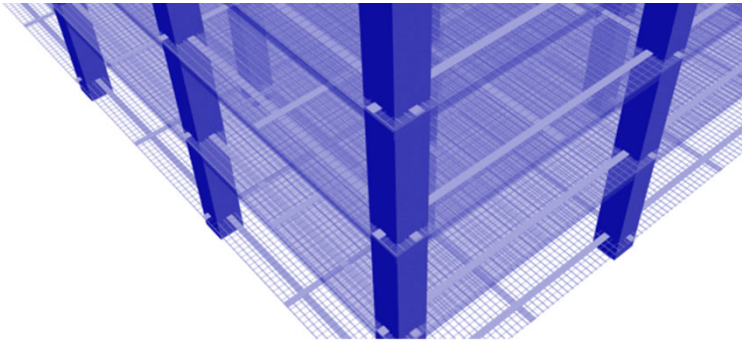
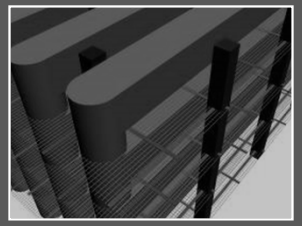
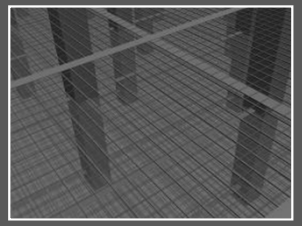
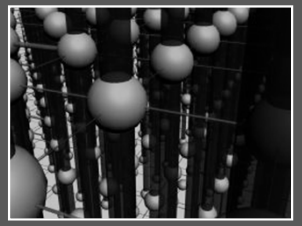


Heterogeneous 3D ICs: Benefits, Challenges, and Future Prospects

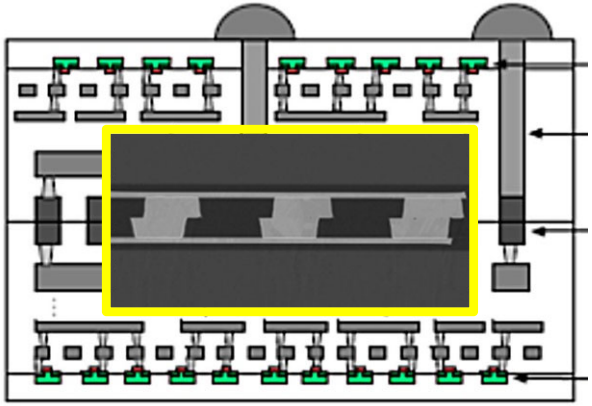
Prof. Sung Kyu Lim

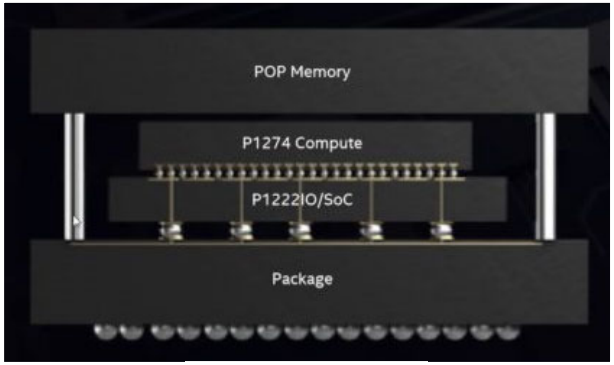
Georgia Institute of Technology

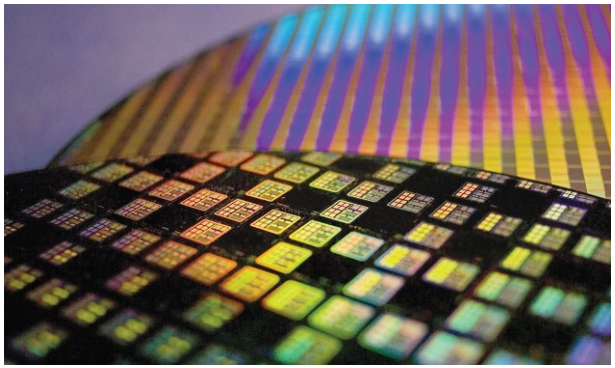
9/17/2019

ARM Research Summit

Recent Race for 3D Chips (Logic)

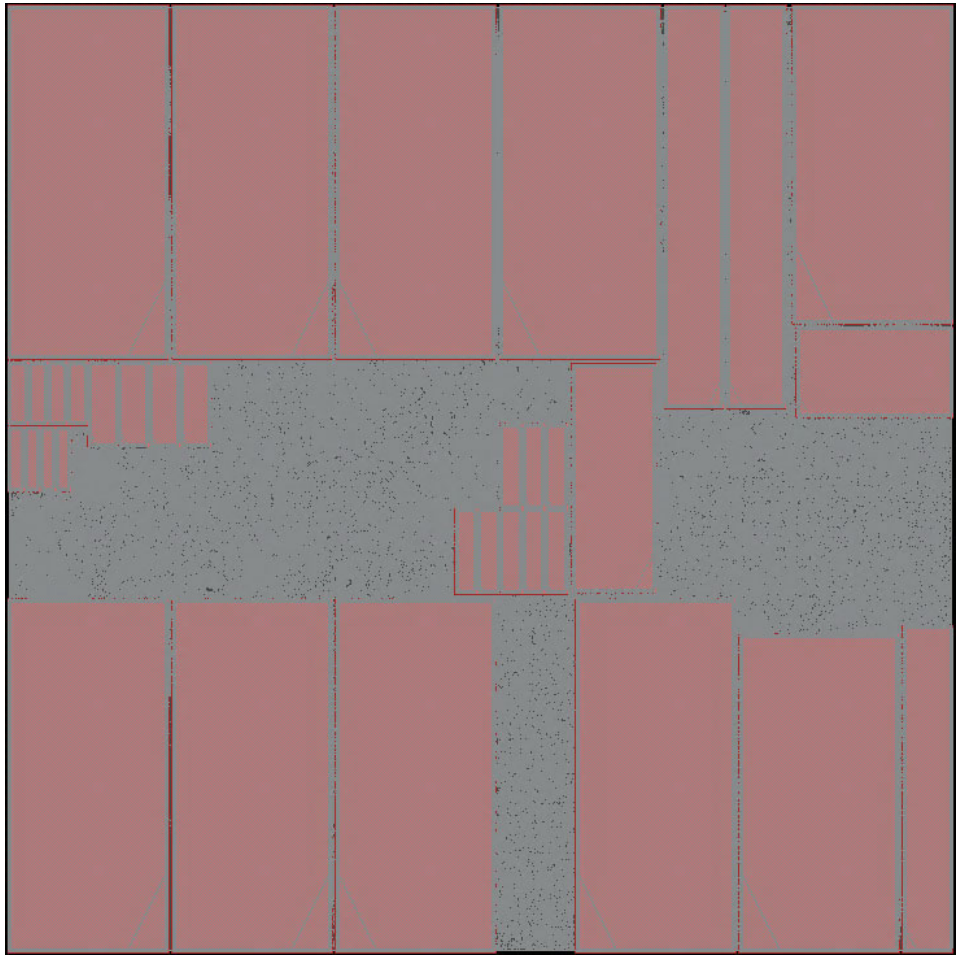




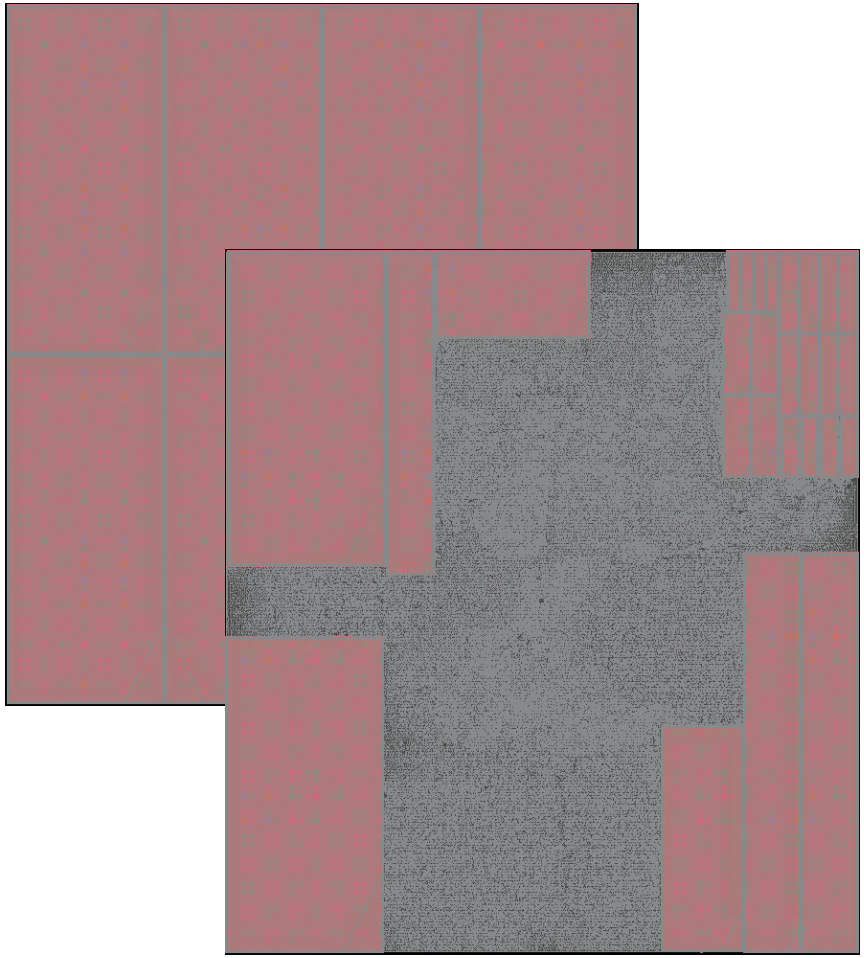


arm GLOBALFOUNDRIES intel tsmc

Logic-On-Memory with R-arch



2D IC (TSMC 28nm)



