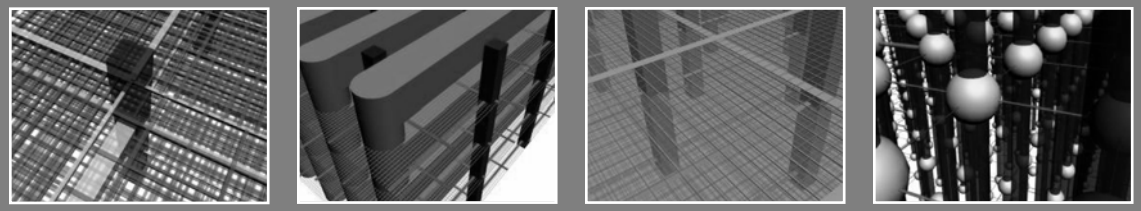
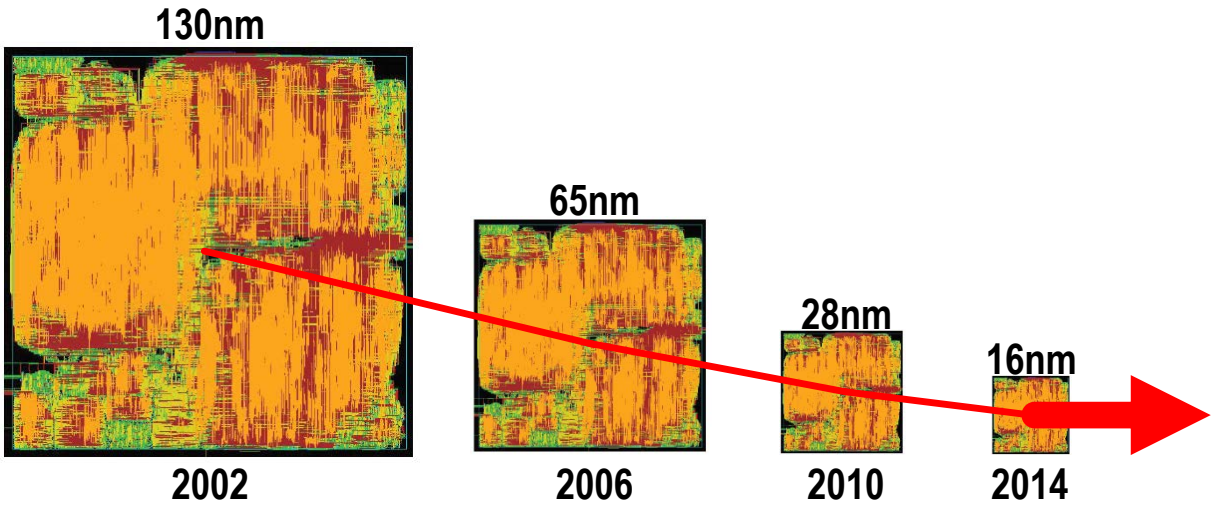
Thermal Analysis on Face-to-Face(F2F)-bonded 3D ICs



Kyungwook Chang, Sung-Kyu Lim
School of Electrical and Computer Engineering
Georgia Institute of Technology

Challenges in 2D Device Scaling



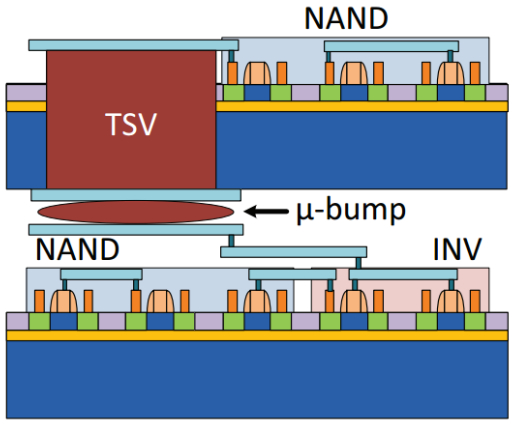
challenges

- channel length scaling
- increasing contact resistivity
- lithography limitation
- increasing wire resistance
- higher manufacturing cost

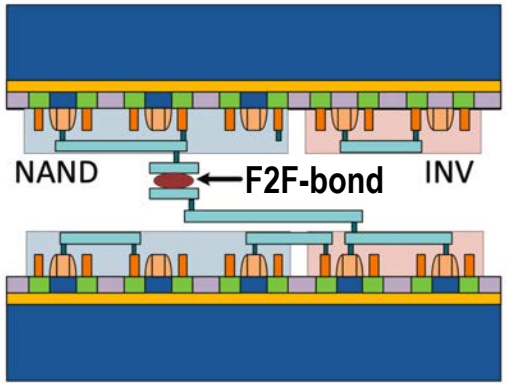
How about placing cells in 3D space?

Three-dimensional (3D) ICs

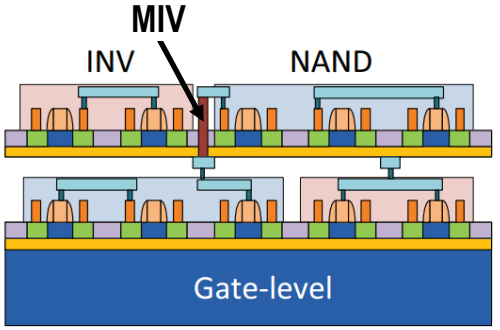
- **Wafers are stacked vertically**
  - **Benefits from short vertical connections**
- **Three flavors of 3D ICs**



TSV-based 3D IC[1]



F2F-bonded 3D IC[1]



Monolithic 3D IC[1]

easier and reliable vertical connection

higher vertical integration density

[1] S. Panth et al., "Design and CAD Methodologies for Low Power Gate-level Monolithic 3D ICs," IEEE International Symposium on Low Power Electronics and Design, 2014

Challenges in 3D ICs

- **Tools**
  - No EDA tools supporting 3D cell placement

    Requires tricks to place cells in 3D space

- **Heat dissipation**
  - Higher temperature due to increased power density

    Requires thermal-aware packaging and 3D cell placement

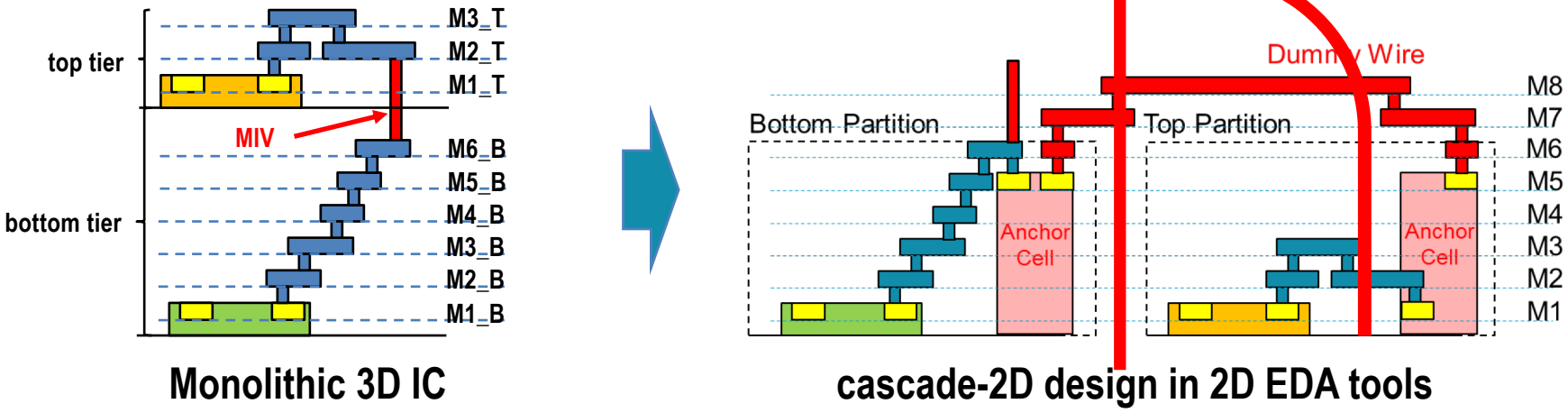
Cascade-2D Flow

- How to place cells in 3D space?

