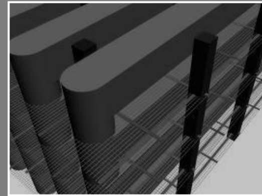
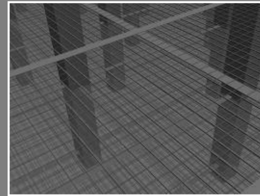
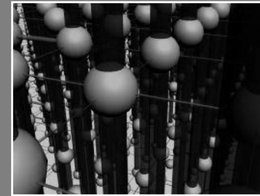
High-Density Integration of Functional Modules Using Monolithic 3D-IC Technology

Shreepad Panth[1], Kambiz Samadi[2], Yang Du[2] and Sung Kyu Lim[1]

[1]Dept. of Electrical and Computer Engineering, Georgia Tech, Atlanta GA, USA
[2]Qualcomm Research, San Diego, CA, USA

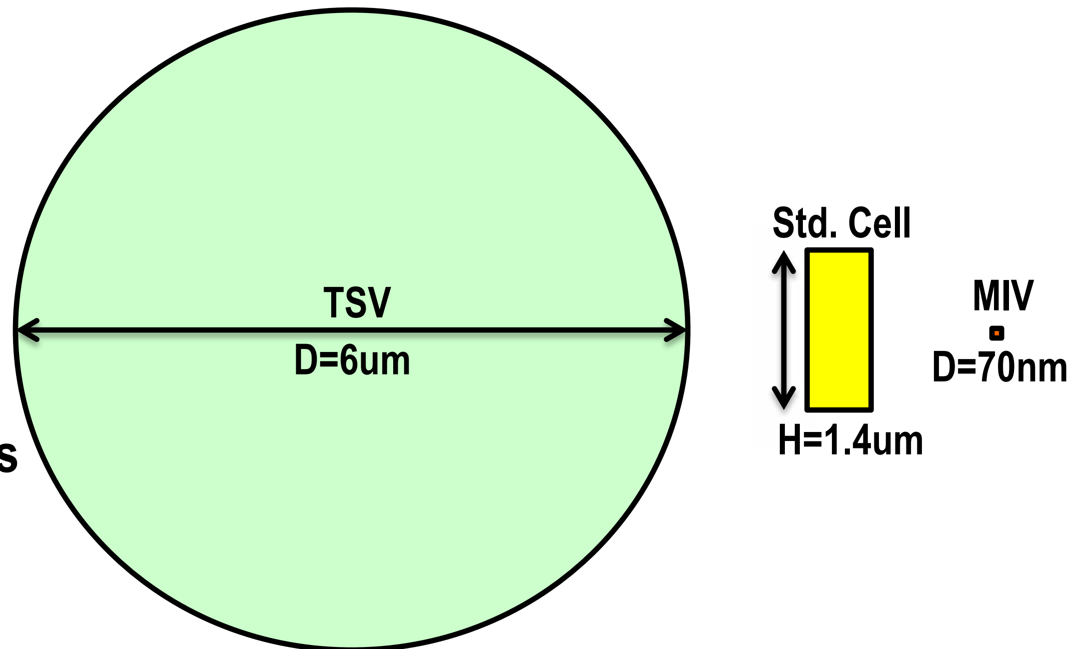
Outline

- Introduction to monolithic 3D
- Prior work and contributions
- Design and analysis methodology
- Post-layout comparisons with 2D and TSV-based 3D
  - Footprint, Wirelength, Timing and Power
- Conclusions

Extending Moore's Law – 3D-ICs

- 3D-ICs have emerged as a promising solution to continue scaling
- Many possible solution – SiP, PoP, Through Silicon Via , etc.

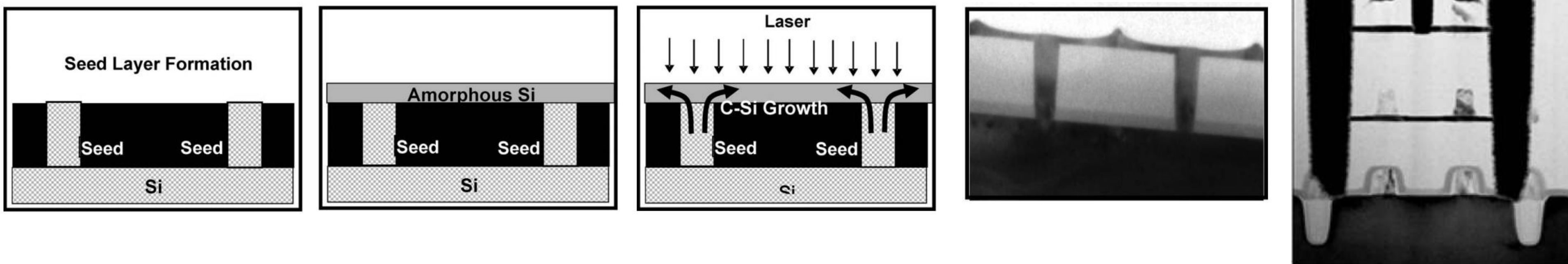
- TSV-based 3D
  - Dies fabricated separately
  - Wafer thinned
  - Aligned and bonded
  - Pitch is limited by microbumps and alignment accuracy



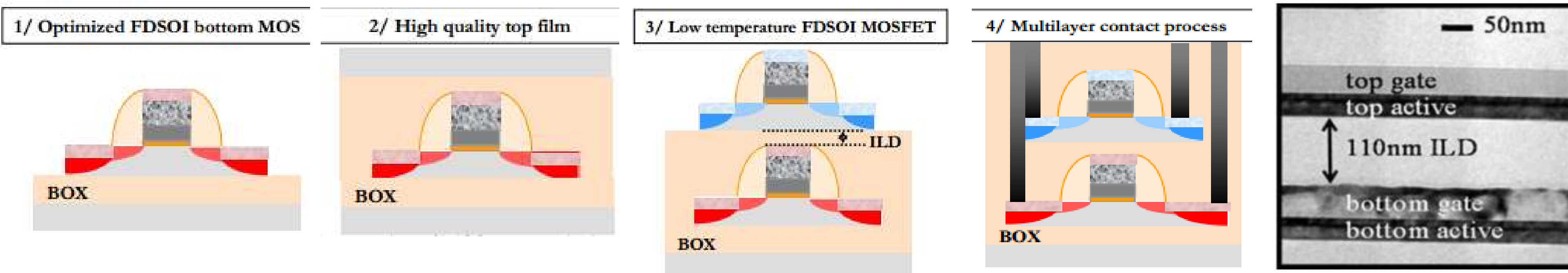
- Monolithic 3D is emerging as an alternative
  - Tiers fabricated sequentially → no alignment issues
  - Monolithic inter-tier vias (MIVs) are the same size as local vias

Monolithic 3D – Fabrication Methods

- **Epitaxial growth[2]**



- **Wafer bonding[3]**



[2] S.-M. Jung, H. Lim, K. Kwak, and K. Kim, " 500-MHz DDR High-Performance 72-Mb 3-D SRAM …" in IEEE Trans. on Electron Devices, 2010.
[3] P. Batude et al., "Advances in 3D CMOS Sequential Integration," in Proc. IEEE Int. Electron Devices Meeting, 2009.

Prior Work

- Transistor-level monolithic 3D[4][5][6]
  - Separate PMOS and NMOS onto separate tiers (limited to two tiers)
- Gate-level monolithic 3D[4][7]
  - Each std. cell has PMOS and NMOS on the same tier, and std. cells are distributed onto multiple tiers
- No existing work on block-level monolithic
  - Due to the extensive use of IP blocks, this design style is likely to be the first to utilize this technology

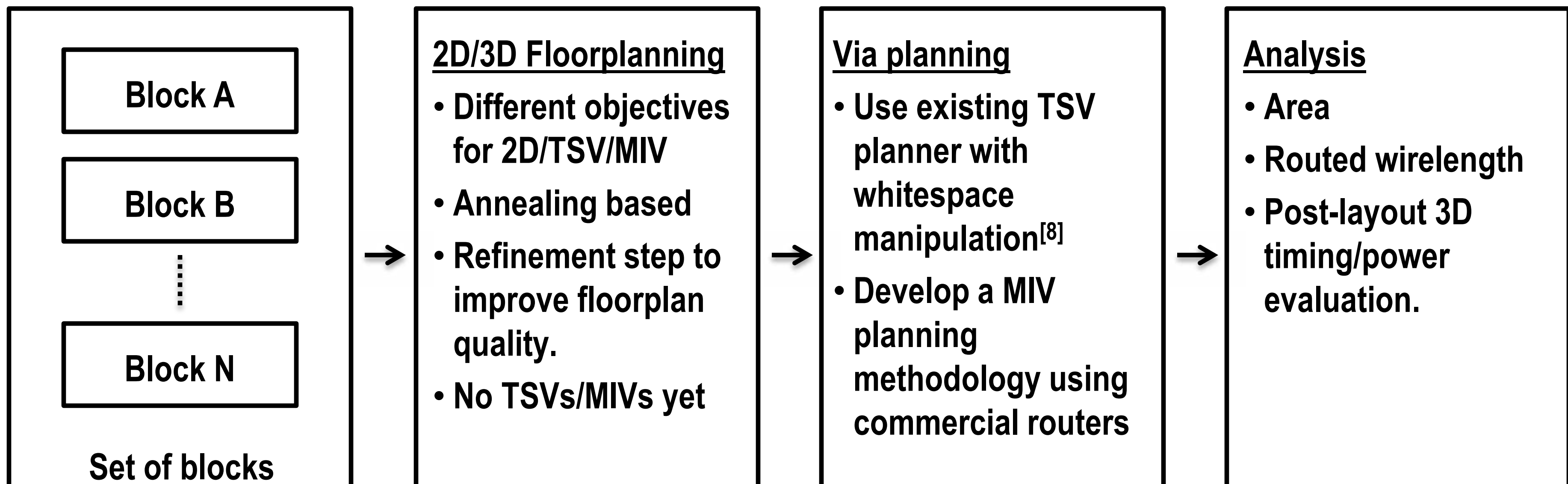
[4] S. Bobba, et al., “CELONCEL: Effective design technique for 3-D monolithic integration targeting high perf. integrated circuits,” in ASPDAC, 2011.
[5] C. Liu and S. K. Lim, “Ultra-High Density 3D SRAM Cell Designs for Monolithic 3D Integration,” in IITC 2012.
[6] Y. J. Lee, P. Morrow, and S. K. Lim, “Ultra High Density Logic Designs Using Transistor-Level Monolithic 3D Integration,” in ICCAD, 2012.
[7] C. Liu and S. K. Lim, “A Design Tradeoff Study with Monolithic 3D Integration,” in ISQED, 2012.

Contributions

- First work to consider block-level monolithic 3D
- We develop a RTL → GDSII methodology for block-level monolithic 3D-ICs
  - We develop a floorplanning framework for monolithic 3D-ICs
  - We develop a MIV planning methodology
  - We perform post-layout analysis on block-level monolithic 3D-ICs
- We show that monolithic 3D-ICs have huge benefits over 2D-ICs
  - Negligible total silicon area penalty (Max 2%)
  - Up to 42% reduction in the inter-block WL
  - Up to 33% reduction in the longest path delay
  - Up to 82% reduction in the total negative slack
  - Up to 43% reduction in the inter-block net power

Design Flow

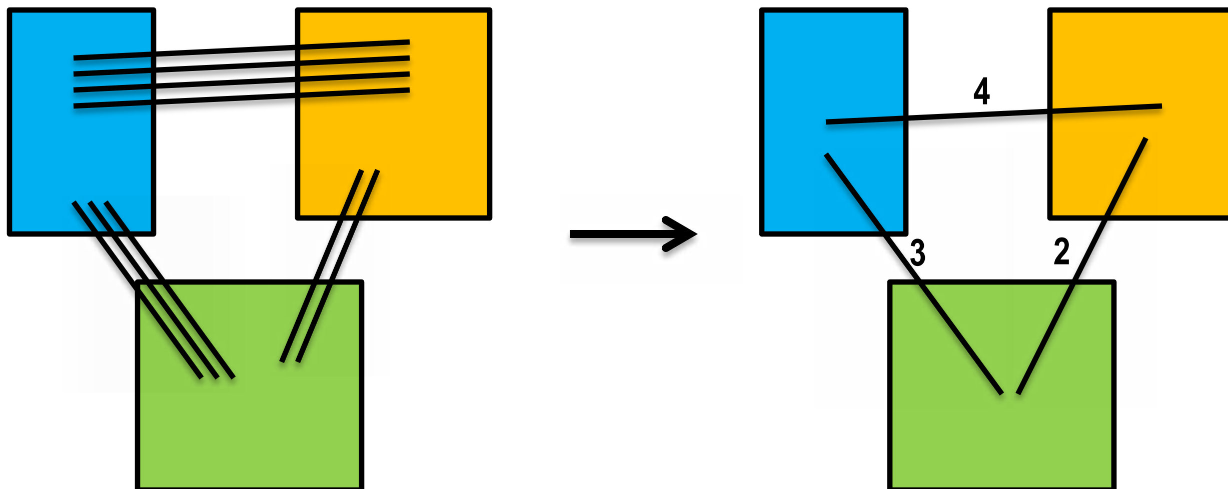
- Given a set of hard-blocks (fixed GDSII)
- Evaluate the benefits of implementation in monolithic 3D vs that in 2D or TSV-based 3D



[8] D. H. Kim, R. O. Topaloglu, and S. K. Lim, "Block-Level 3D IC Design with Through-Silicon-Via Planning," ASPDAC 2012.