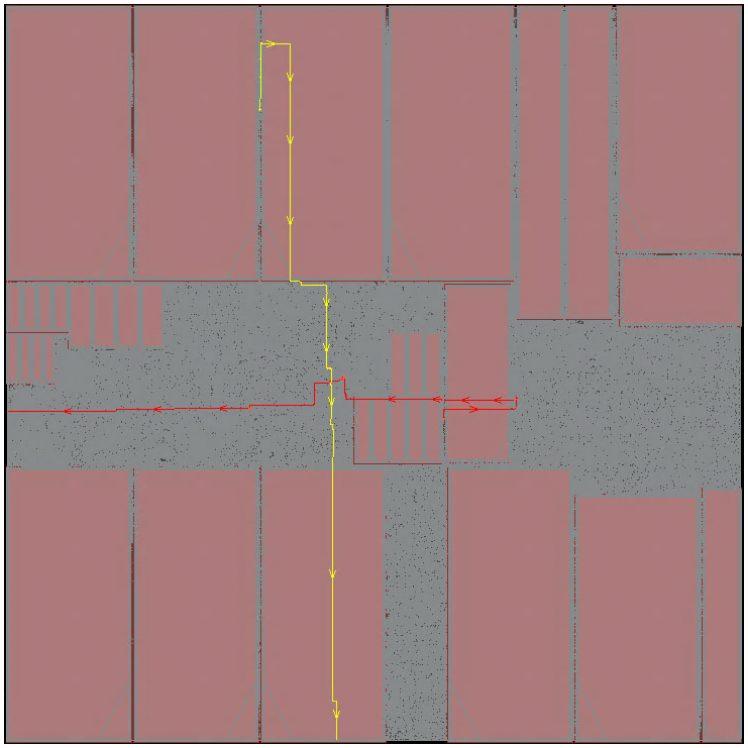
3D IC (TSMC 28nm)

PPA Comparison

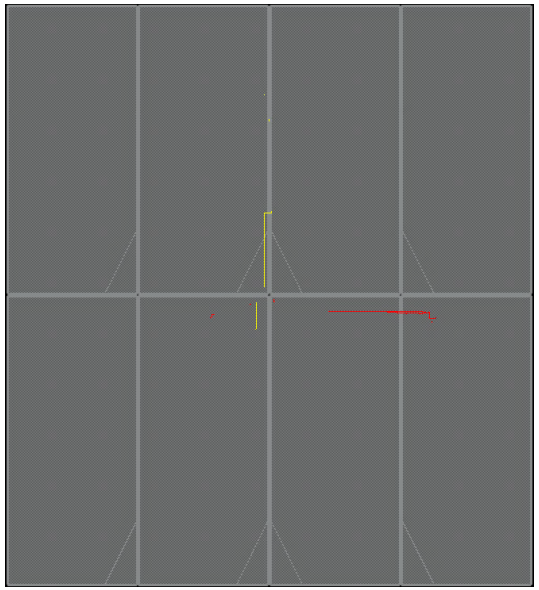
- **37% performance improvement over 2D**
  - **300K-gate 64-bit OOO RISC-V core**
  - **TSMC 28nm**

	2D	M3D	%
Clock Frequency (MHz)	475	650	36.8
Footprint (um)	1970x1970	1322x1466	-50.1
WL (m)	12.09	10.96	-9.3
Total power (mW)	292.7	399.3	36.4
WNS (ps)	-10	-22	-
Design time (hours)	19	21	