  Mimic vertical connections

  with sets of ”anchor cells” and “dummy wires”

- Implement two tiers simultaneously in a single emulated 2D design (cascade-2D design)

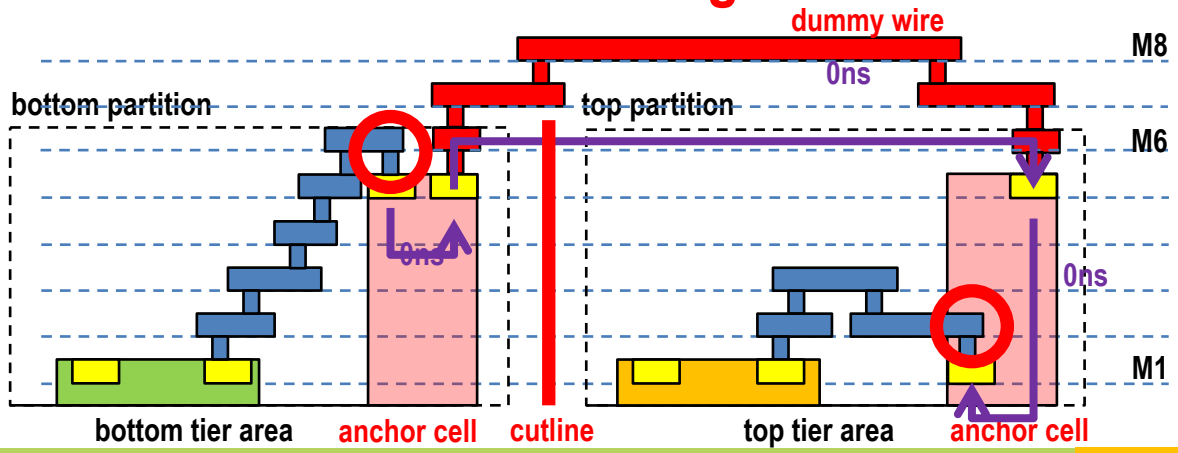


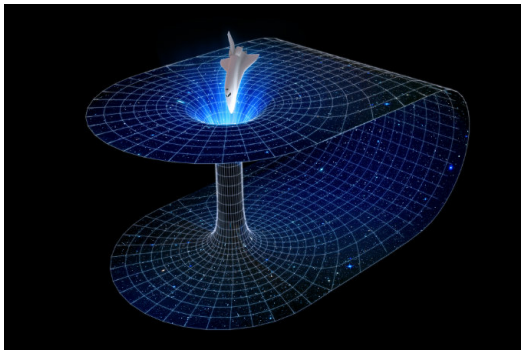
[2] K. Chang et al., "Cascade2D: A Design-Aware Partitioning Approach to Monolithic 3D IC with 2D Commercial Tools," IEEE International Conference on Computer-Aided Design, 2016

Cascade-2D Flow Details

1. **Design-aware partitioning (cell z-location)**
   - Perform block-level partitioning, maximizing timing paths crossing tiers
2. **Vertical connection planning (vertical connection location)**
   - Determine vertical connection location, minimizing wire-length btw blocks
3. **Cascade-2D design implementation (cell x-y location & routing)**
   - Implement the emulated 2D version of an M3D IC

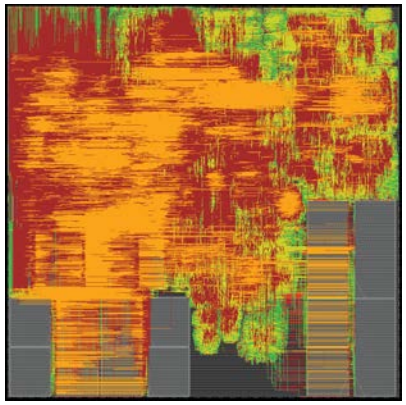
Anchor cells and dummy wires act as a wormhole emulating vertical connections



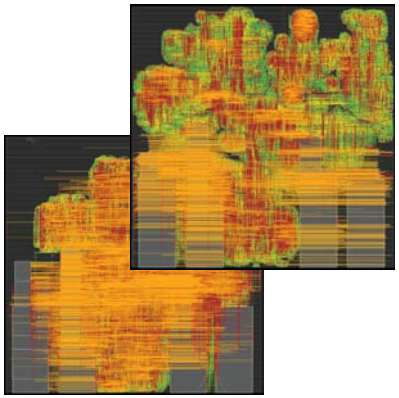


Iso-performance Power Comparison

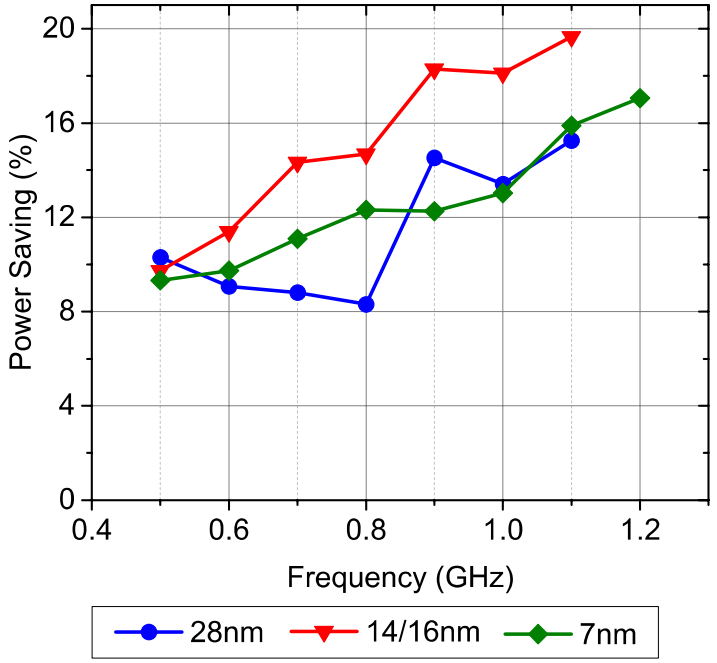
- Design: Arm® Cortex®-A7 core
- Technology: Foundry 28nm, 14/16nm, and predictive 7nm technology
- 3D IC flavor: Monolithic 3D IC (M3D)