Floorplanning (1/2)

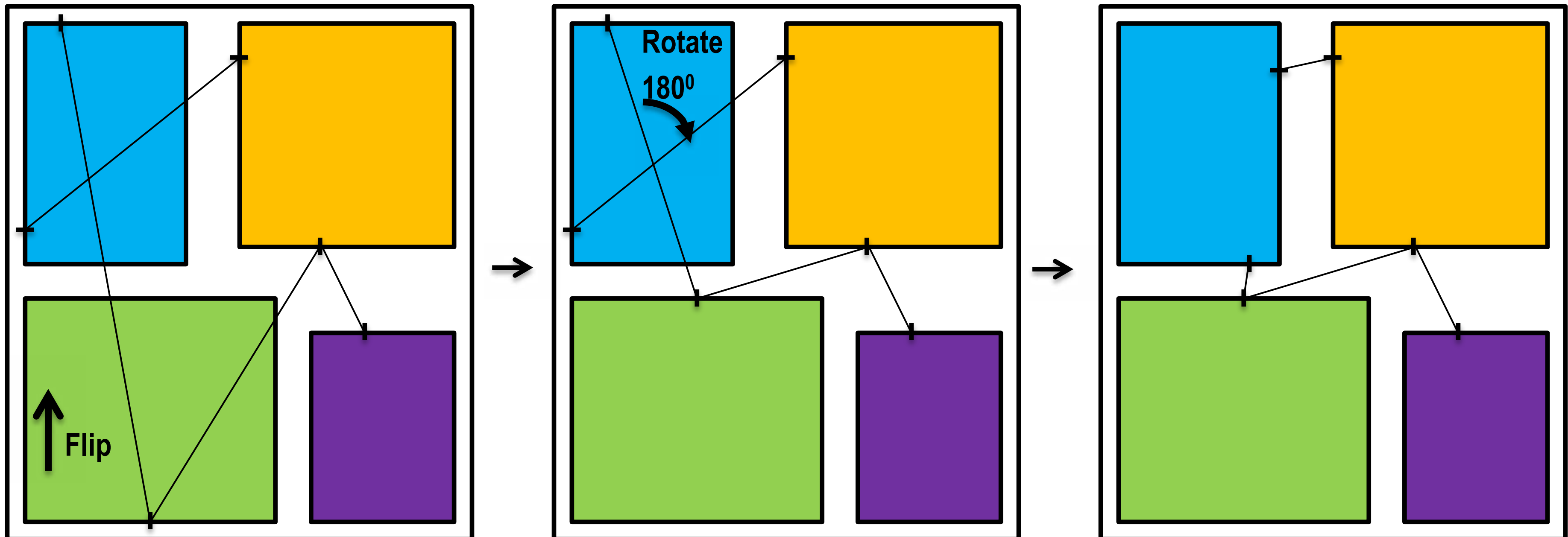
- We first perform annealing with the HPWL measured from block center-block center
- The floorplanner performs both intra-die and inter-die moves
- We group all the two-pin nets between a given pair of blocks into a single net, and increment its weight.



- Different objective functions:
  - TSV-based 3D : $\alpha.HPWL + \beta.Area + \gamma.\#TSV$
  - 2D / Monolithic 3D : $\alpha.HPWL + \beta.Area$

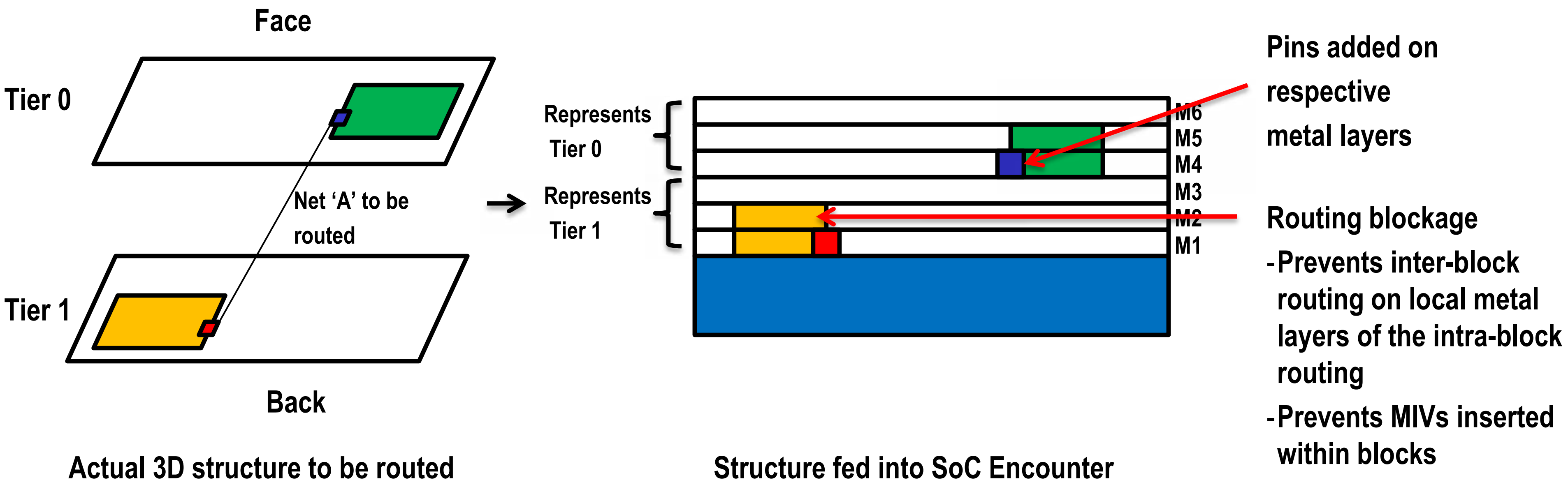
Floorplanning (2/2)

- After the relative locations of blocks are fixed, update them with pin locations.
- Each block has four possible orientations without changing the floorplan.
- Perform an annealing based refinement step to pick the best orientation of each block.



MIV Planning (1/2)

- Use an existing 2D router that can route to pins on multiple metal layers (SoC Encounter)
- Current tools can only handle 15 metal layers – 4 tiers at 3 layers per tier (for inter-block 3D routing).
- Create a netlist and DEF file to trick the 2D tool to do 3D routing.

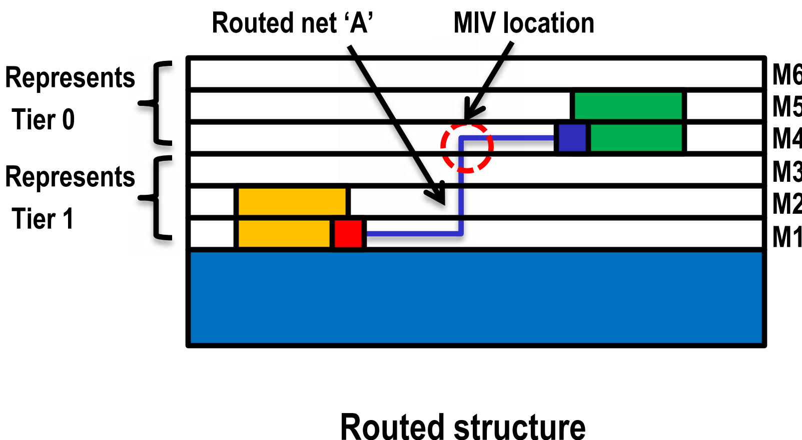


Actual 3D structure to be routed

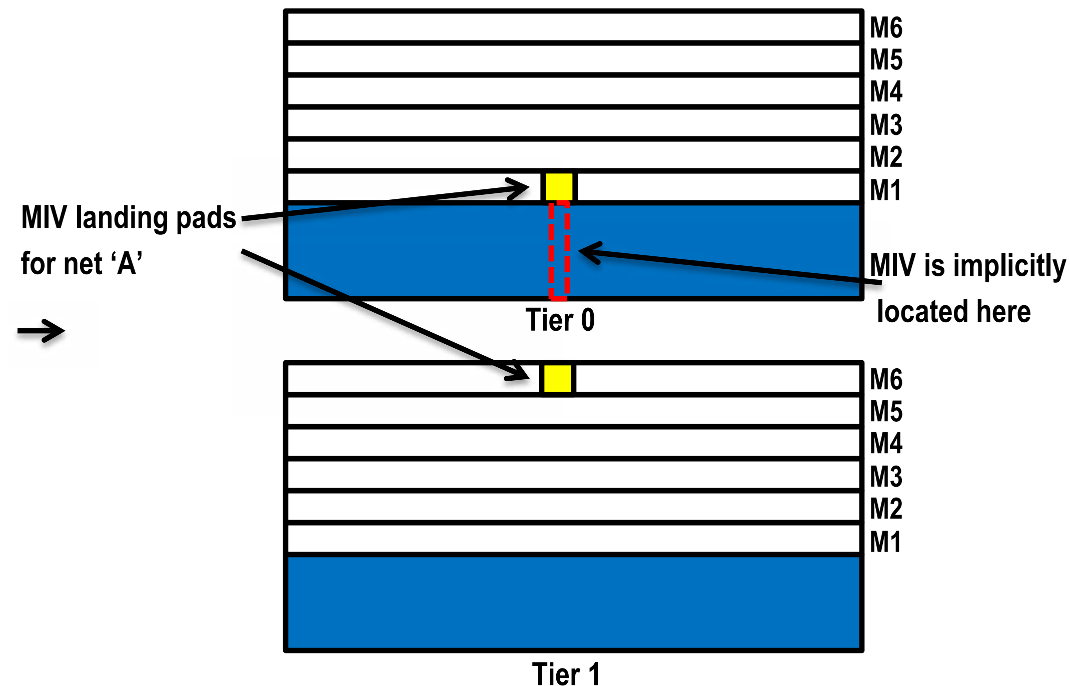
Structure fed into SoC Encounter

MIV Planning (2/2)

- All 3D nets are routed simultaneously -> prevents congestion issues.
- Once structure is routed, extract via locations and create separate Verilog/DEF file for each die.
- Each die is then routed with required number of metal layers (6 in this case).