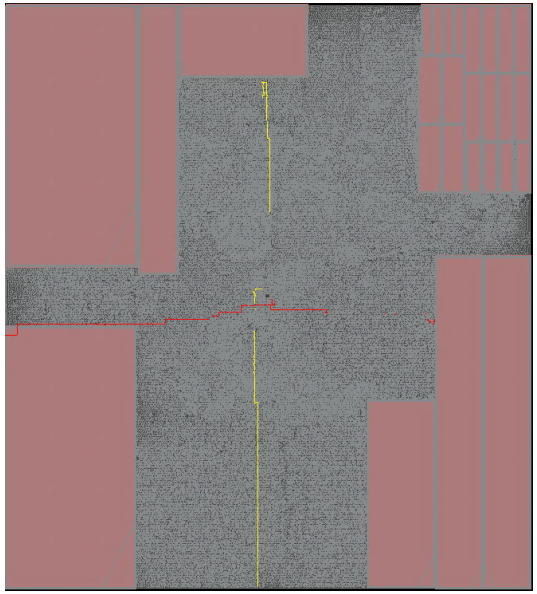
Key Recipe #1: Logic Path



memory die



logic die



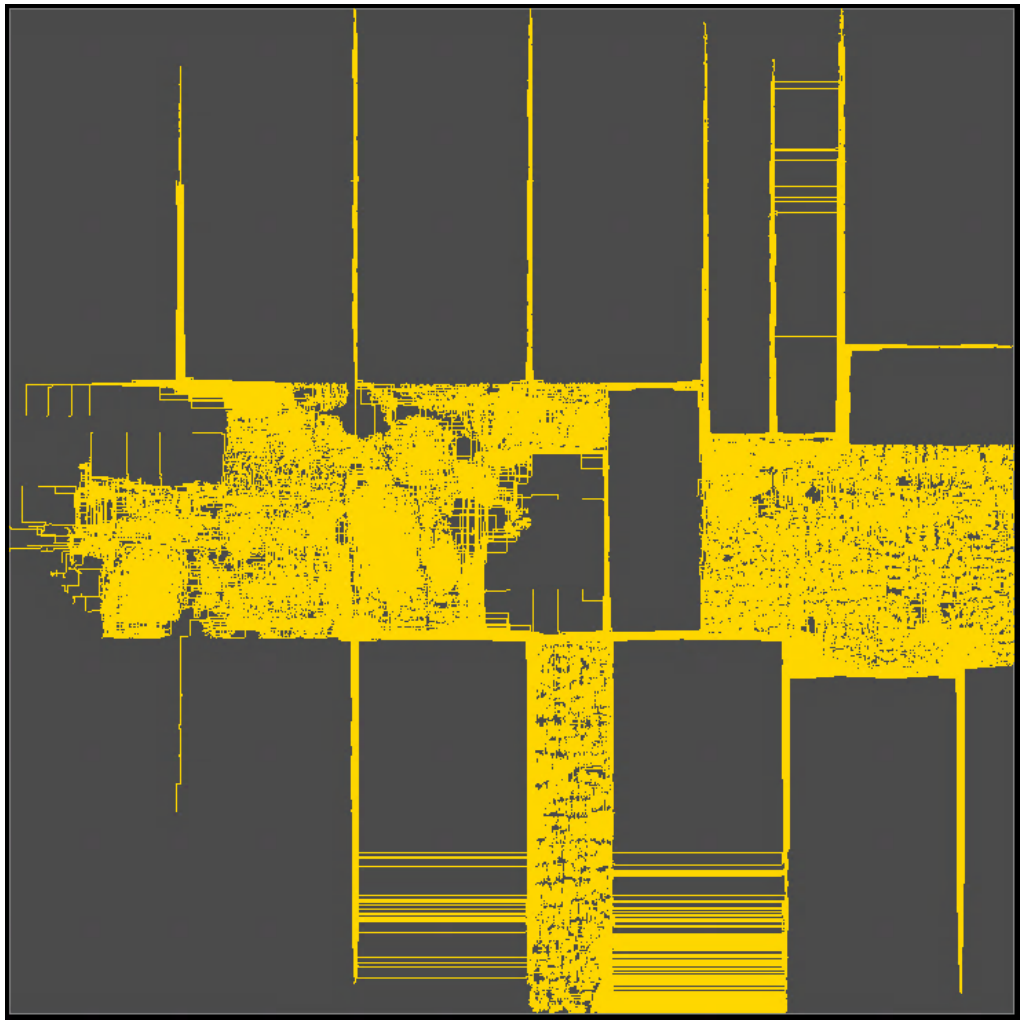
2D (yellow): 2.12ns

3D (red): 2.01ns

3D (red): 1.56ns

yellow: 1.51ns

Key Recipe #2: Clock Path

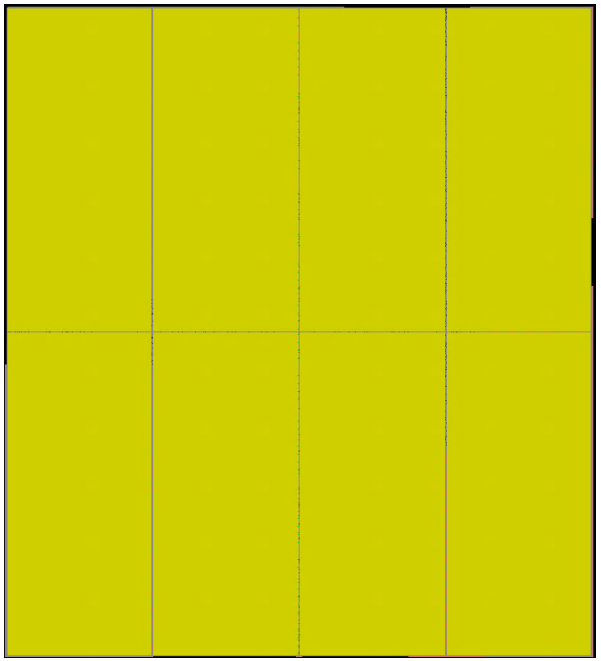


2D

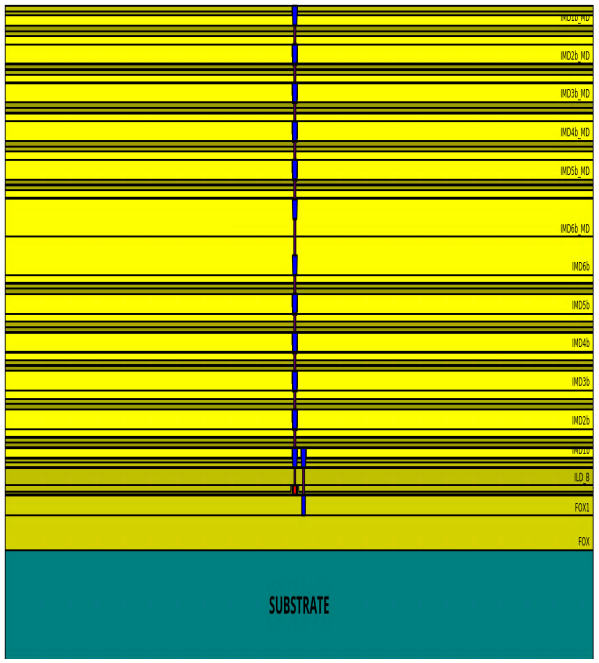
M3D (logic tier): better CLK

Macro-3D Flow

- **Idea: Use 2D commercial tools**
  - **Design the memory tier manually first and save pin locations**
  - **Conduct 2D P&R for both memory & logic tiers with double metal stack**



1. Macro Placement

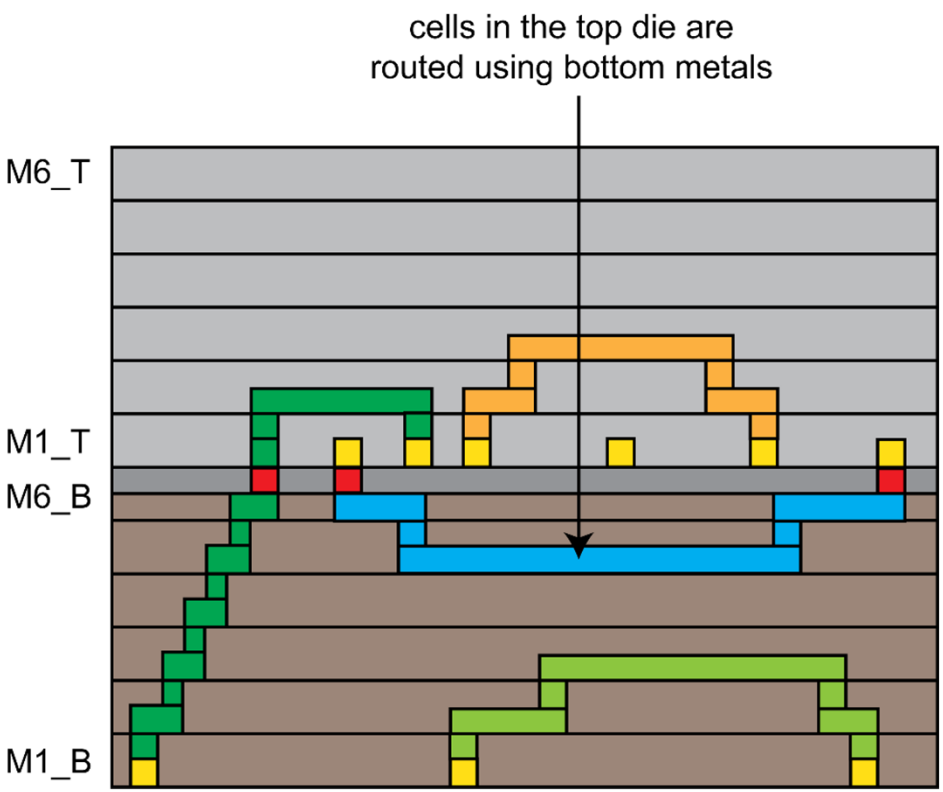


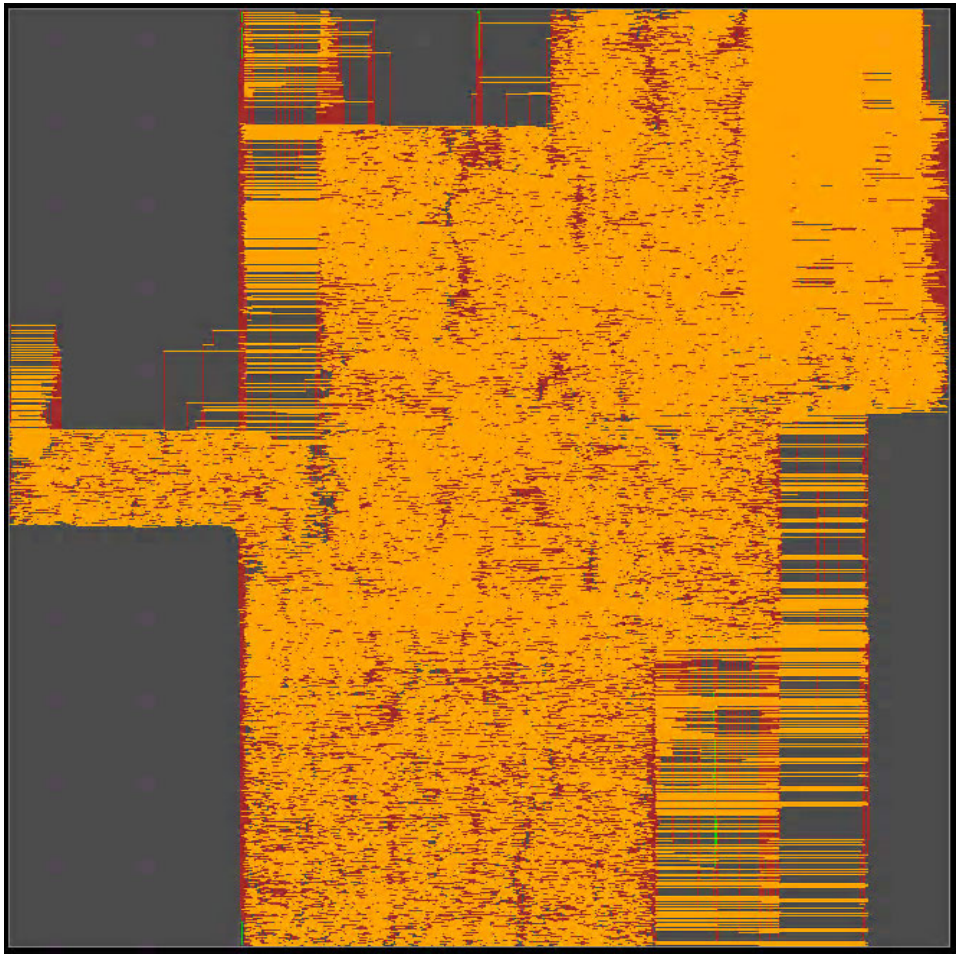
2. Project Pins

3. Logic Tier P&R

Macro-3D Flow (cont)

- **Observations**
  - **Pro: cross-tier metal sharing**
  - **Con: high F2F usage (252K)**

