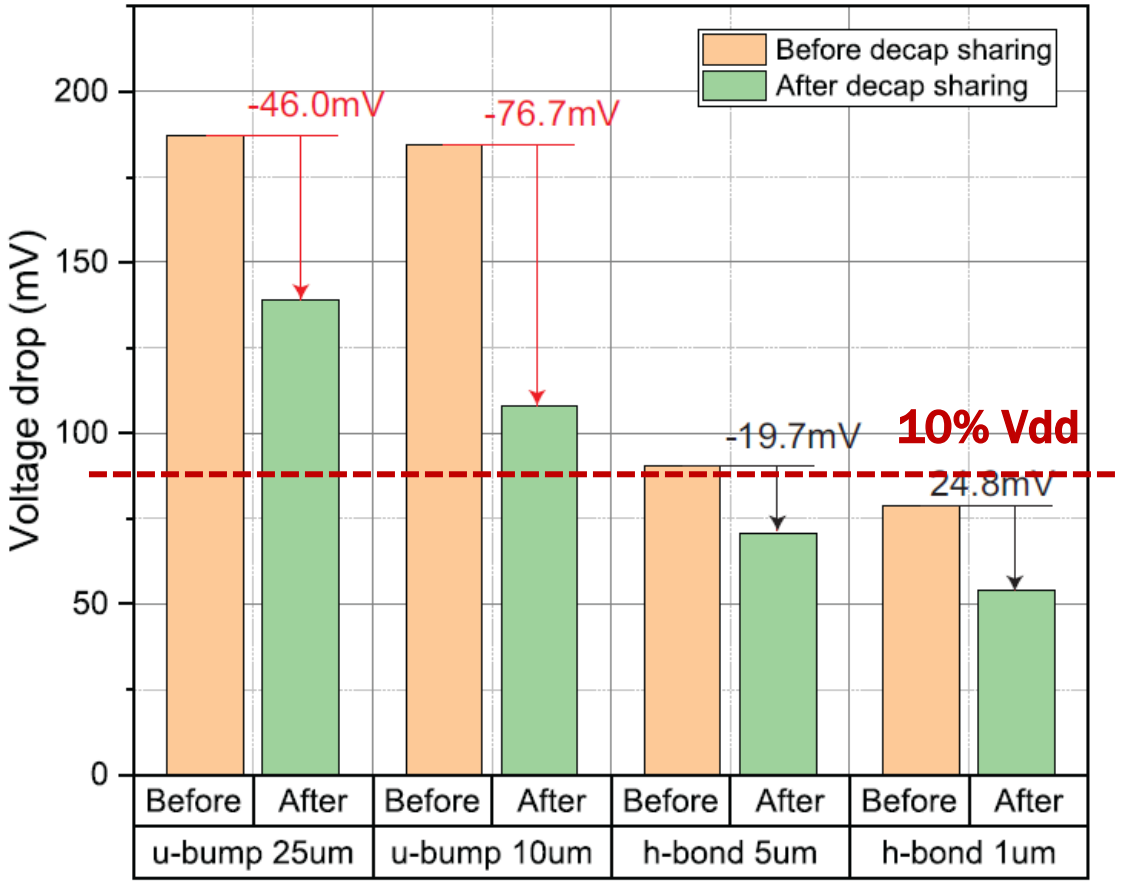


H-bond 1um

MIM Decaps Help

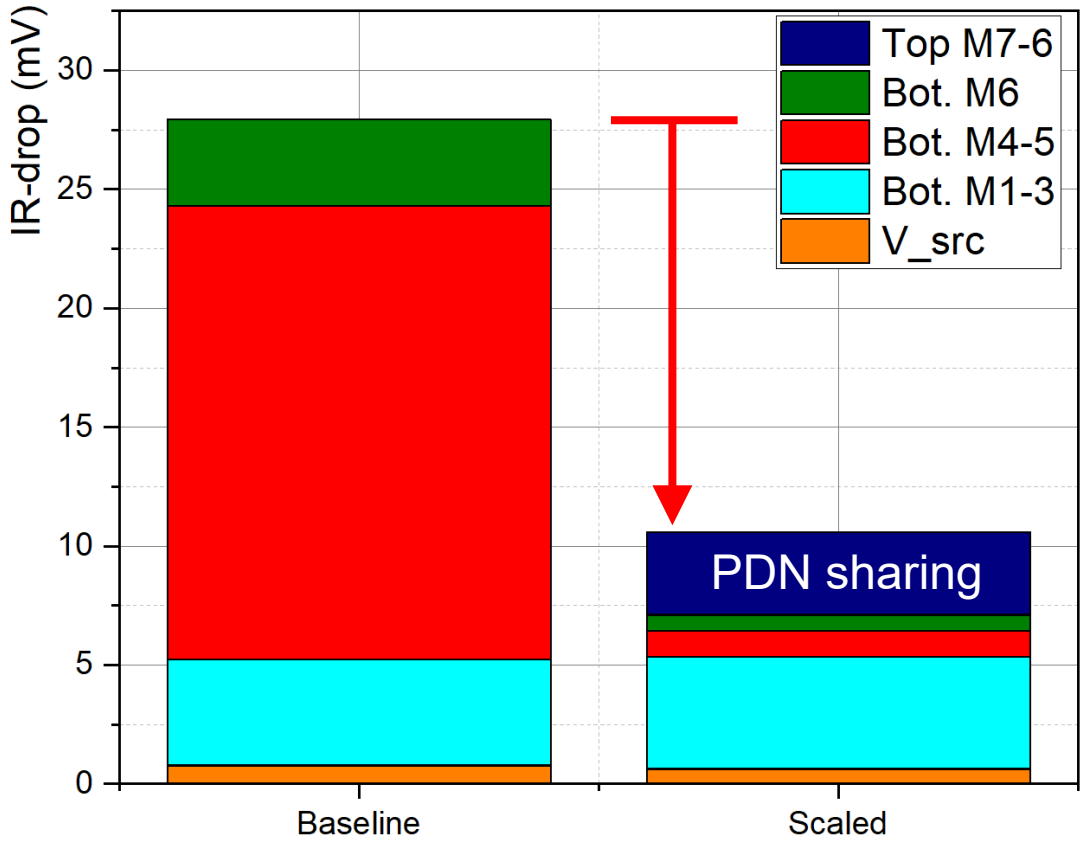


In logic tier BEOL (M5-M7)
Cap. density = 40fF/mm^2
Total inserted: 678nF



What If M45 Resistivity Reduced?

- Scale wire resistivity
  - 50% on M4 & M5 (U-bump 25um)
  - Vdd drop reduces from 27mV to 11mV
  - More inter-tier rail sharing occurs



Conclusion

- Power delivery is harder for 3D ICs
  - Mainly due to smaller footprint (= fewer C4s)
  - Micro-bumping vs. hybrid bonding: it depends on several factors
- 3D IC IR-drop under iso-performance
  - Micro-bumping was worse: due to fewer bumps
- 3D IC IR-drop under max-performance
  - Hybrid bonding was worse: due to faster chip