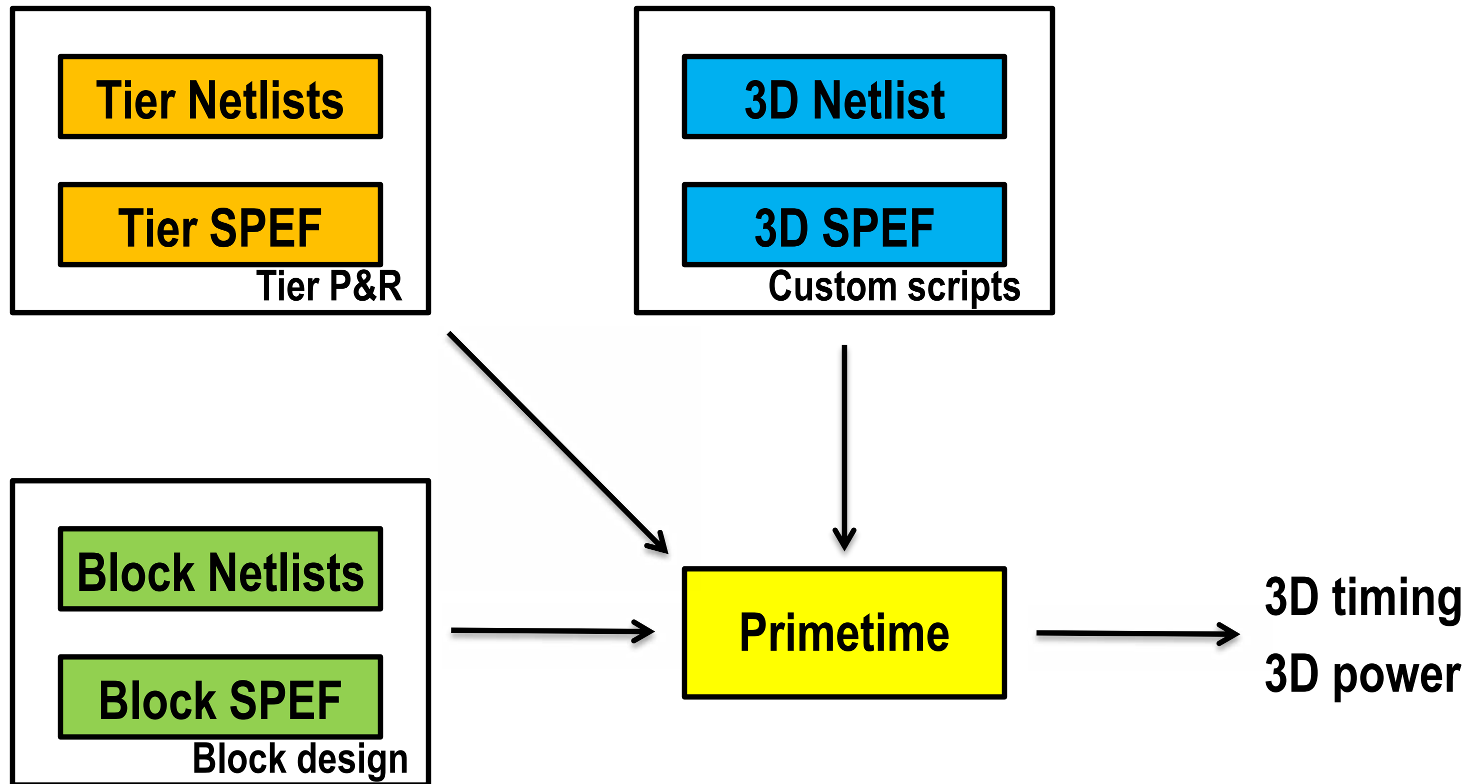
Routed structure



Separate DEF files for each tier

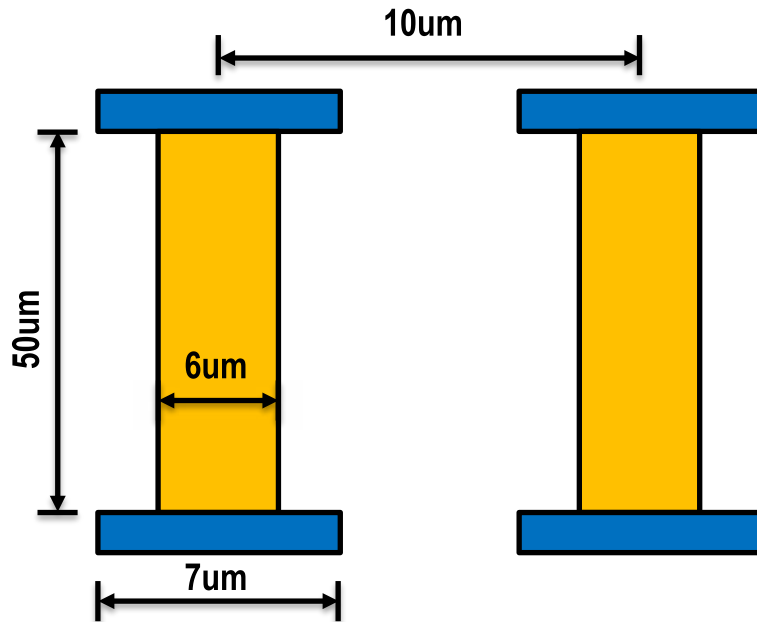
Analysis Flow

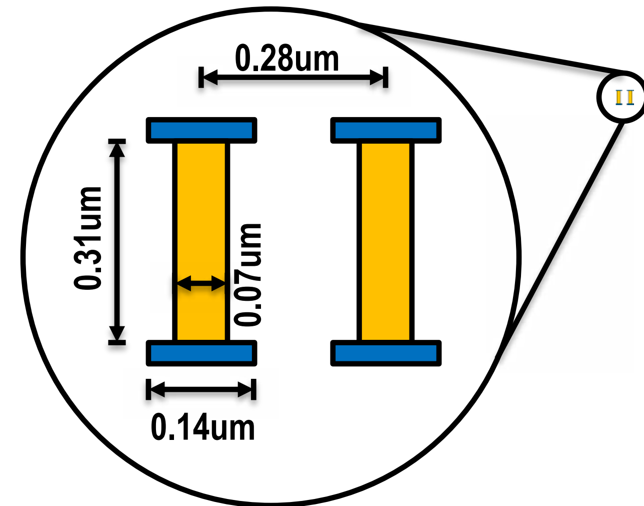
- Use Primetime to perform 3D timing and power analysis



Technology Assumptions

- Nangate 45nm std. cell library





Through-silicon-via [9]

R = 50 $m\Omega$

C = 122 fF

Monolithic inter-tier via

R = 4 Ω

C = 1 fF

[9] X. Wu et al., "Electrical Characterization for Inter-tier Connections and Timing Analysis for 3-D ICs," in TVLSI 2012

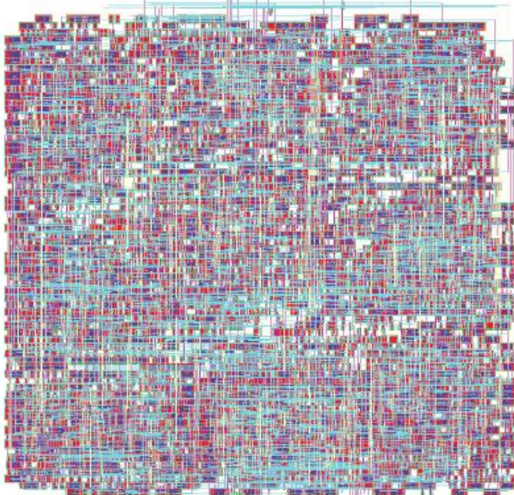
Benchmarks Statistics

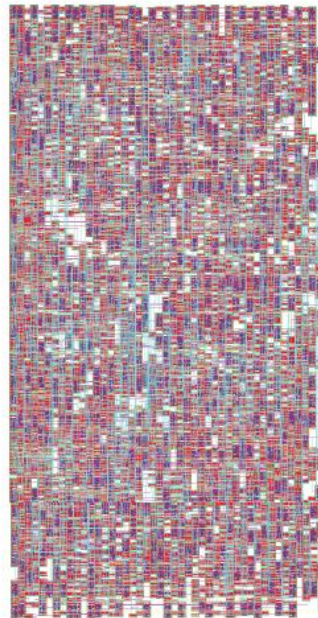
Design	Description	#Gates	#Blocks	#inter-block nets	Target period (ns)
des	Encryption core	33,024	38	2,378	0.9
rca	Reconfigurable array	146,542	95	3,135	1.3
fft	256 bit Fast fourier transform	288,145	49	1,402	1.5
mult	256-bit integer multiplier	1,639,050	127	49,471	0.845

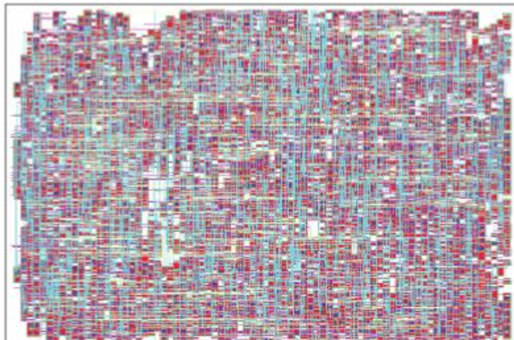
- **The first three designs are taken from the OpenCores benchmark suite**
- **The last one is a custom-built 256-bit pipelined multiplier**

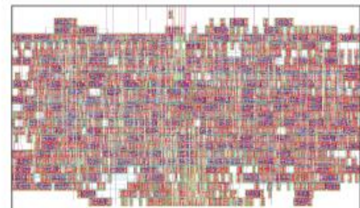
Block Design Snapshots

- **Layout snapshots for four blocks from the 'rca' benchmark.**


Block A


Block B

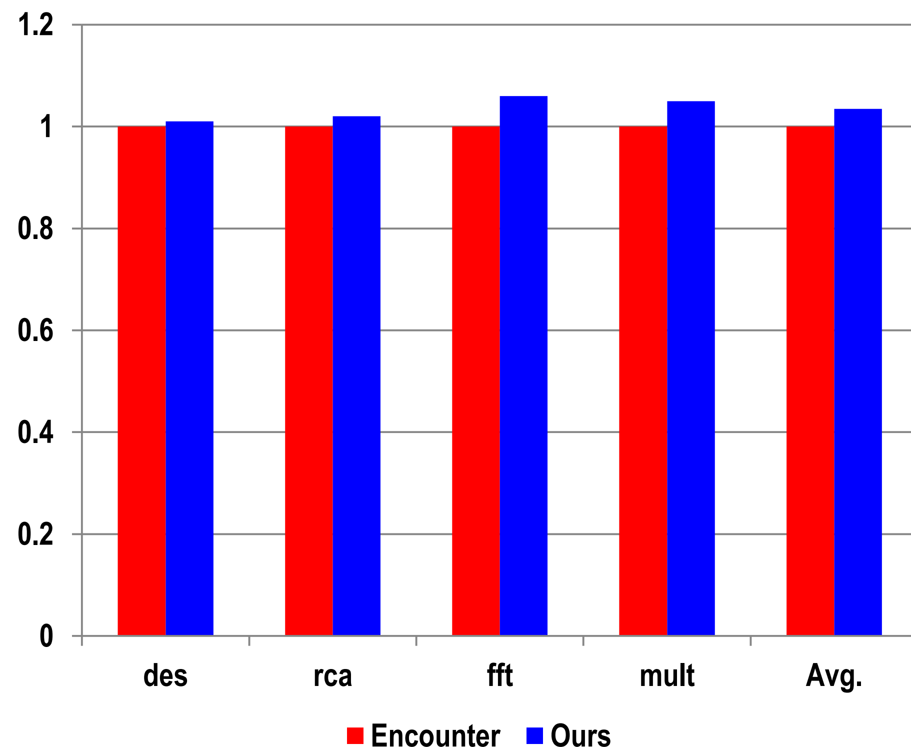

Block C

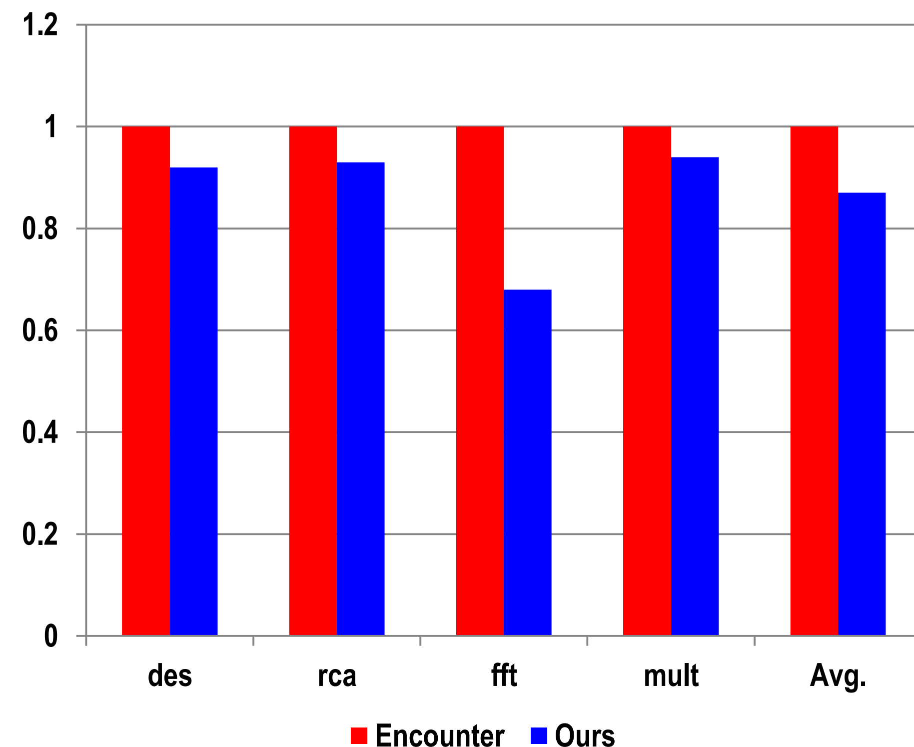

Block D

	Footprint (um x um)	#Gates	Wirelength (um)
Block A	103x100	5,066	47,190
Block B	125x64	3,950	37,380
Block C	73x110	3,725	34,150
Block D	63x36	1,128	8,932