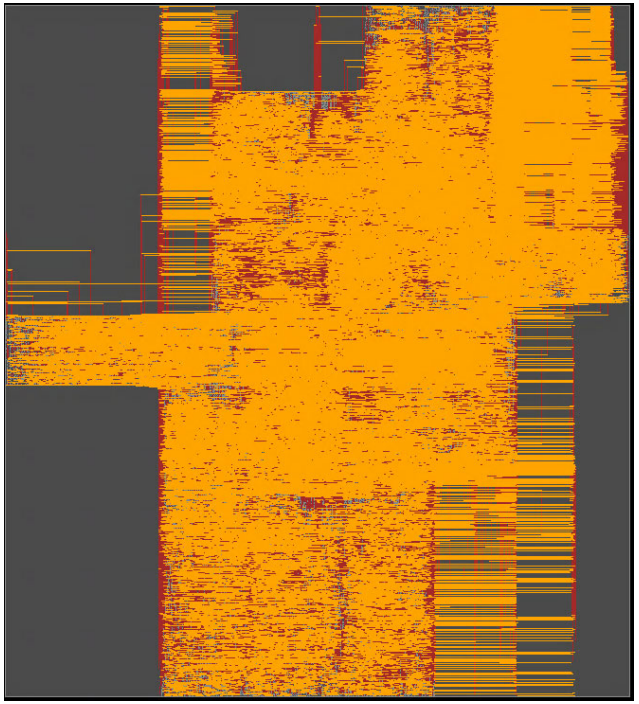




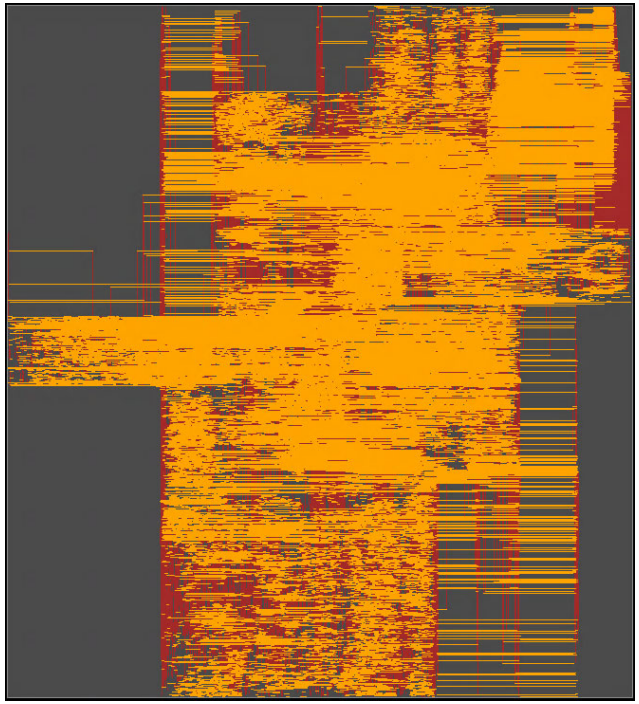
M5/M6 of memory die (bottom) used for logic die (top) routing

Controlling F2F Usage

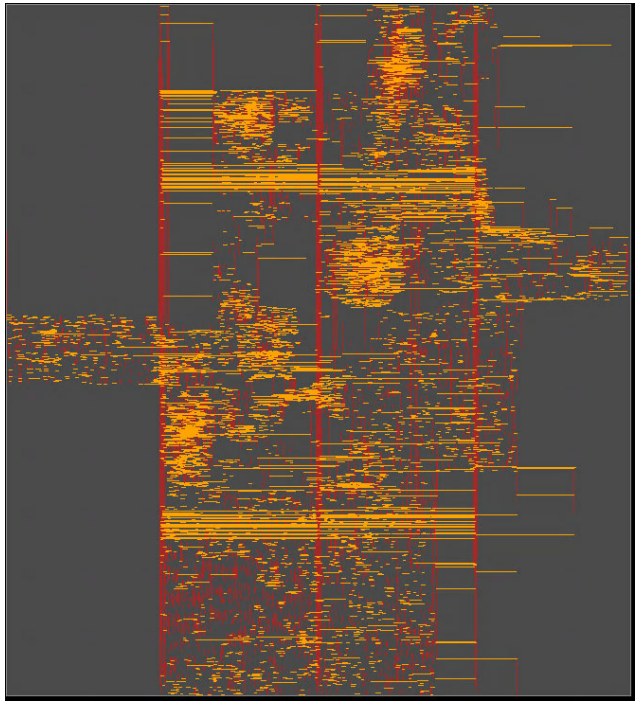
- **Found a way to control F2F usage**
  - **F2F count vs. delay tradeoff exists**



252K F2F
WNS = – 22ps

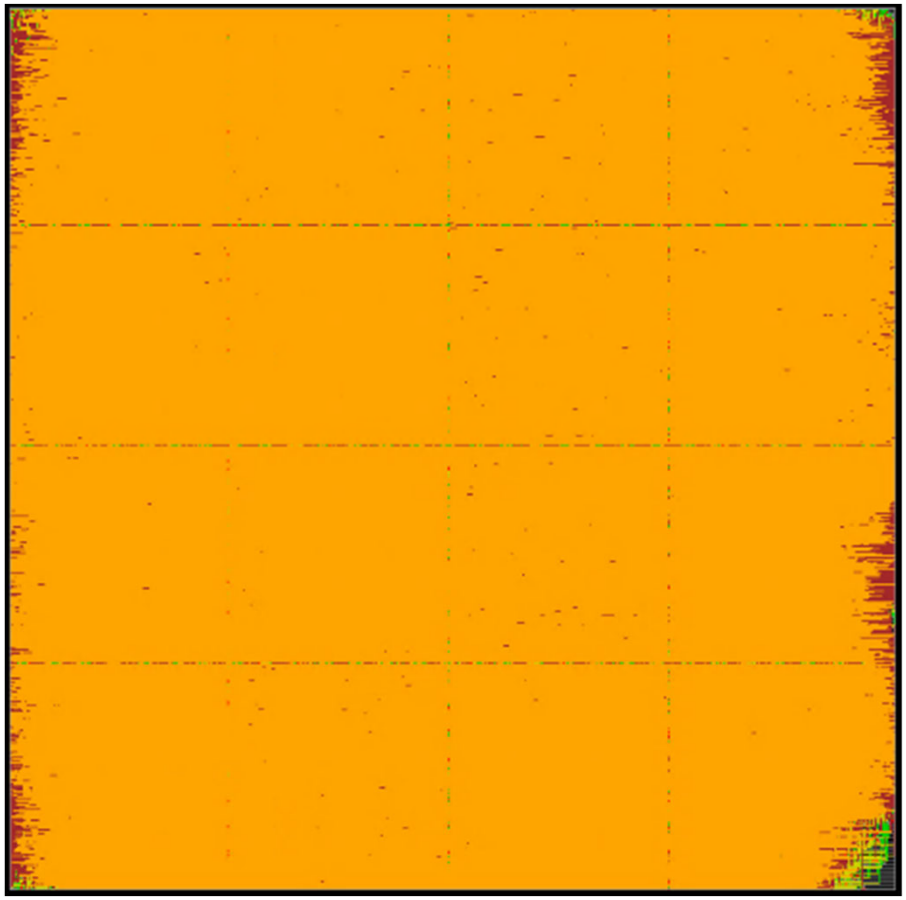


105K F2F
WNS = – 32ps

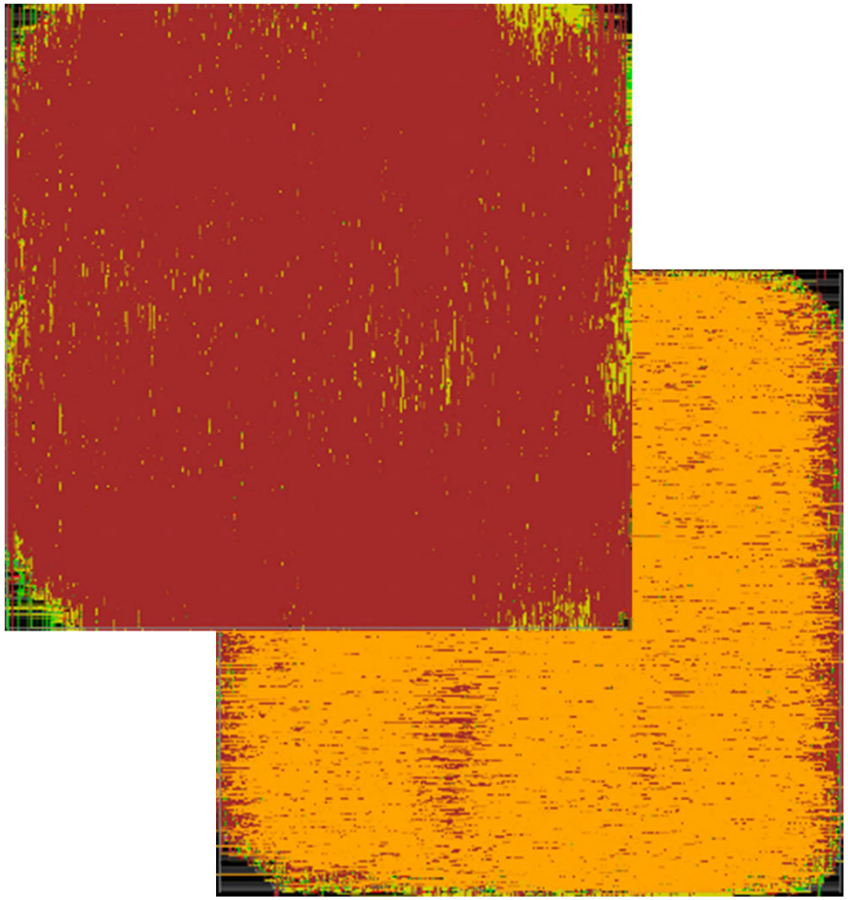


35K F2F
WNS = – 47ps

Logic-On-Logic for LDPC



2D IC (TSMC 16nm)