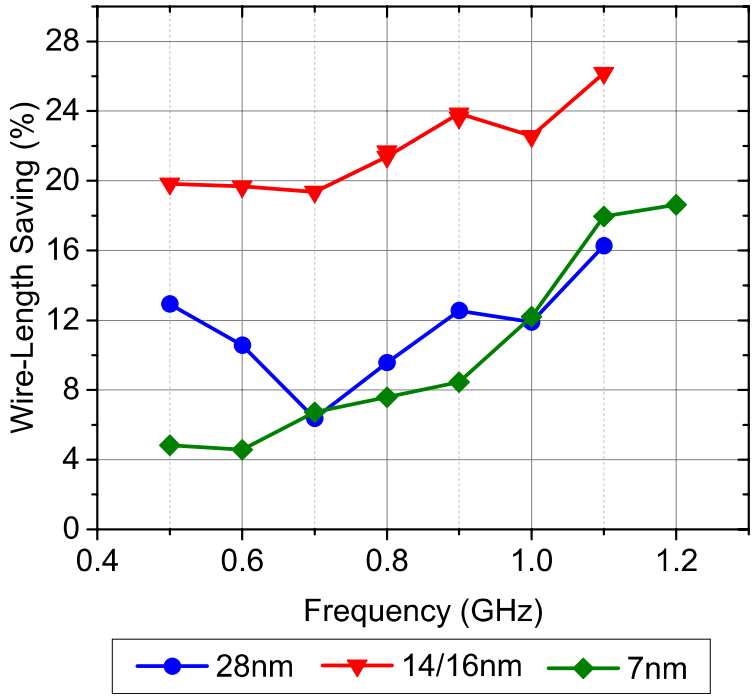
28nm 2D

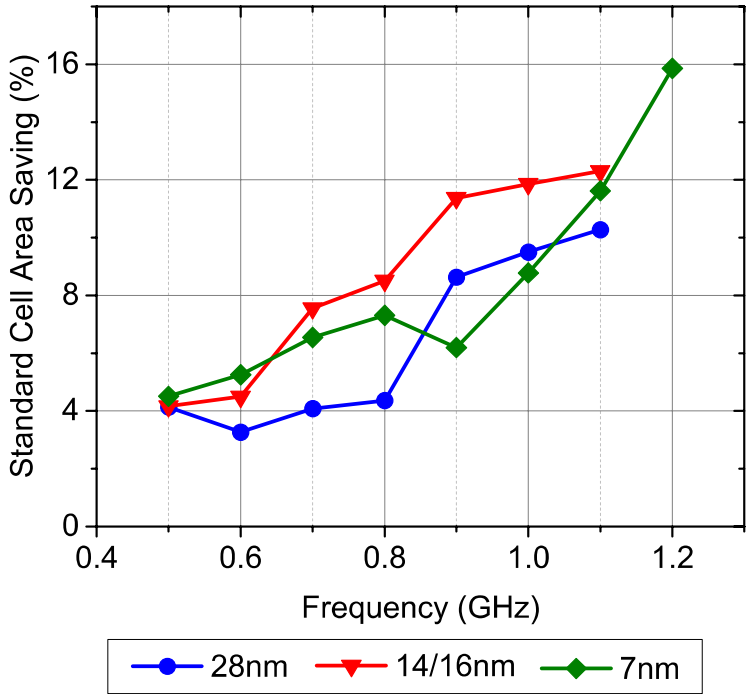


28nm M3D



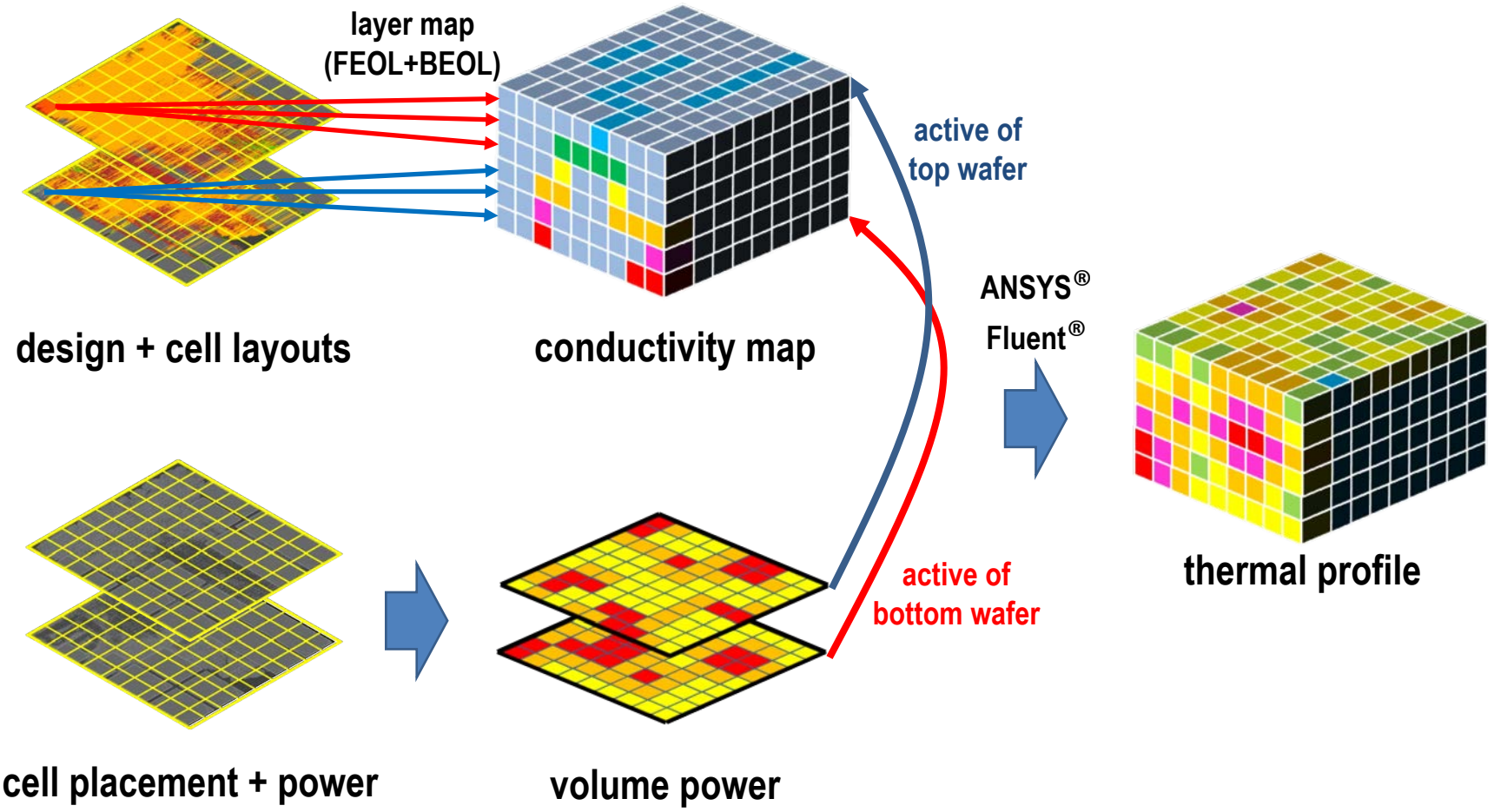
Source of Power Savings



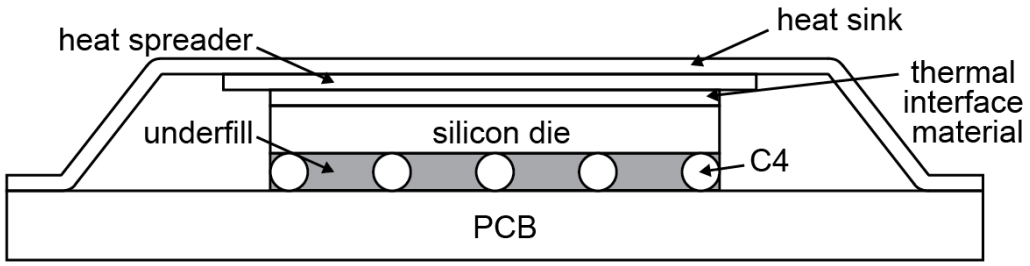


3D IC offers both wire-length and standard cell area savings

Thermal Analysis Methodology

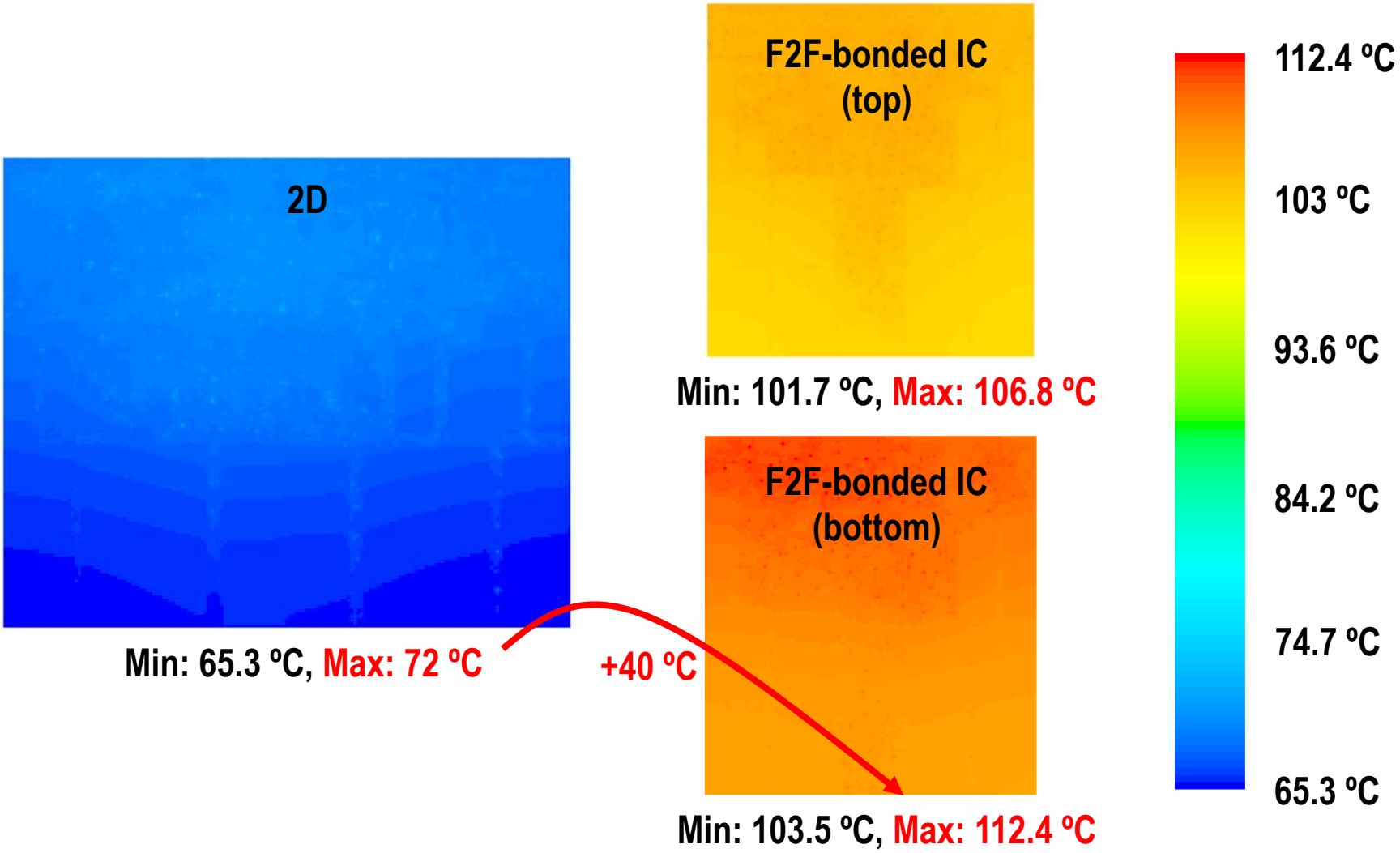