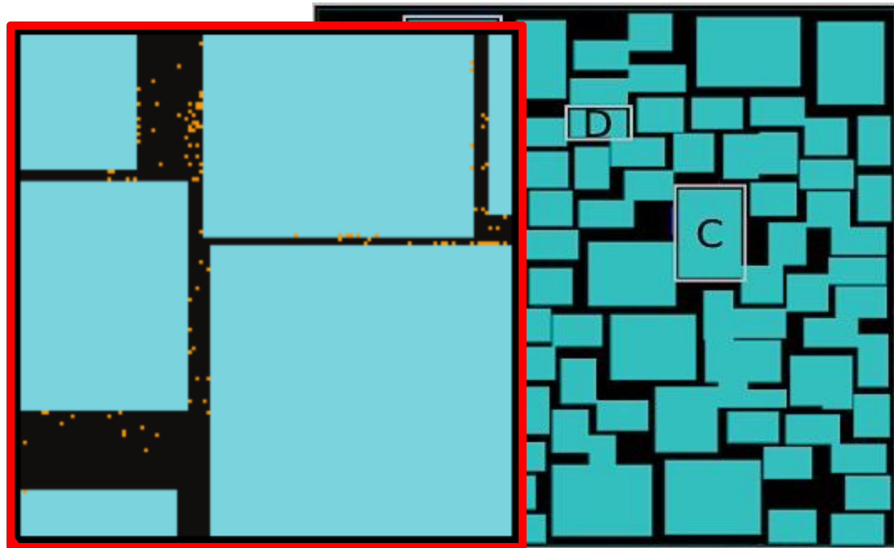
Floorplanner Validation

- A 2D implementation using our tool gives us routed wirelength within 3.5% (on average) of that of Cadence SoC Encounter.
- The area of our 2D implementation is 13% less than that of Encounter (7% excluding fft)

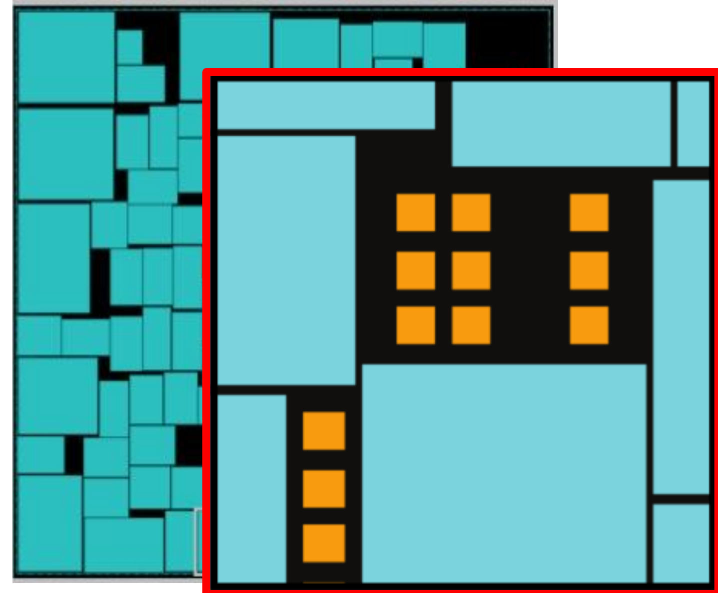




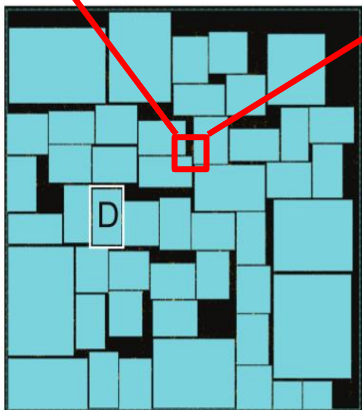
Floorplan Screenshots (rca benchmark)