3D IC (TSMC 16nm)

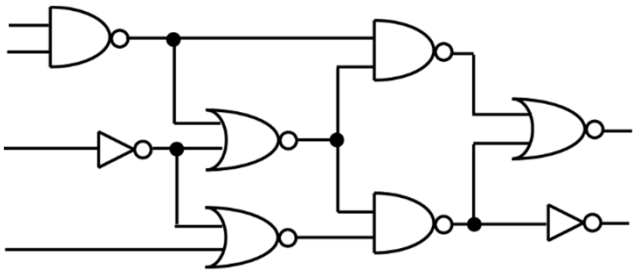
LDPC PPA Details

- **39% power improvement over 2D**
  - **With a 50K-gate error-correction benchmark @ TSMC 16nm**

metric	2D	3D	Δ%
target freq (MHz)	2500	2500	0.0
cell count	55,276	45,653	-17.4
WL (m)	1.389	0.913	-34.3
pin cap (pF)	161.3	109.3	-32.2
wire cap (pF)	263.3	173.3	-34.2
total power (mW)	214.2	130.9	-38.9

Pseudo vs. True 3D IC Tools

circuit



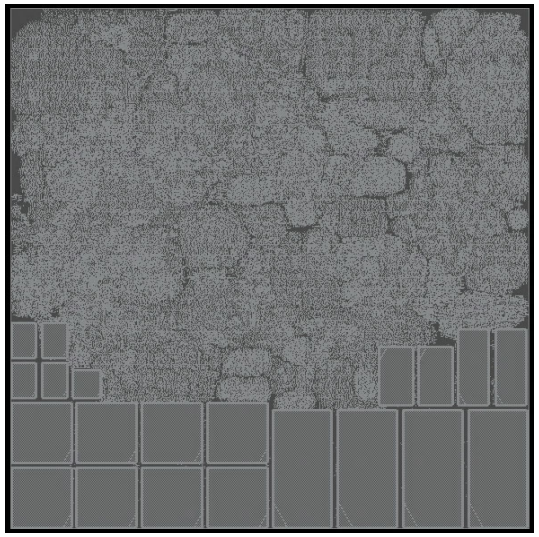
1

true-3D placement

commercial tool not ready

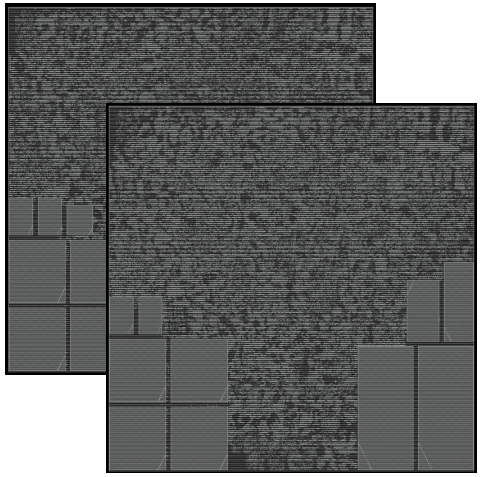
2

pseudo-3D placement



intermediate 2D (commercial tool ready)

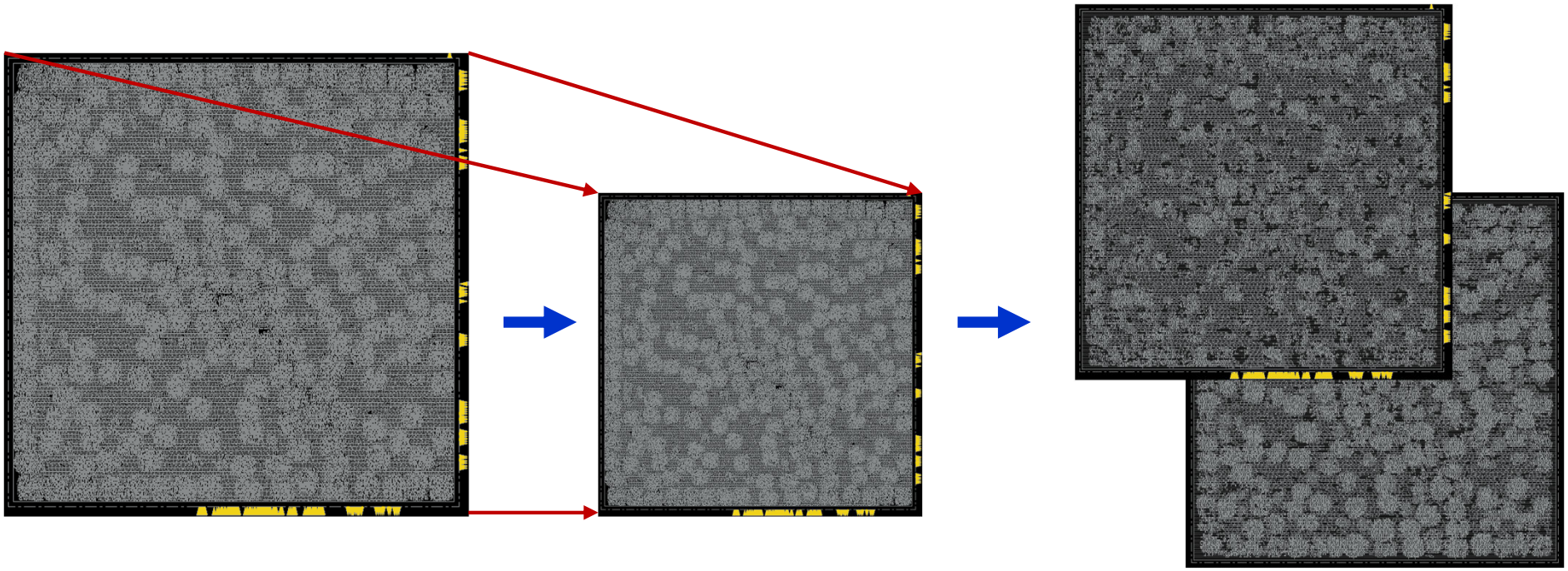
tier partitioning



final 3D

Placement Projection and Partitioning

- **ST28nm, AES**



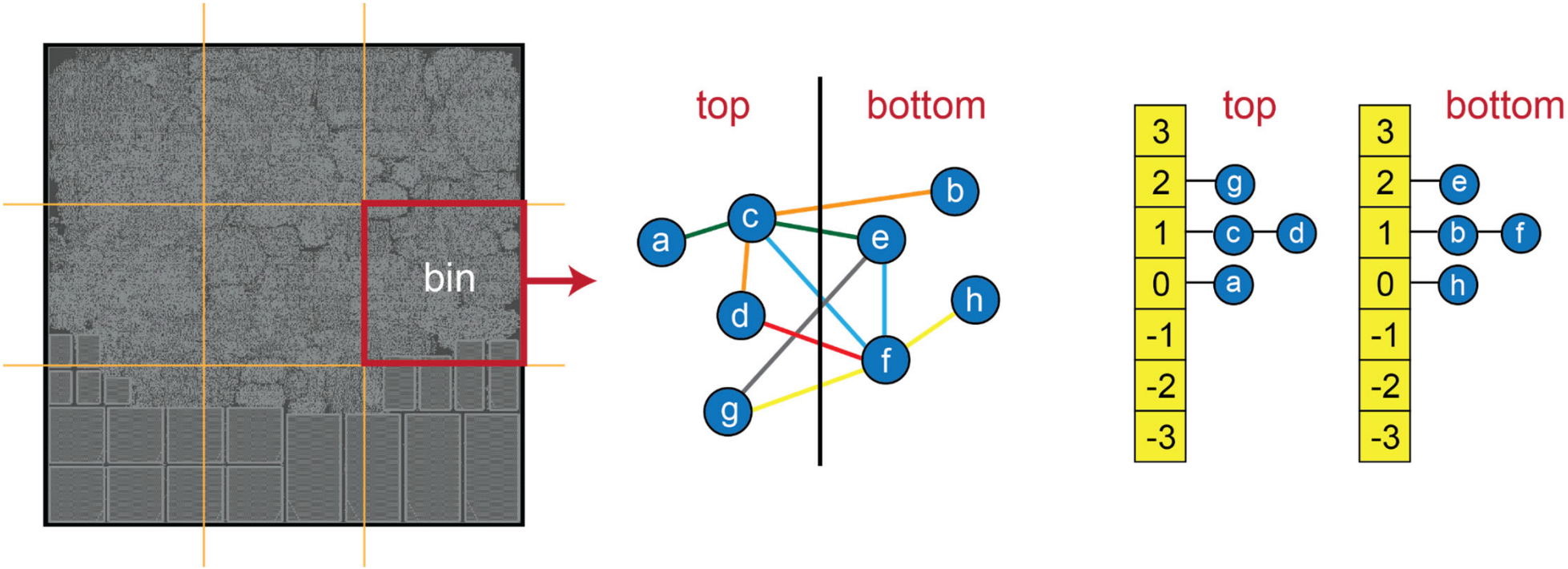
Compact-2D
(interconnect RCs derated by 0.7)