Assumed Thermal Stack Configuration

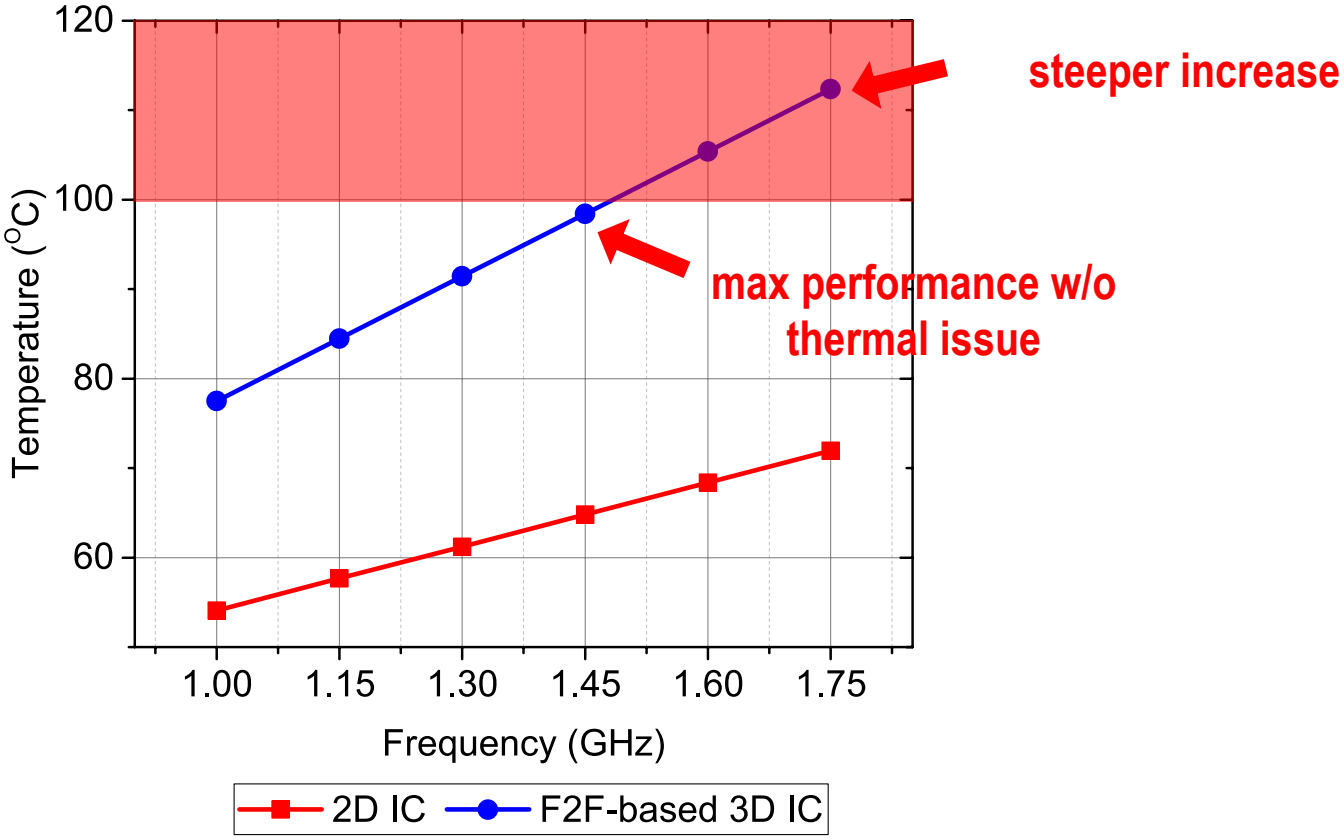


	thickness (um)	material
heat sink	1,000	Cu
heat spreader	100	Cu
TIM	25	Grease
substrate	45	Si
underfill	25	SiO_2
C4	25	Cu
PCB	1,200	PCB

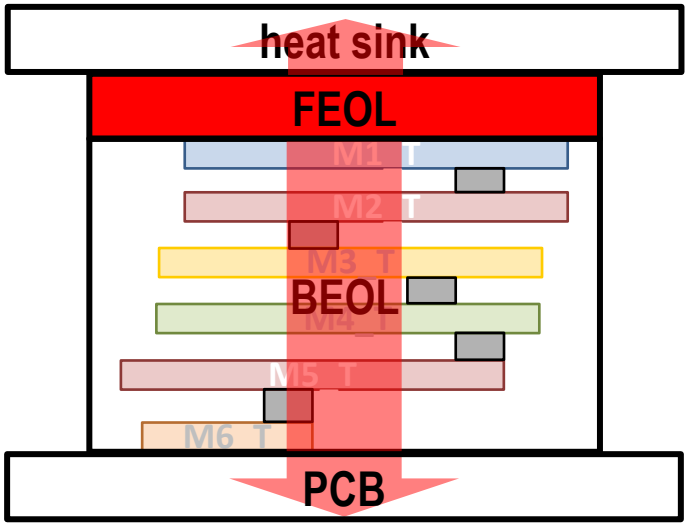
Temperature Comparison: 2D vs. F2F-bonded 3D IC