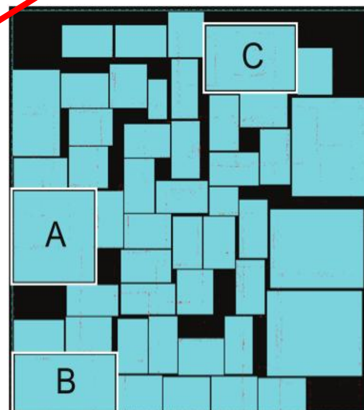
2D Encounter



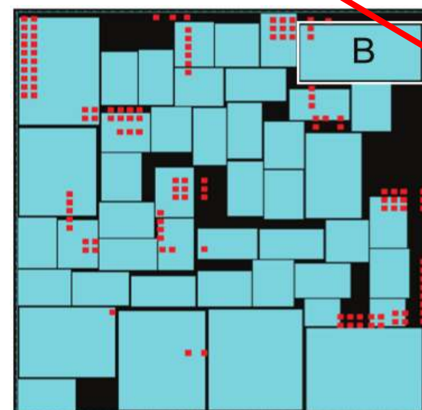
2D Ours



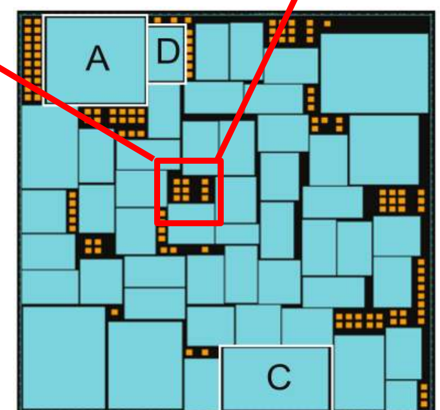
MIV – Tier 0



MIV – Tier 1

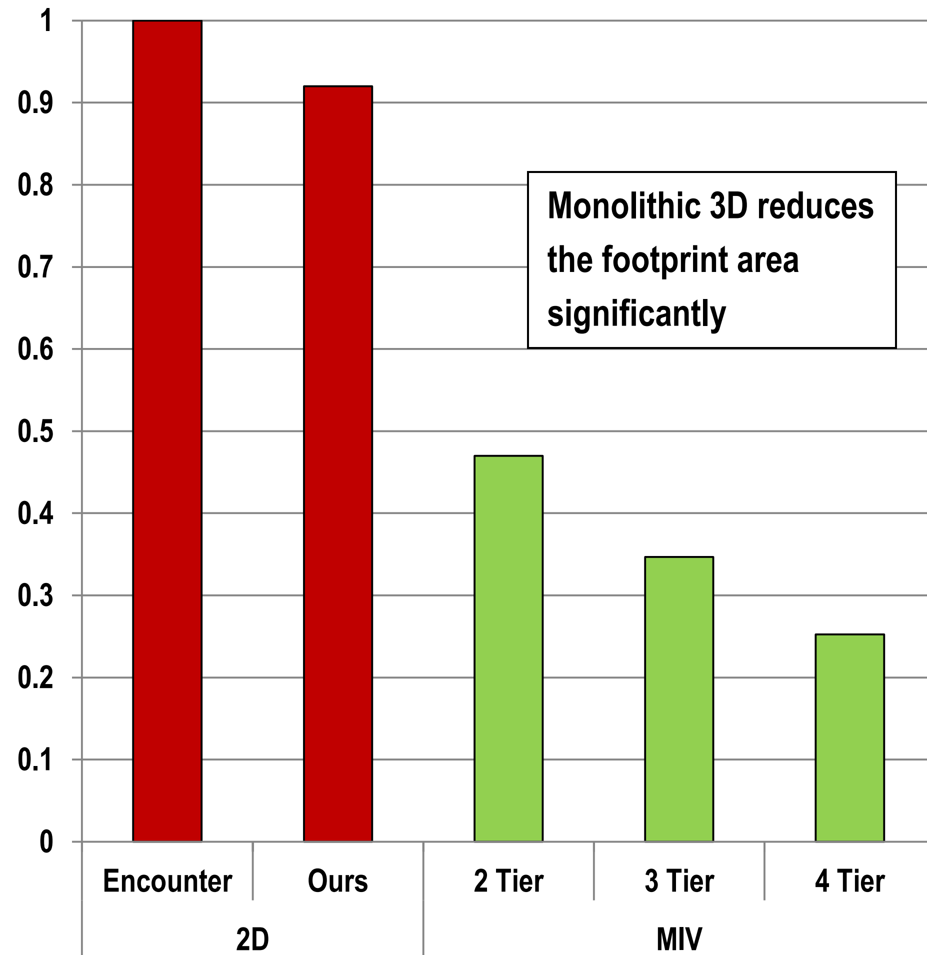


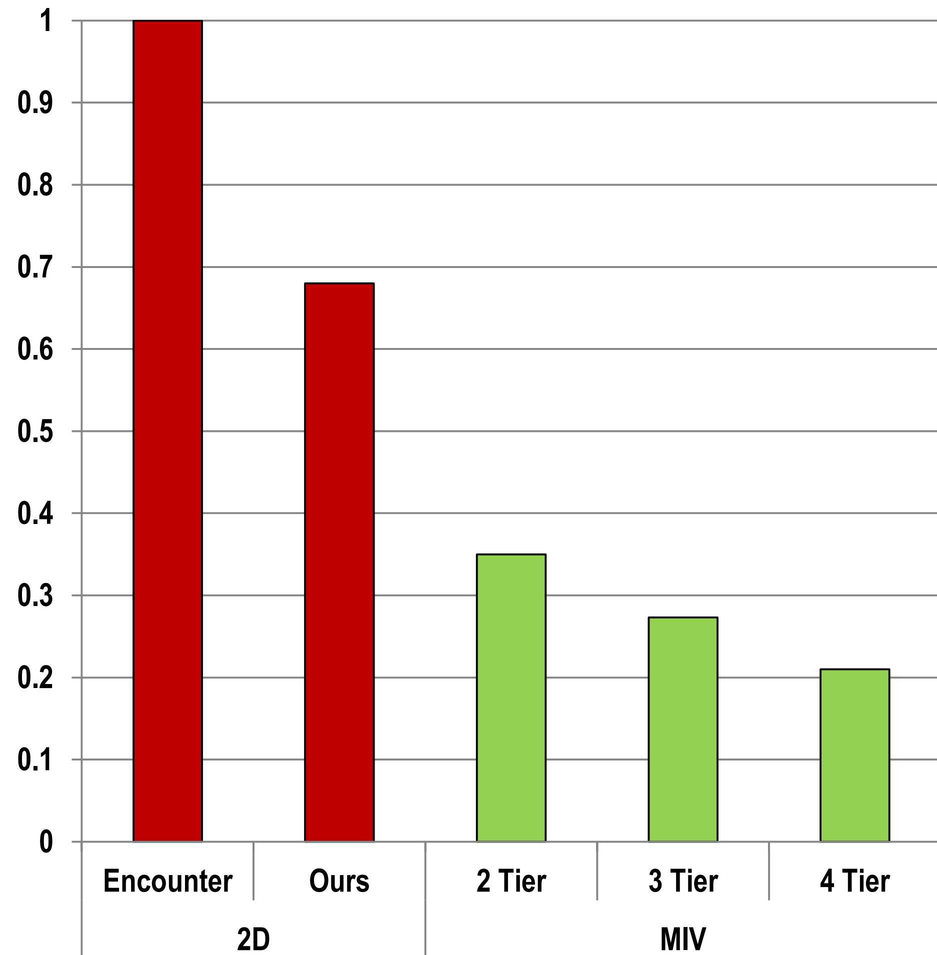
TSV – Die 0



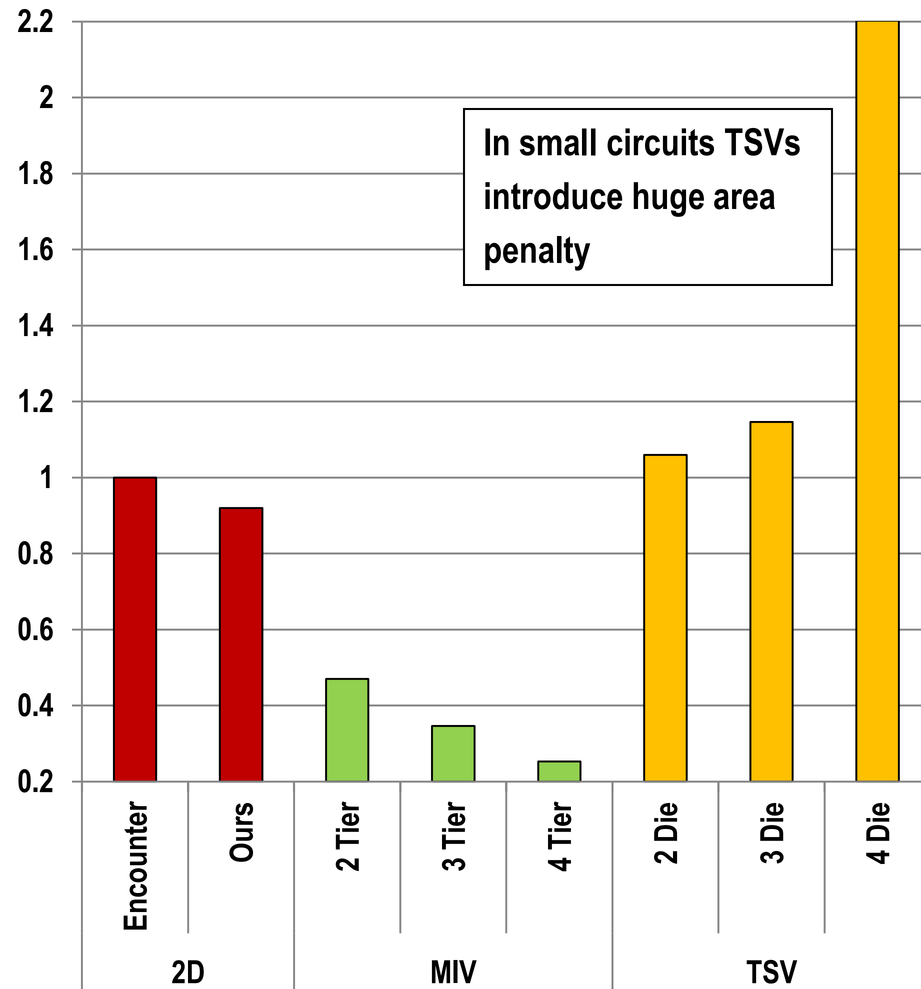
TSV – Die 1

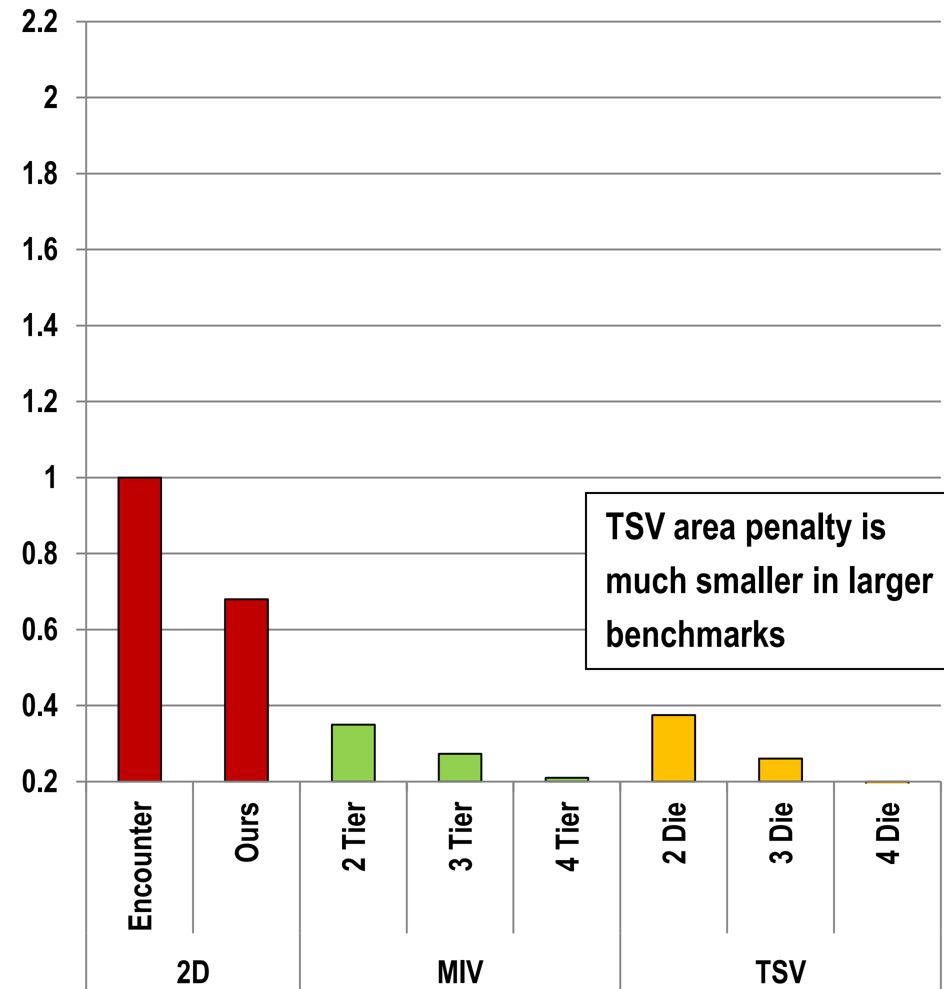
Footprint Area Comparisons (1/2)



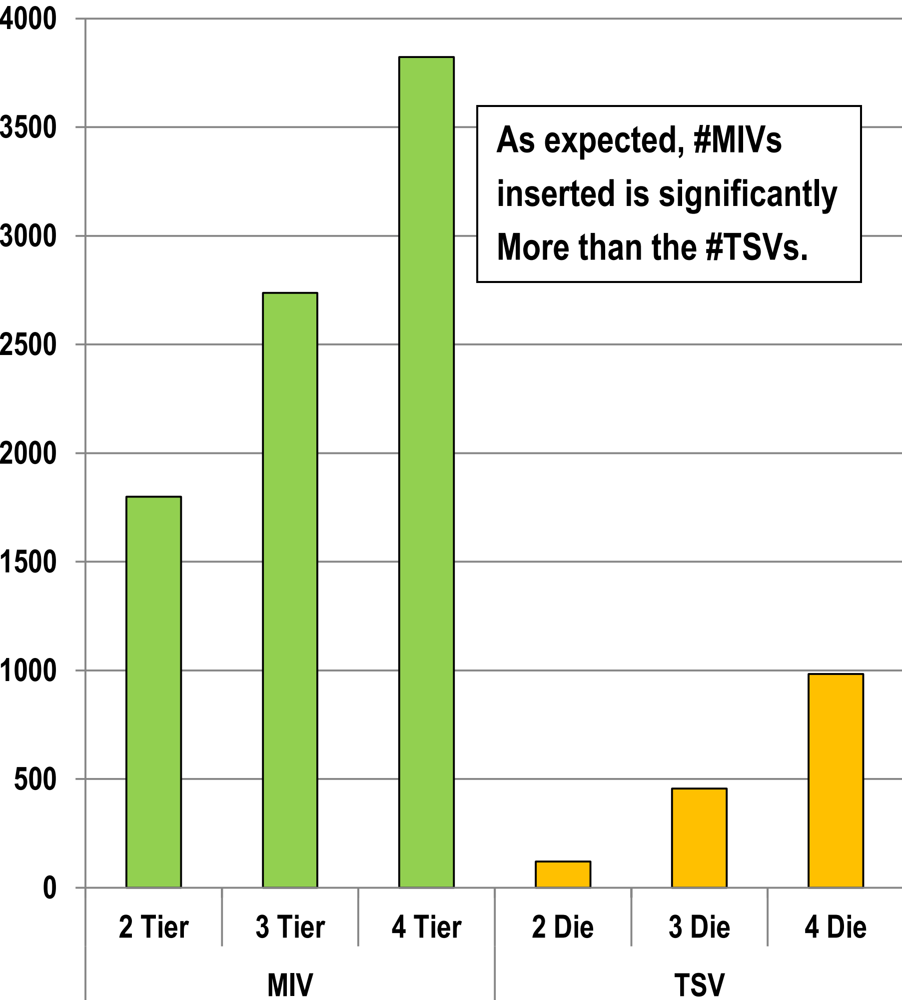


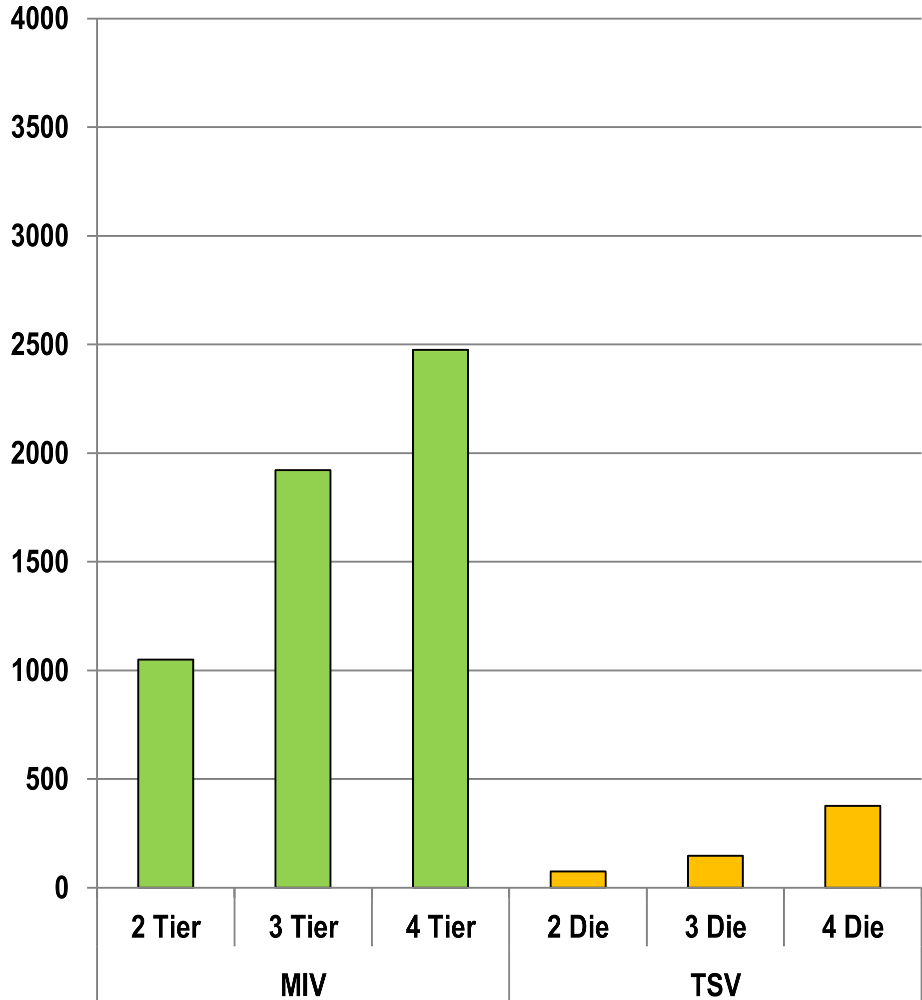
Footprint Area Comparisons (2/2)