Placement Projection
= 0.7 X x/y-coordinates
(have lots of overlap)

Tier Partitioning
(overlap removed)

Our Automatic Tier Partitioner

- **Bin-based hypergraph partitioning**
  - Divide 2D into bins, and partition each bin
  - Bi-partitioning engine is Fiduccia-Matheyses algorithm [1982]

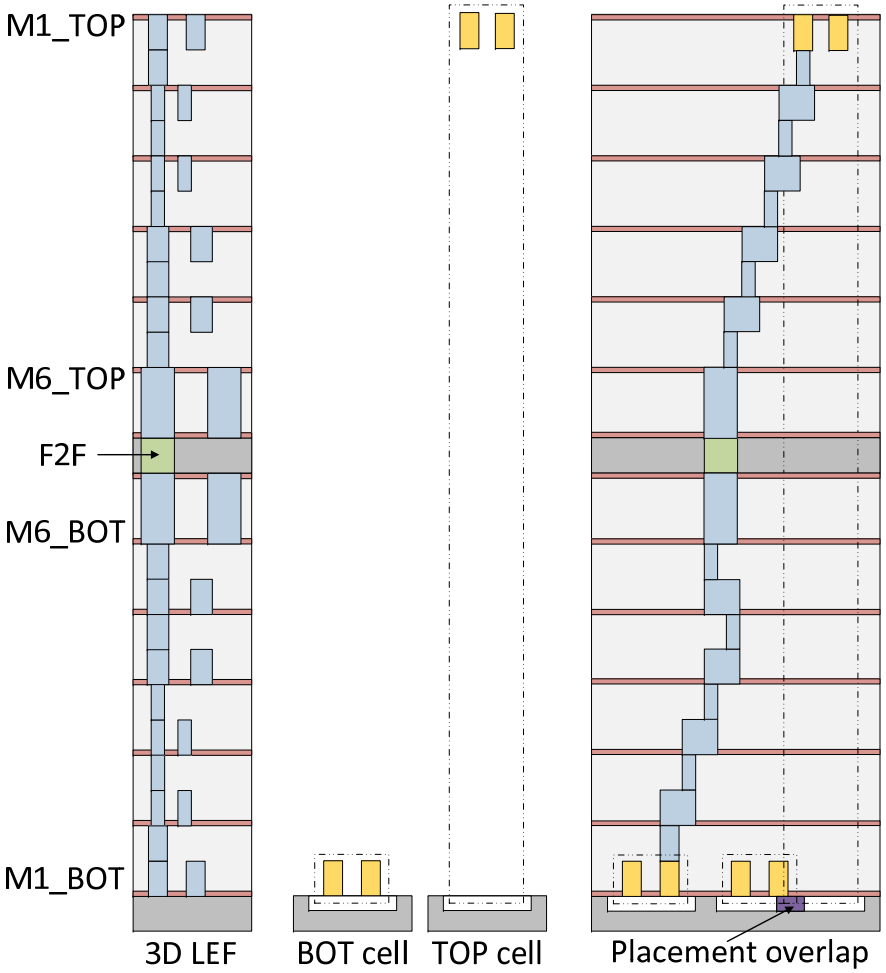


intermediate 2D | hypergraph model | bucket sorting of "gain"

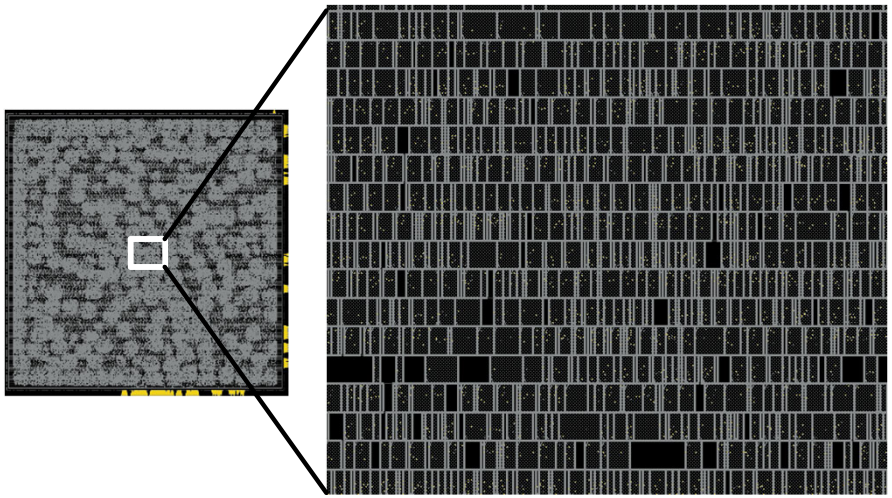
Why binning? Bin size determines MIV usage!!!

Use 2D Optimizer for 3D Timing Closure

- **We still need to use a 2D IC optimizer**



F2F stack-up view

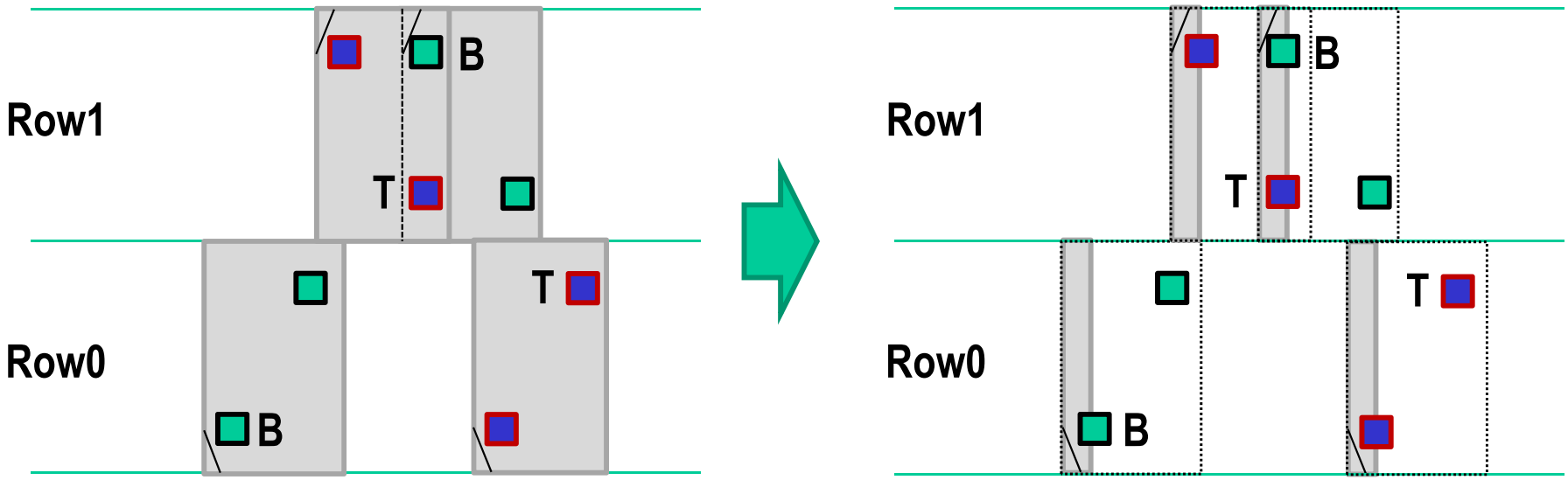


Overlapped Top/Bottom cell placement

Optimization engine will legalize the overlap: placement is DAMAGED!