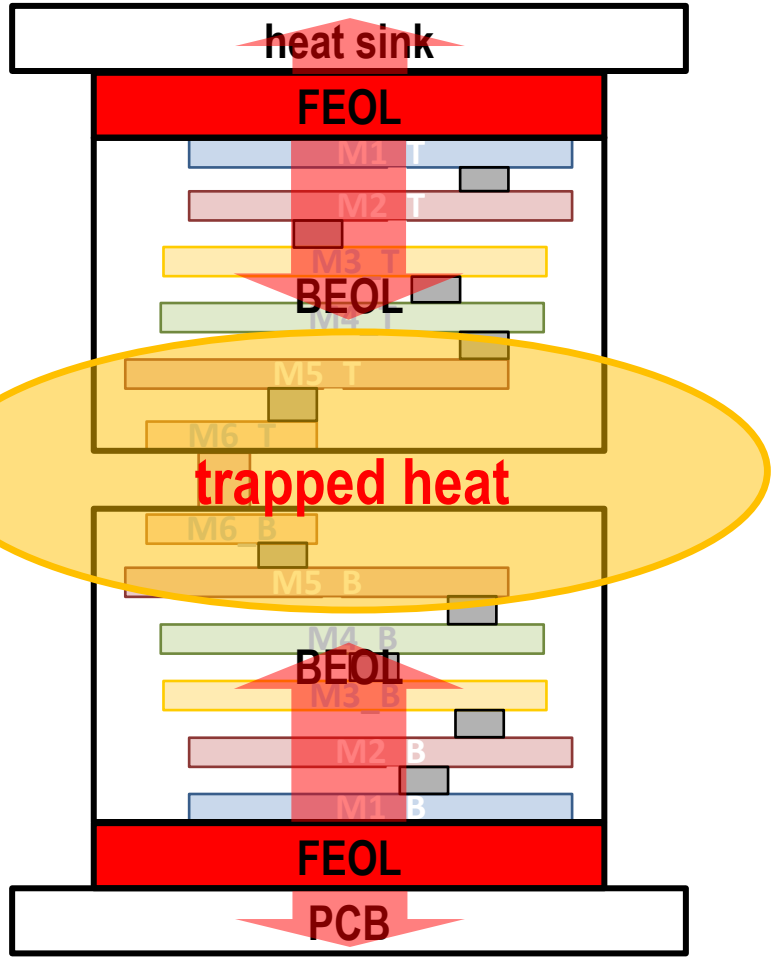
Impact of Frequency on Temperature



Thermal Issue of F2F-bonded 3D ICs

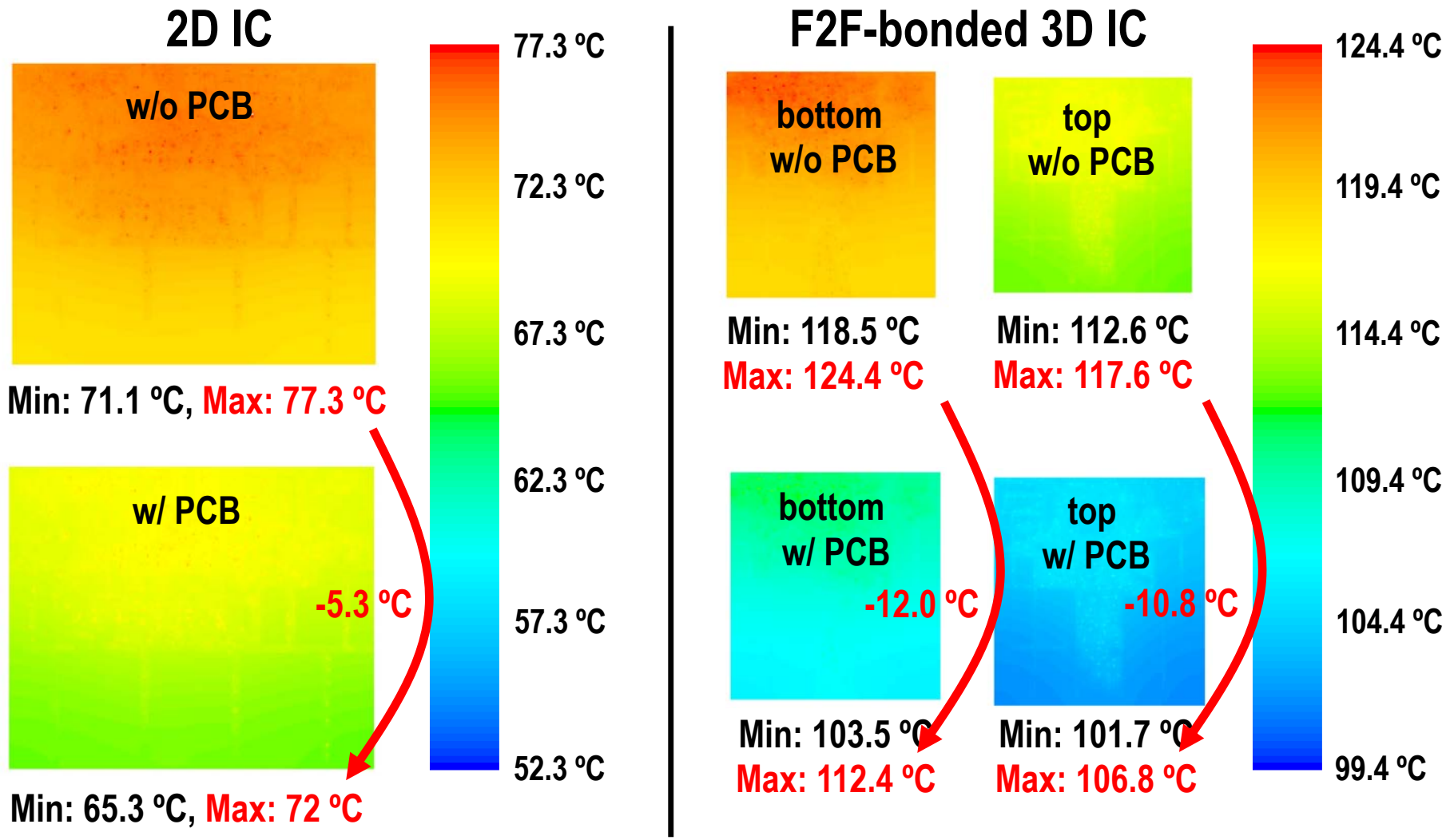


2D IC



F2F-bonded 3D IC

Impact of Packaging on Temperature



Packaging affects more in F2F-bonded 3D ICs

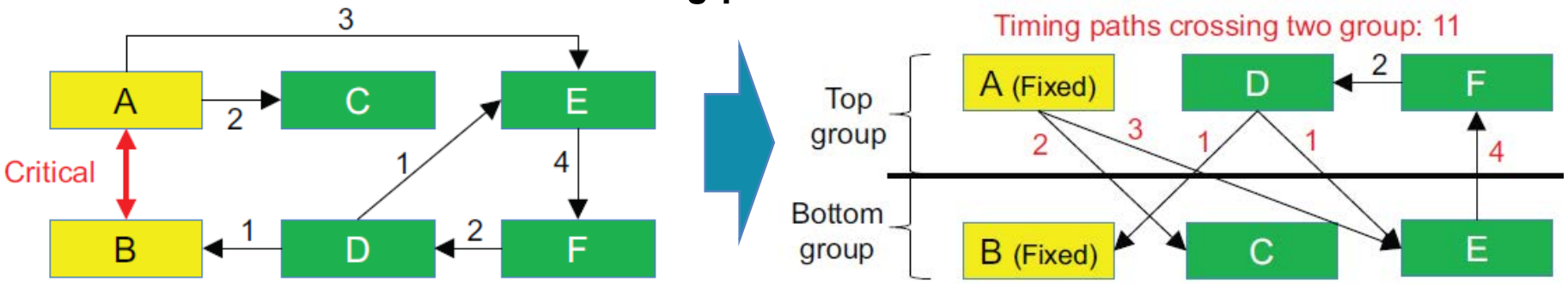
Future Works

- Impact of packaging
  - Find thermal-aware packaging for F2F-bonded 3D ICs
  - How to reduce trapped heat in the middle?
- Thermal-aware cell partitioning
  - Place power hungry cells on top tier
- Impact of power-delivery network
  - Utilize TSVs for better conduction to PCB
- Commercial tool
  - Use commercial EDA tools for accurate thermal analysis