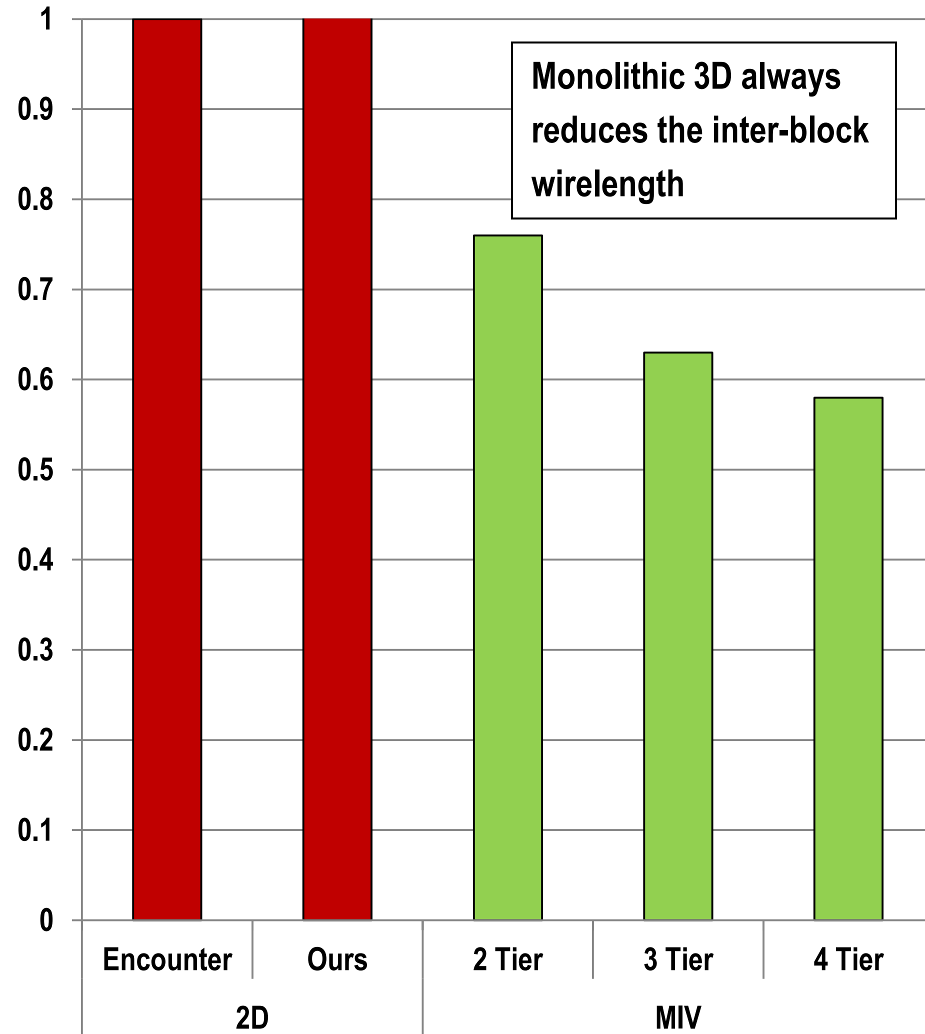


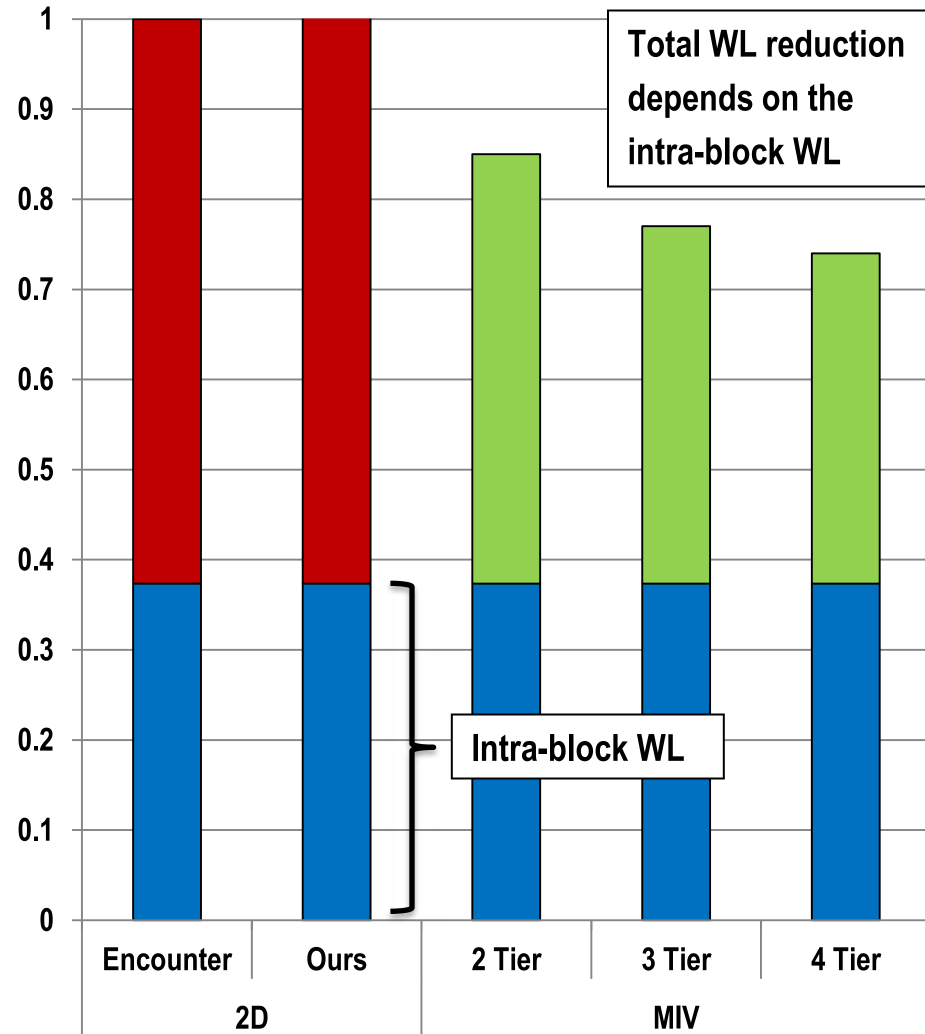
Number of Inserted Vias



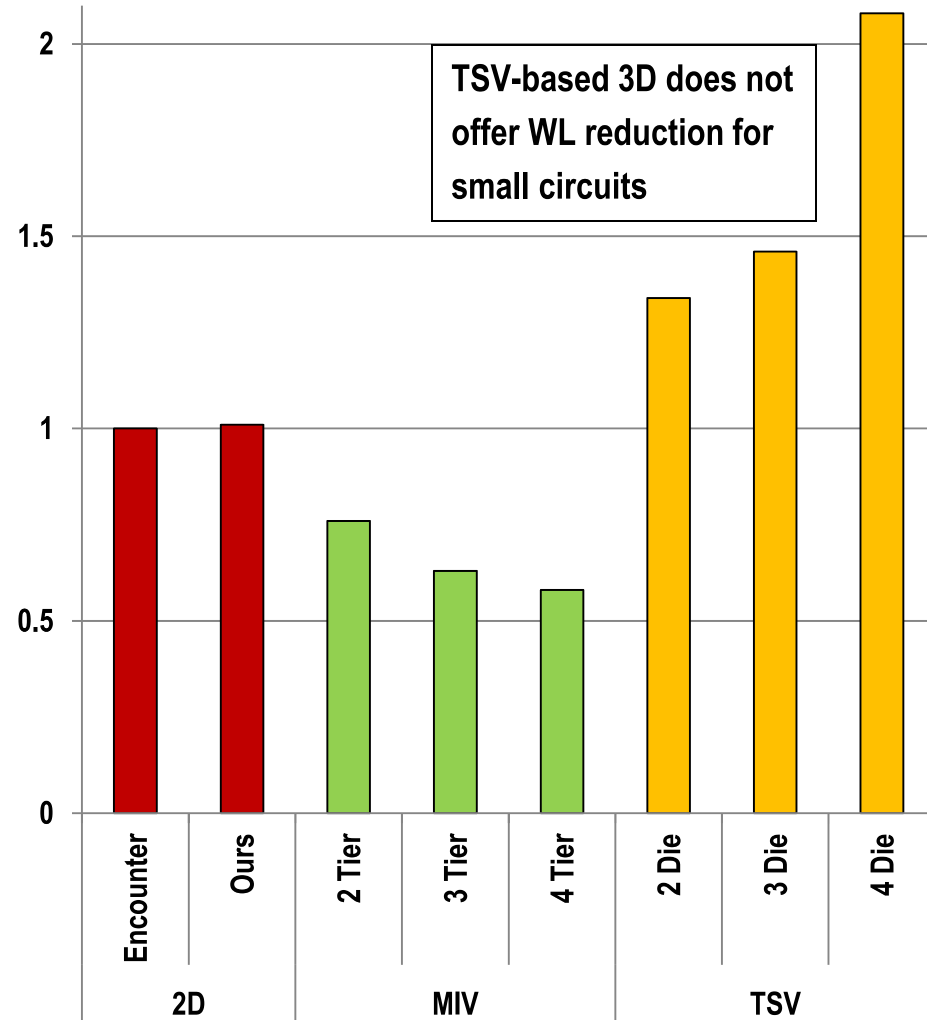


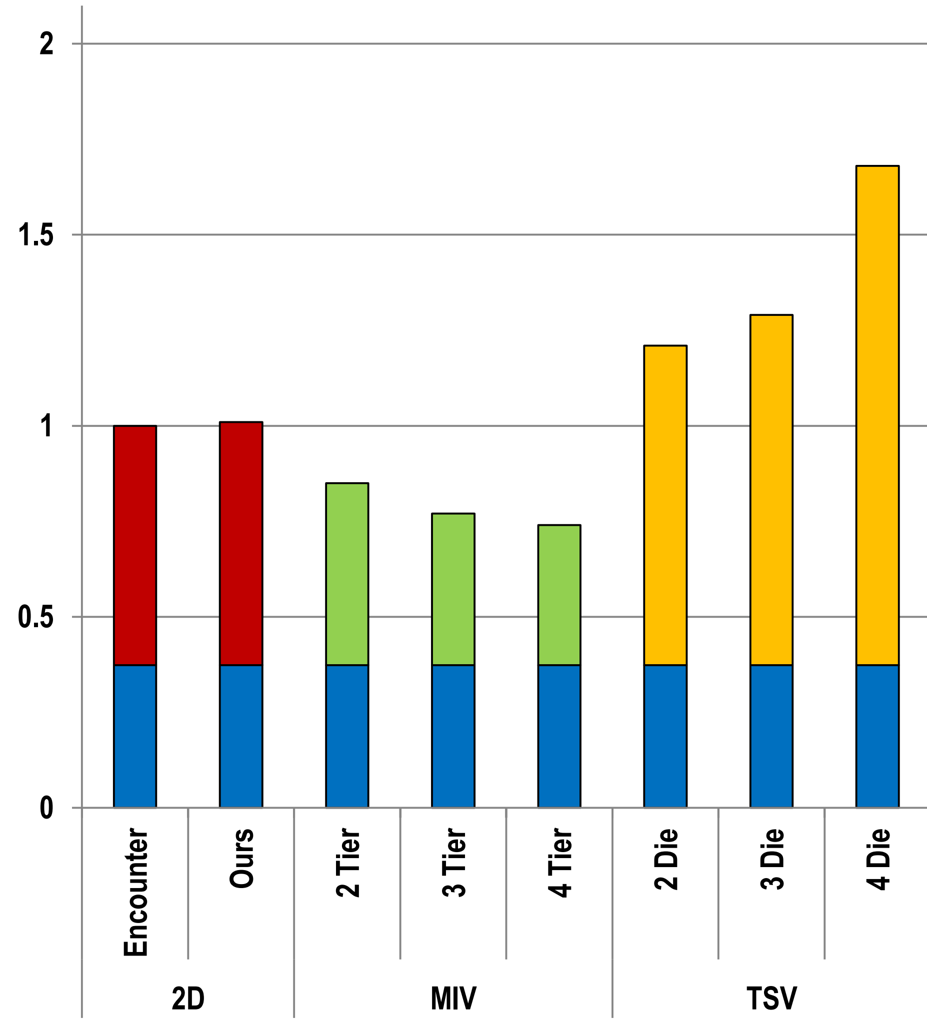
Wirelength Comparisons – des (1/2)