Our Trick

- **Idea: cell narrowing (site-sized MACRO LEF)**
  - **To temporarily remove overlap just to do timing closure**

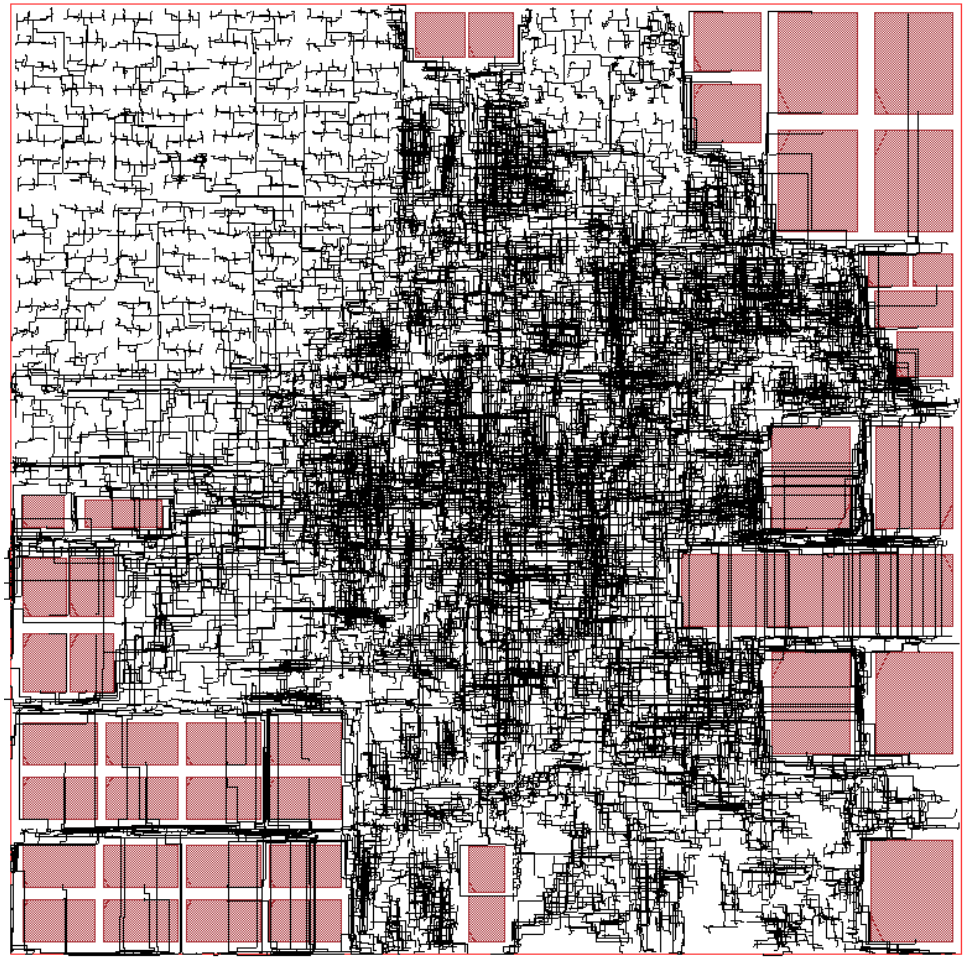


Pins are fine: no overlap
Cells are not fine: overlap

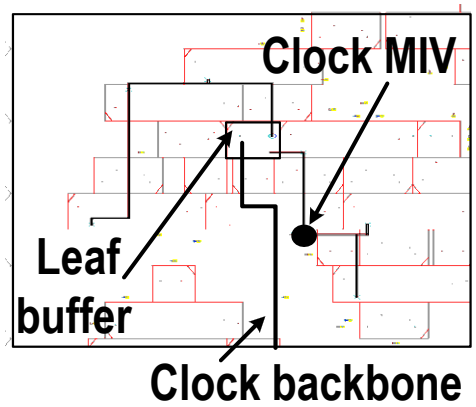
Pins are not overlapping
Cells are not overlapping
Optimization works
And placement is not damaged

Use 2D Clock Router for 3D Clock Tree

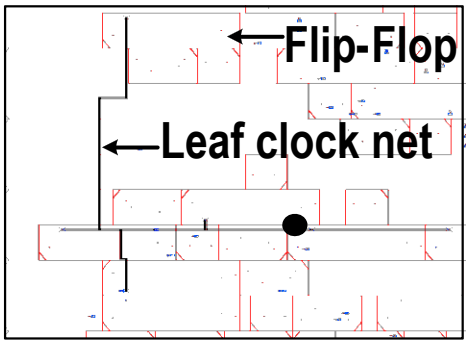
Clock back-bone on tier 0



nanoroute last-level buffer and its FFs



Leaf clock net on tier 0

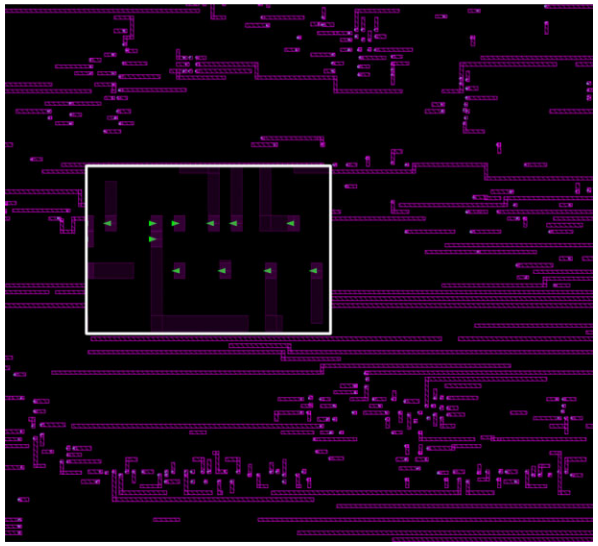


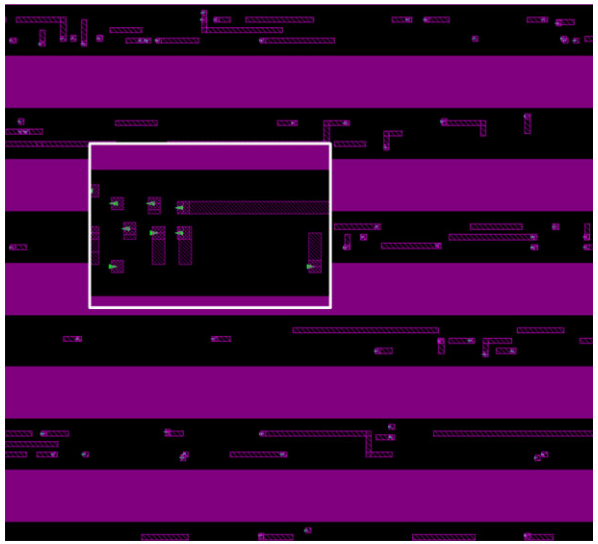
Leaf clock net on tier 1

	Die-by-die	Our method
WL (m)	1.03	0.80 **(-21.67%)**
Power (mW)	68.40	48.00 **(-29.82%)**