Thank You

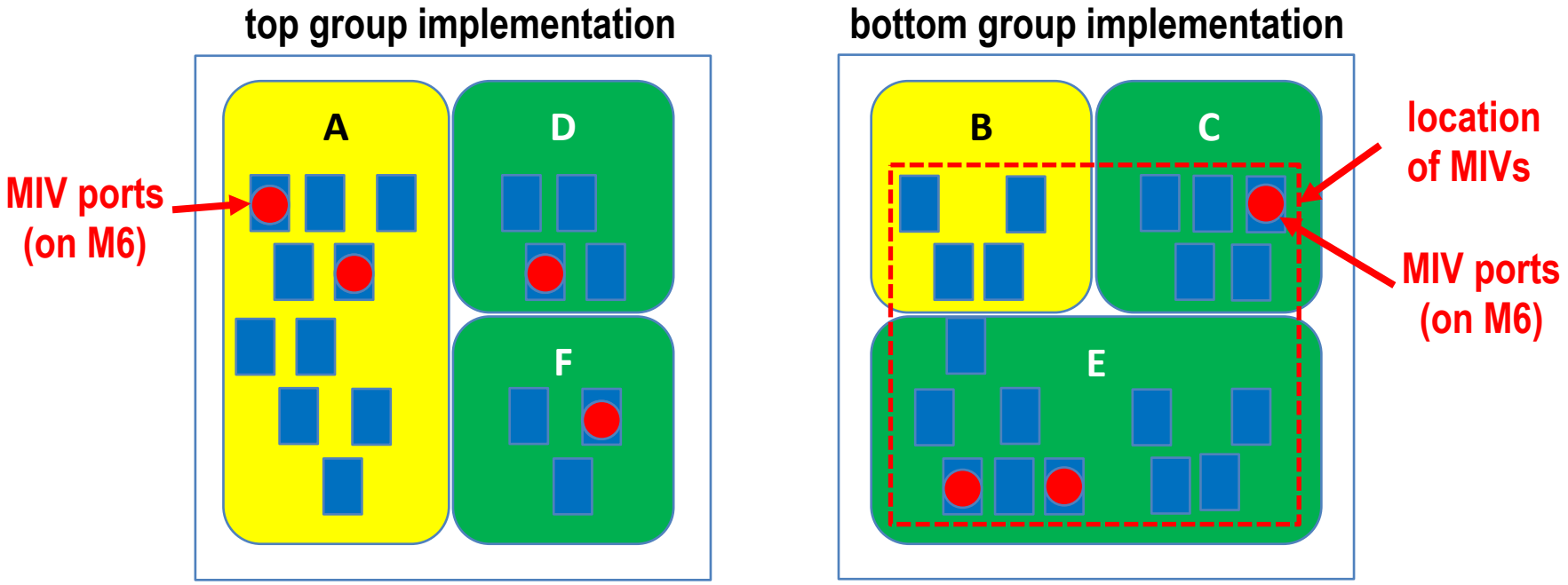
Design-Aware Partitioning Stage (cell z loc.)

- **Partition functional blocks into two groups**
- **Microarchitecture organization** ■
  - → **Floorplan timing-critical functional blocks on different tiers**
- **Information from 2D implementation** ■
  - **Functional block area**
  - **The number of timing paths crossing each pair of functional blocks**
- **Partition functional blocks into two groups**
  - **Balance area**
  - **Maximize the number of timing-path crossing two groups**



MIV Planning Stage (MIV loc.)

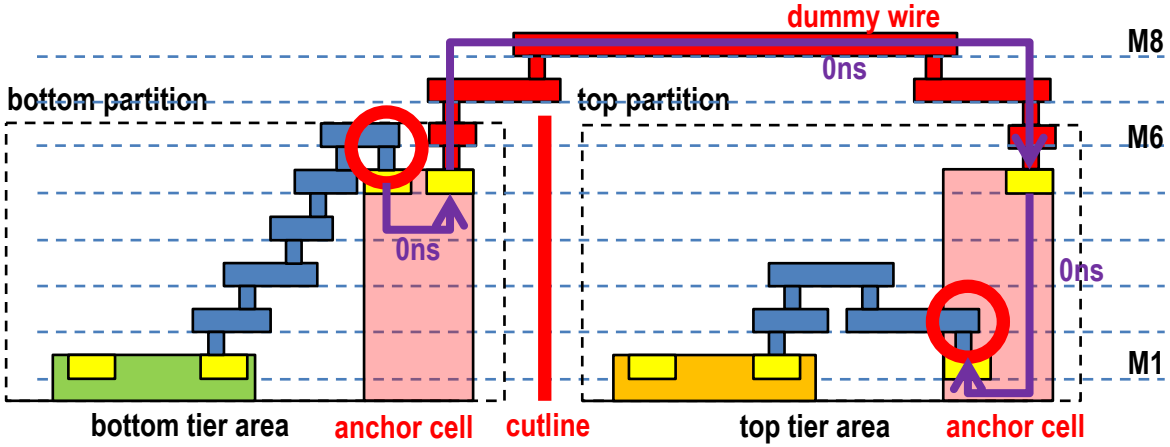
- **Minimize the distance** between functional blocks with large number of timing paths

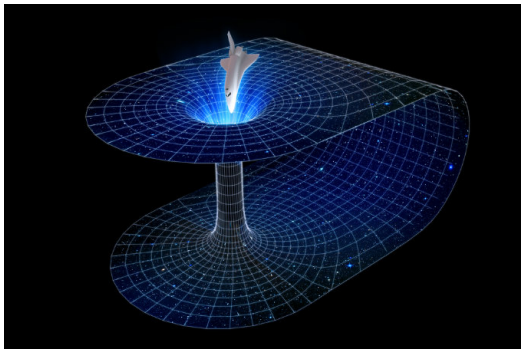


Cascade-2D Stage (cell x-y loc. & routing)

- **Dummy wires**
  - Connect MIV port pairs in two partitions
- **Anchor cells**
  - → Connect MIV ports to M1 in top partition, M6 in bottom partition
- **Perform place & route with hard partition between tiers**