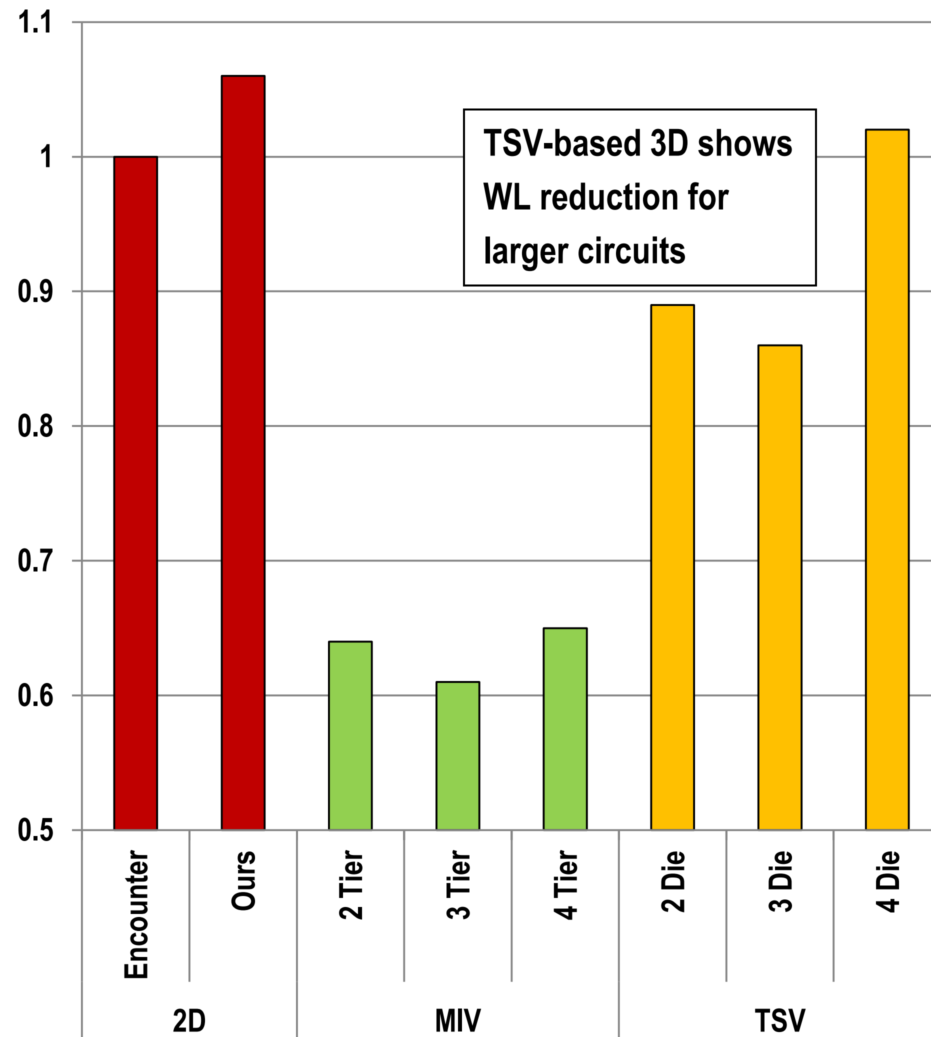


Wirelength Comparisons – des (2/2)





Wirelength Comparisons – fft



Total

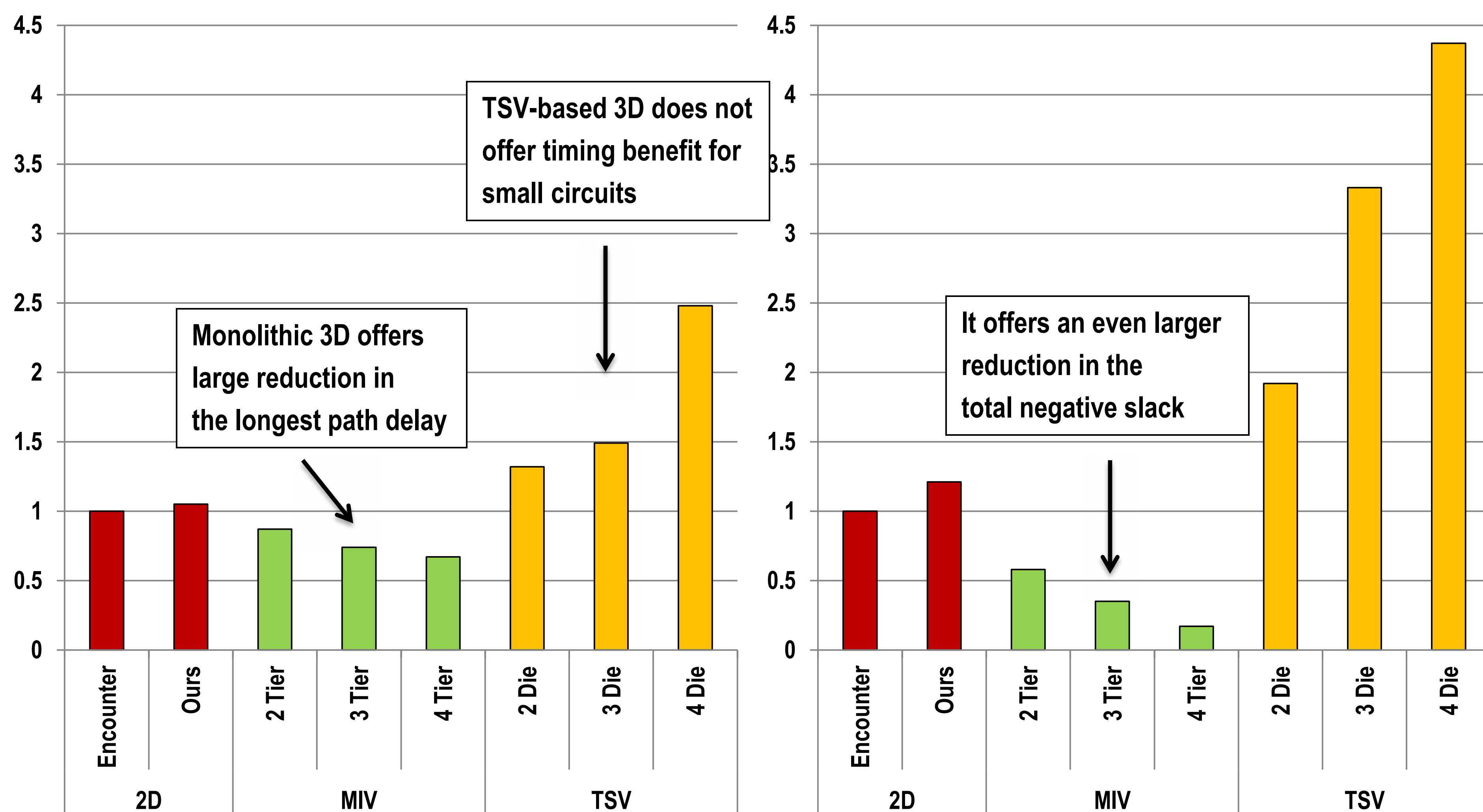
1.1
1
0.9
0.8
0.7
0.6
0.5

Large intra-block WL
limits the total WL
reduction

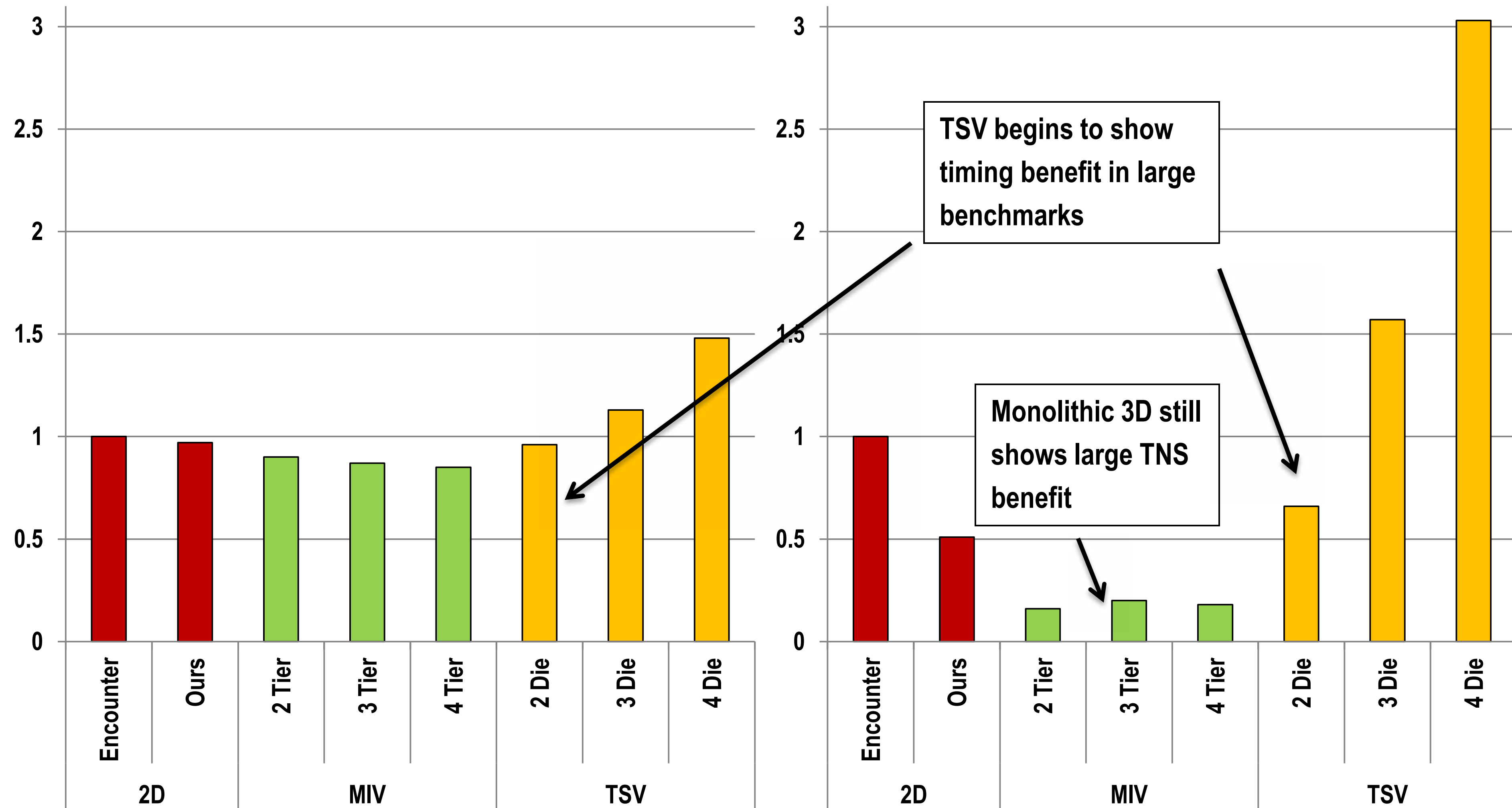
Encounter
Ours
2 Tier
3 Tier
4 Tier
2 Die
3 Die
4 Die