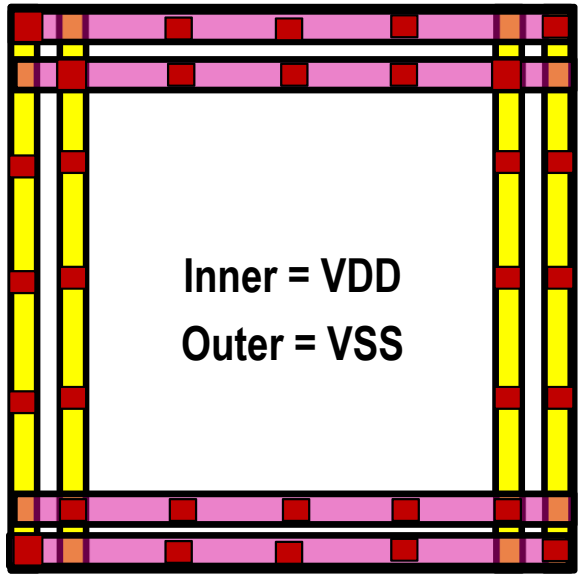
3D PDN Gets In the Way

metal 4



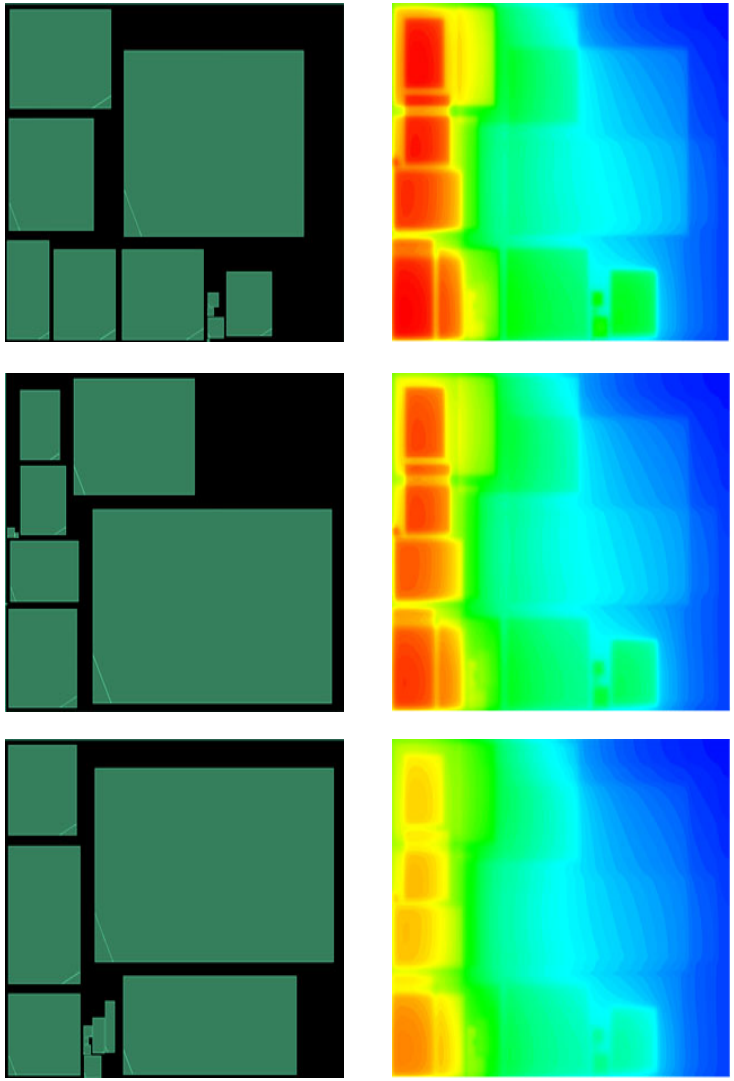


metal 7

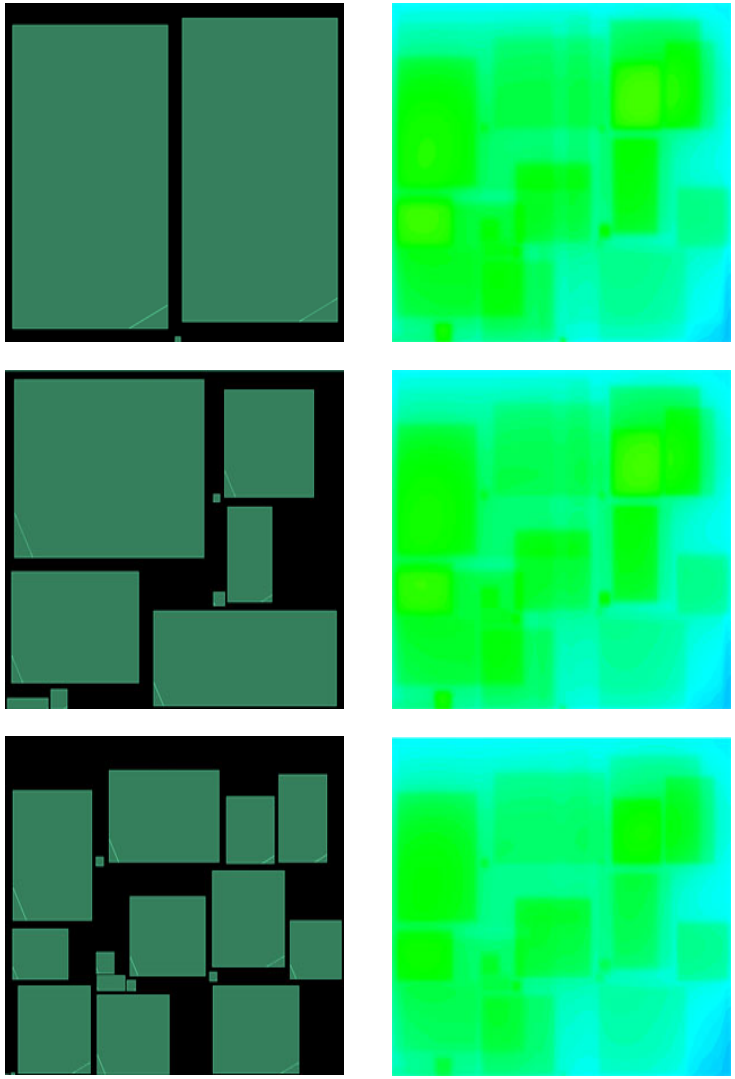


P/G F2F arrays placed along the boundary

Floorplanning Optimization



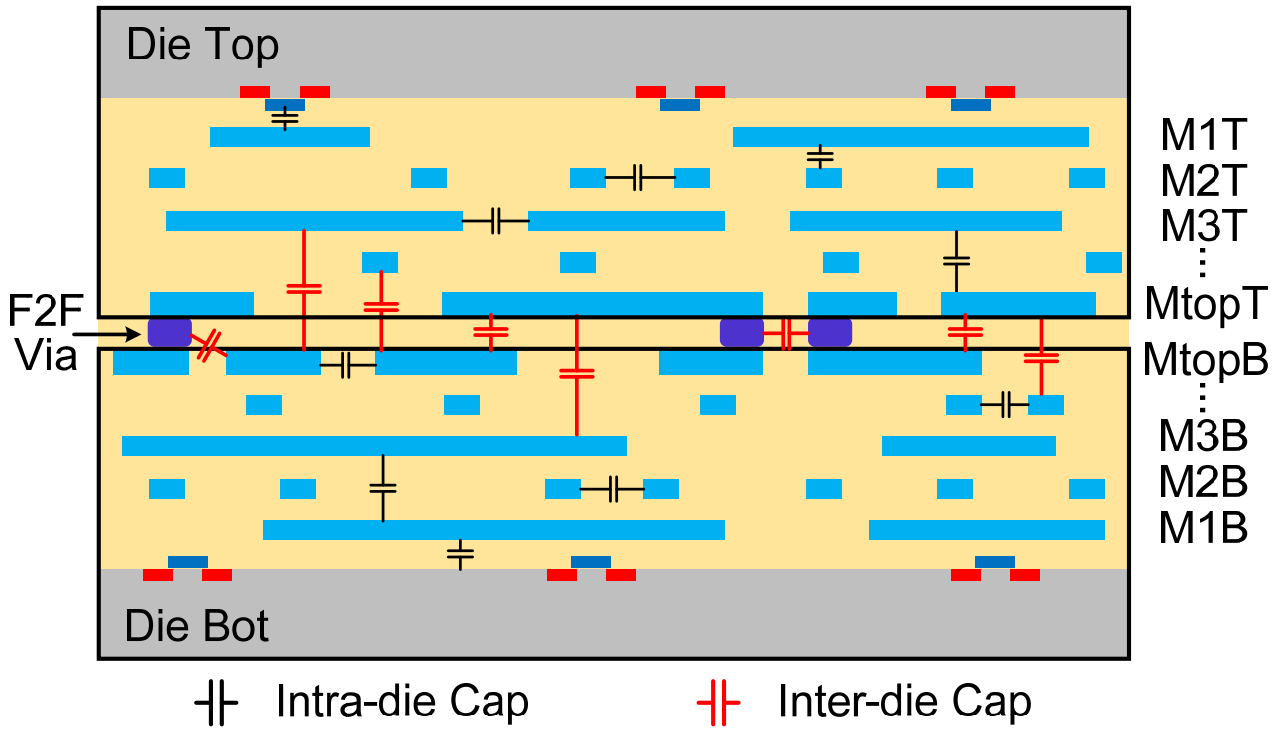
Conventional (Tmax = 60.89C)



Thermal-aware (Tmax = 52.03C)

Coupling Cap in F2F 3D ICs

- **New parasitics**
  - **No commercial tool exists today**
  - **Timing, coupling, noise calculations are WRONG without them**
  - **Magnitude depends on geometries, materials, etc**



Impact on Full-Chip Metrics

- **If you do not extract F2F parasitics**
  - **Calculations become wrong: delay by 6.2%, coupling 32%, noise 26%**

	Extract	Ignore
longest path delay (ns)	3.90	3.66 (-6.2%)
3D net switching power (mW)	1.05	1.01 (-3.5%)
total switching power (mW)	12.1	11.9 (-1.7%)
total coupling cap on 3D nets (fF)	4.37	2.96 (-32%)
total wire cap on 3D nets (fF)	10.8	9.35 (-13%)
ave # of aggressors on 3D nets	285	200 (-30%)
max noise on 3D nets (mV)	41.3	30.4 (-26%)

Concluding Remarks

- **Our best 3D IC results so far**
  - Performance: 37% better than 2D (TSMC 28nm)
  - Power: 39% better than 2D (TSMC 16nm)
  - EDA tools are a key enabler
- **My (long but desperate) wish list**
  - How about better benchmark?
  - How about better EDA?
  - How about better RTL?
  - How about better cells?
  - How about better devices?
  - How about better interconnects?
  - How about better synthesizer?
  - How about better compiler/OS?