Anchor cells and dummy wires act as a wormhole emulating MIVs



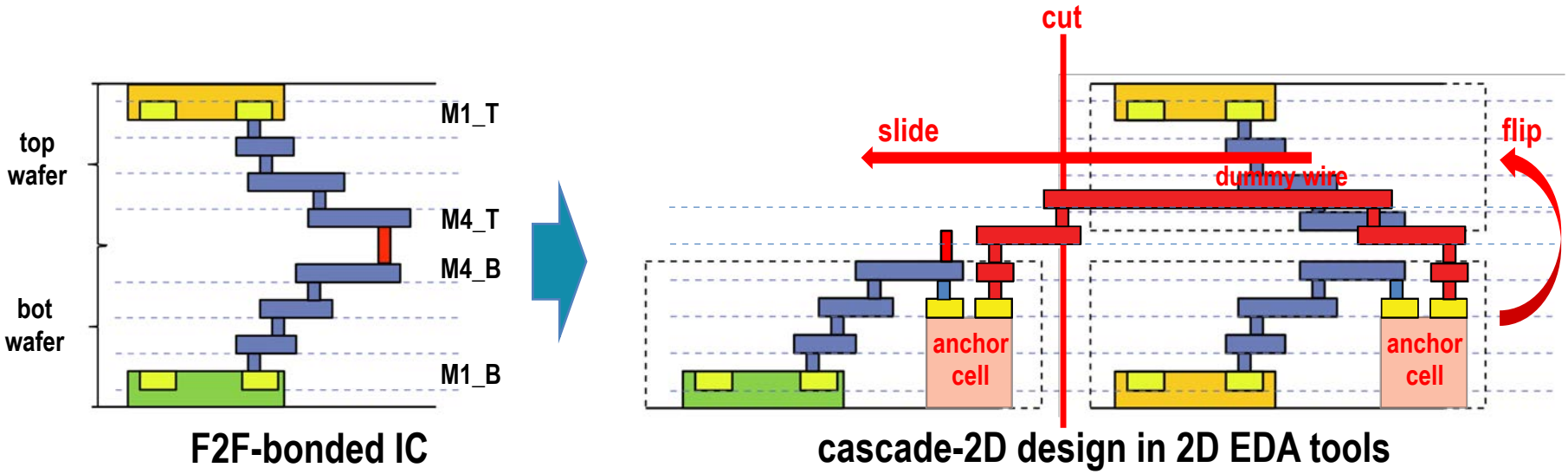


Cascade-2D Flow

- No EDA tools supporting 3D cell placement

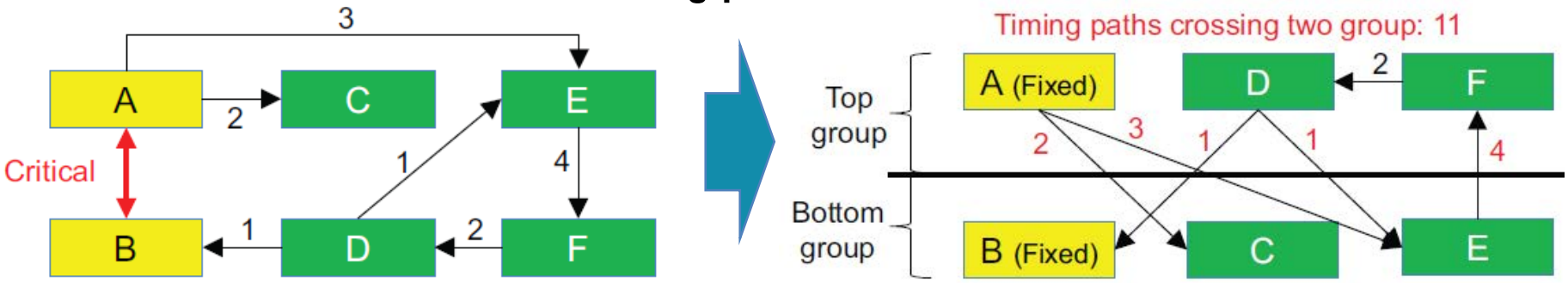
  Use tricks to place cells in 3D space with 2D tools

- Implement two wafers simultaneously in a single emulated 2D design (cascade-2D design)



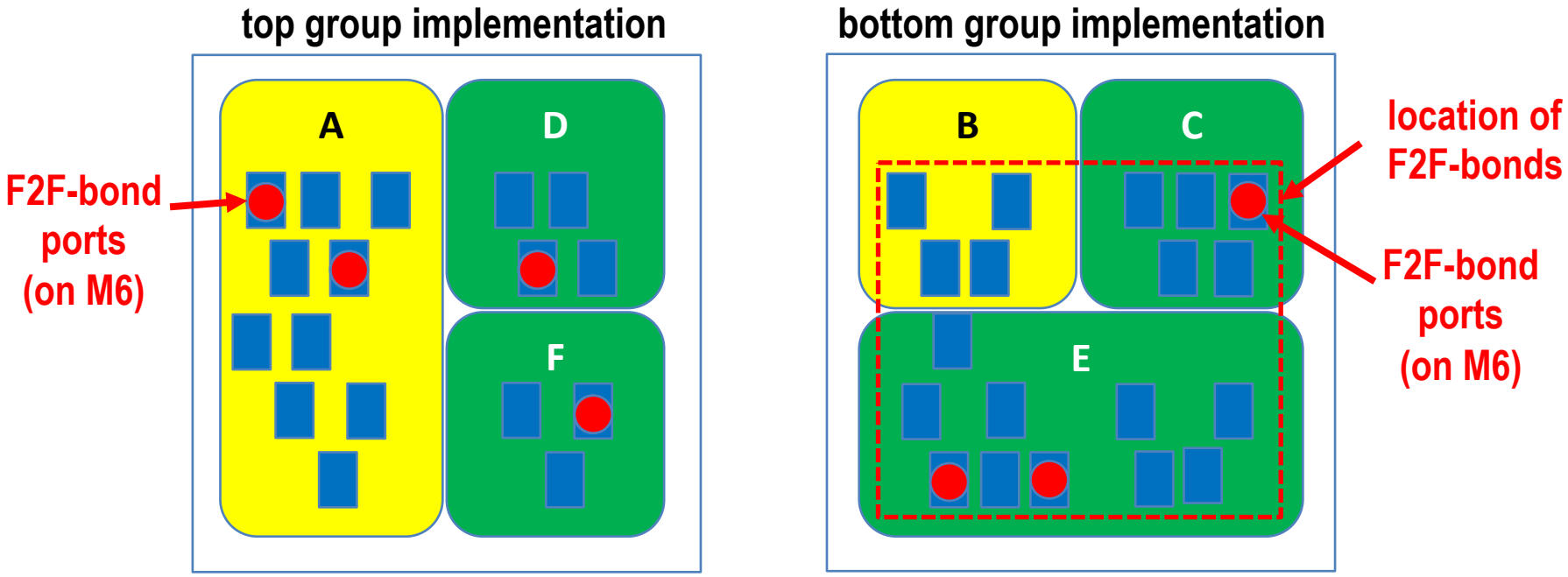
Design-Aware Partitioning Stage (cell z loc.)

- **Partition functional blocks into two groups**
- **Microarchitecture organization**
  - → **Floorplan timing-critical functional blocks on different tiers**
- **Information from 2D implementation**
  - **Functional block area**
  - **The number of timing paths crossing each pair of functional blocks**
- **Partition functional blocks into two groups**
  - **Balance area**
  - **Maximize the number of timing-path crossing two groups**



F2F-bonds Planning Stage (F2F-bond loc.)

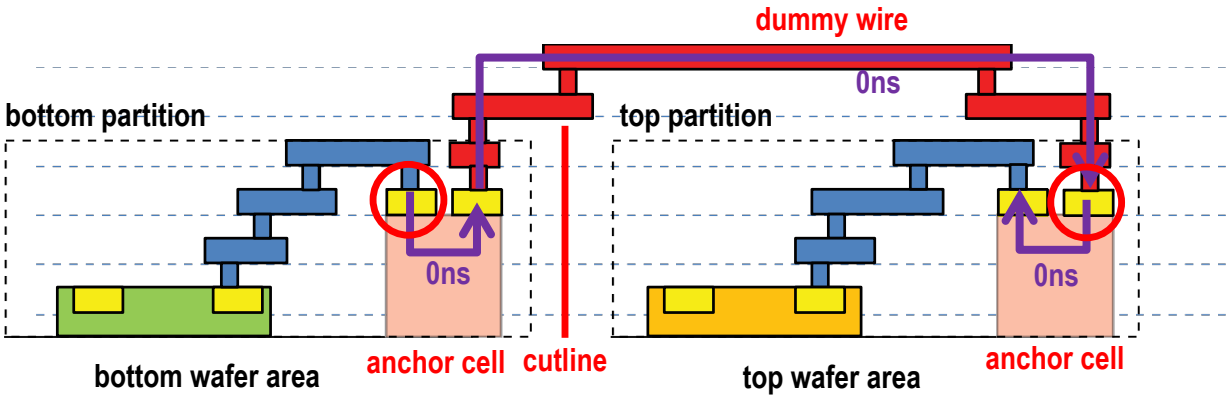
- **Minimize the distance** between functional blocks with large number of timing paths