2D
MIV
TSV

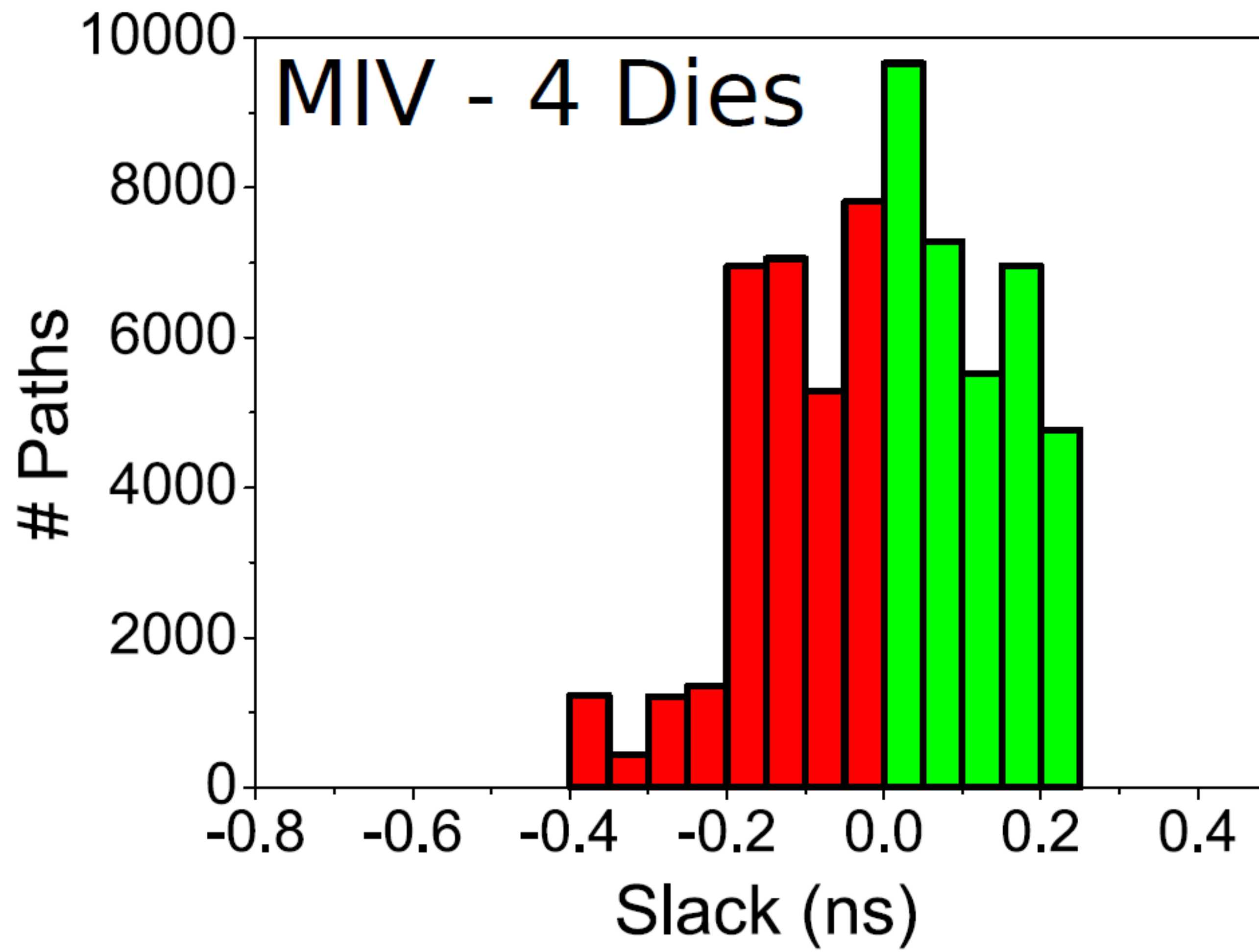
Timing Comparisons - des



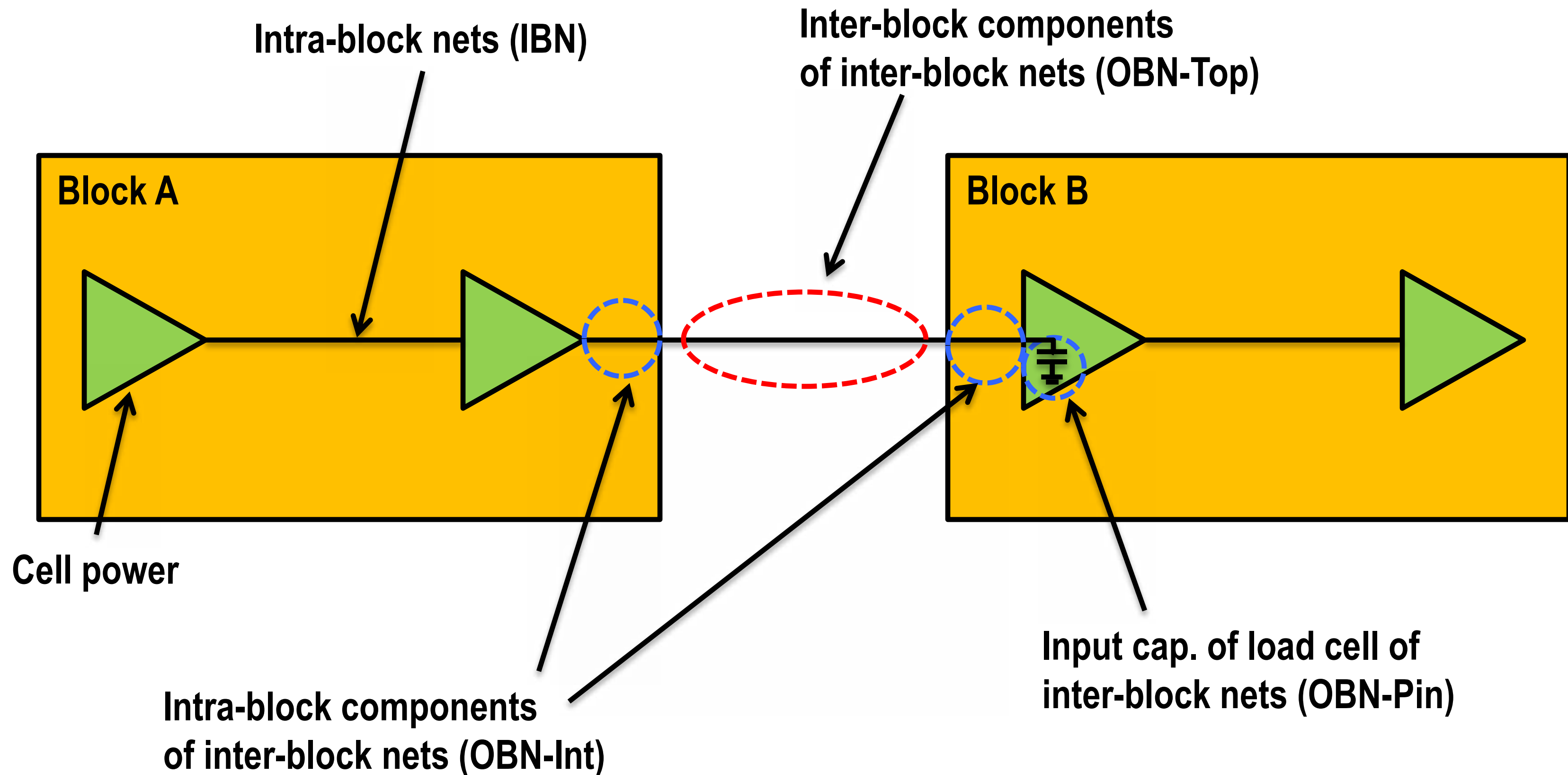
Timing Comparisons - fft



Slack Histograms - FFT Benchmark

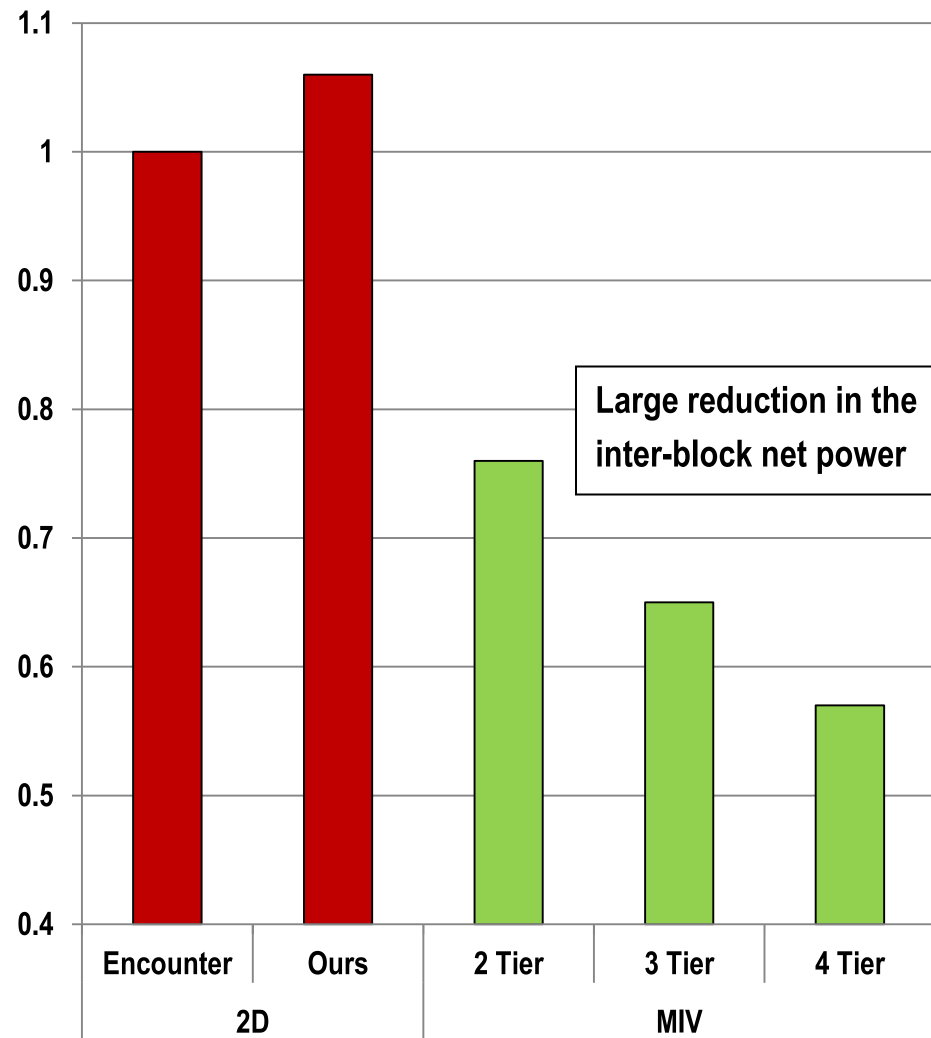


Power Components



- **Ideal implementation: parasitics of OBN-Top = 0**

Power Comparisons - des



Total

1.1
1
0.9
0.8
0.7
0.6
0.5
0.4

This is the theoretical minimum power at the block-level

Total power depends on:
- Cell power
- Other components of net power

Encounter | Ours | 2 Tier | 3 Tier | 4 Tier | Ideal

2D | MIV | Ideal

Power Comparisons - fft

Total

Group	Design	Total
2D	Encounter	1
2D	Ours	1
MIV	2 Tier	0.99
MIV	3 Tier	0.99
MIV	4 Tier	0.99
Ideal		0.98

We close the gap to ideal by 50%

Total

Large intra-block WL Limits the total power savings

Summary of monolithic 3D vs 2D

- Negligible total silicon area penalty (Max 2%)
- Up to 42% reduction in the inter-block WL
- Up to 33% reduction in the longest path delay
- Up to 82% reduction in the total negative slack
- Up to 43% reduction in the inter-block net power

Conclusions

- We have developed a methodology to obtain post-layout results of block-level monolithic 3D using commercial tools.

- We observe almost no area overhead and observe significant reduction in:
  - Inter-block WL (and consequently, total WL)
  - Longest path delay
  - Total negative slack
  - Inter-block net power (and consequently, total power)

- Due to the large reduction in TNS, other power reduction methods are more applicable
  - Multi Vth
  - VDD reduction

Thank you.

Questions ?