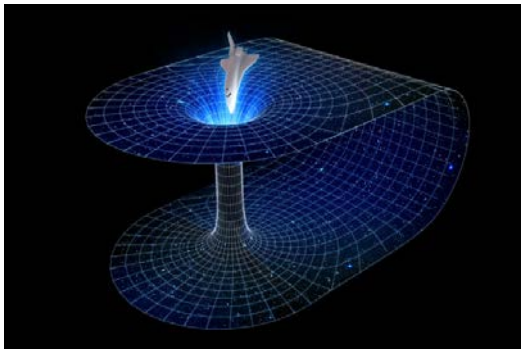


Cascade-2D Stage (cell x-y loc. & routing)

- **Dummy wires**
  - Connect F2F-bond port pairs in two partitions
- **Anchor cells**
  - → Connect F2F-bond ports to M6 of each partition
- **Perform place & route with hard partition between tiers**

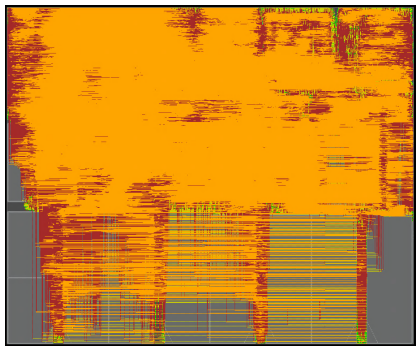
Anchor cells and dummy wires act as a wormhole emulating MIVs



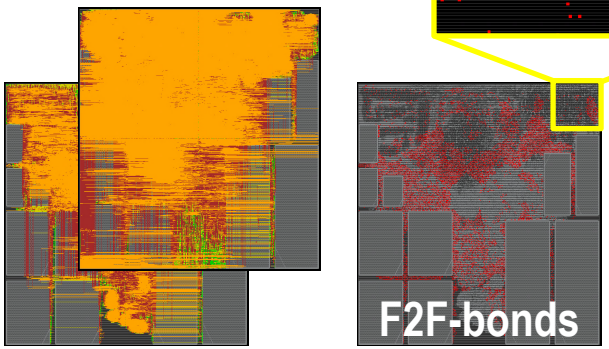


Experimental Setup

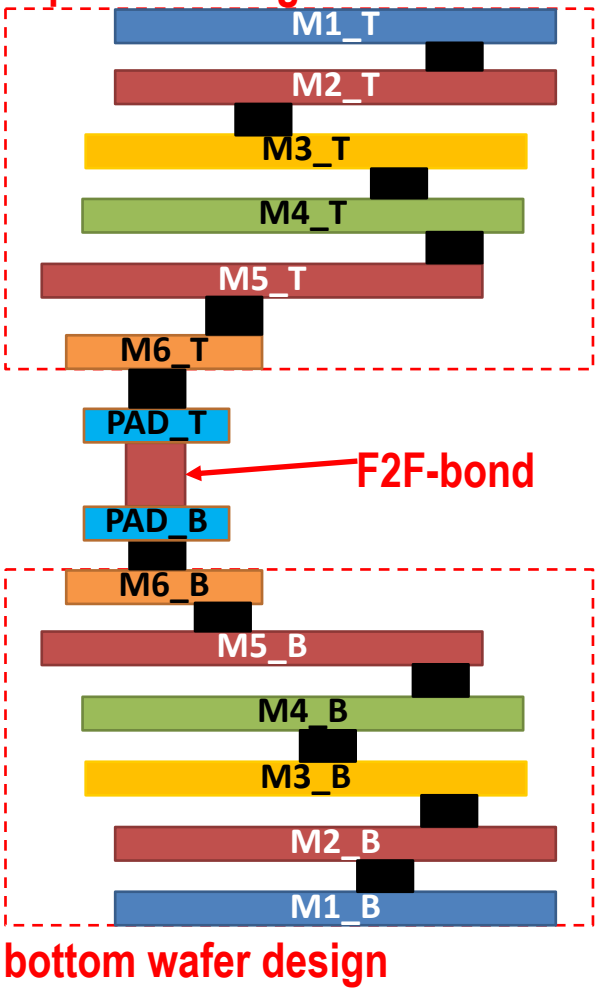
- Design: Arm® Cortex®-A7 core
- Technology: CLN16FCLL ULVT
- Target Frequency:
- F2F-bonds size/pitch: 0.5um/1um



2D IC



F2F-bonded 3D IC

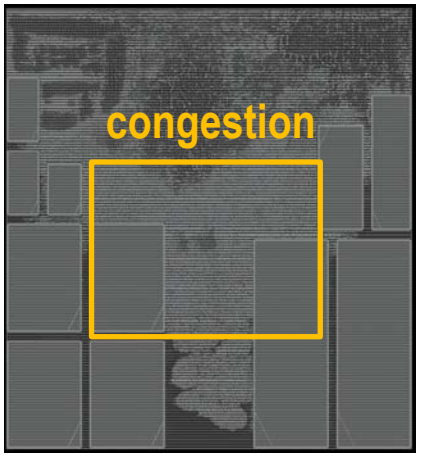


Design / Power Metric Comparison

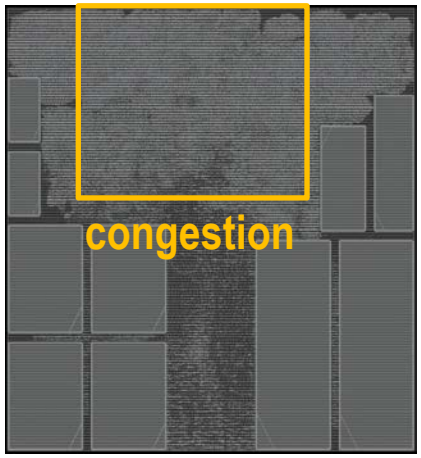
metrics	2D IC	F2F-bonded 3D IC	Δ%
target freq (GHz)	1.75	1.75	0.0
footprint (um)	523x433	323x350	-50.1
cell count	199,765	191,333	-4.2
std. cell area (um^2)	69,009	68,208	-1.2
density (%)	55.0	53.6	-2.5
wire-length (um)	3,025,229	2,885,312	-4.6
pin cap (pF)	347.6	328.0	-5.6
wire cap (pF)	493.0	509.2	3.3
total cap (pF)	840.6	837.1	-0.4
sw power (mW)	138.3	135.9	-1.7
int power (mW)	153.9	150.4	-2.3
lkg power (mW)	27.0	26.2	-3.0
tot power (mW)	319.1	312.5	-2.1

Reason for Low Power Saving

- Slow memory
  - Almost all critical paths start from memory blocks
  - Standard cell area reduction happens only on those paths
- Memory floorplan
- Non-optimal F2F-bond placement

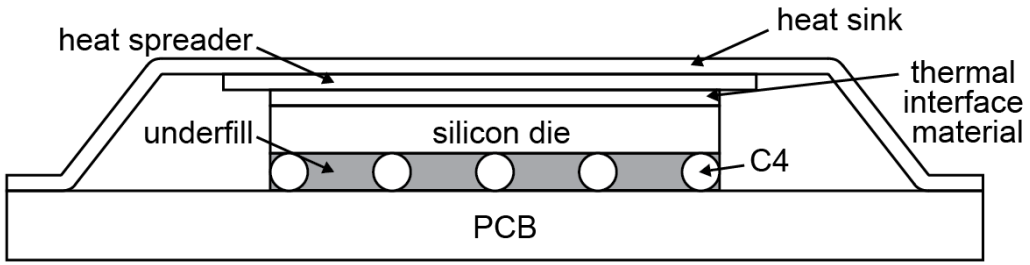


bottom wafer



top wafer

Assumed Thermal Stack Configuration



	width/height	thickness (um)	material	K (W/m-K)
heat sink	wafer width/height * 2	1,000	Cu	401
heat spreader	wafer width/height * 1.5	100	Cu	401
TIM	wafer width/height	25	Grease	5
substrate	wafer width/height	45	Si	149
underfill	wafer width/height	25	SiO_2	1.38
C4	25um (pitch 50um)	25	Cu	401
PCB	wafer width/height *2	1,200	PCB	4.5(V) / 60(L)