SAINT-S: 3D SRAM Stacking Solution based on 7nm TSV technology

Kyoungsun Cho, Jinhong Park, Billy Koo, Sunkyoung Seo, Yoonjae Hwang, Sungcheol Park and Mijung Noh
Samsung Electronics

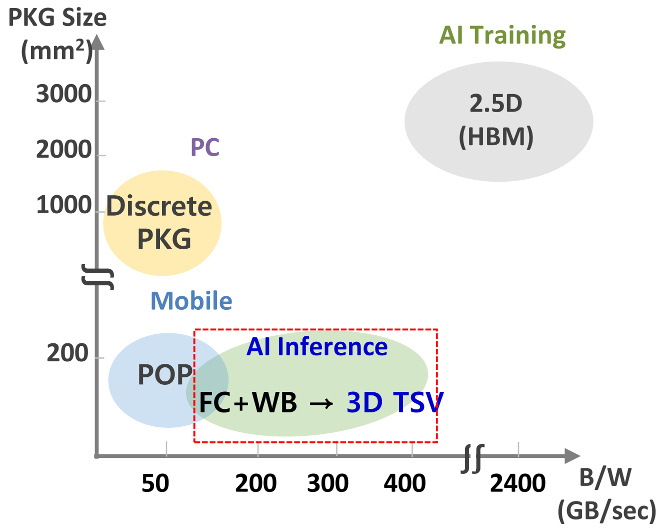
1 July 2020

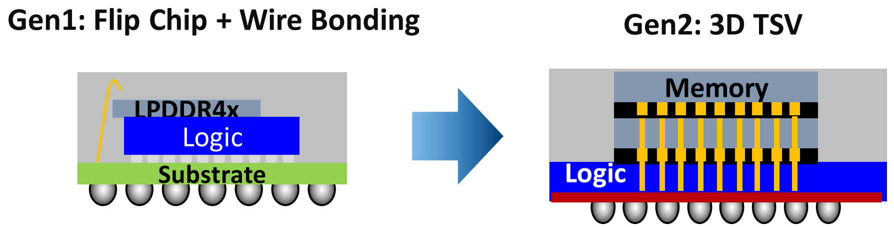
[Abstract]

- Data movement of emerging devices hit the memory wall. AI devices require higher memory bandwidth and higher memory capacity. AR/VR devices also require the lowest latency possible. To fulfill the memory requirements, Samsung Foundry proposed a three-dimension (3D) static random access memory (SRAM) stacking solution (SAINT-S, Samsung Advanced INterconnection Technology with SRAM). The solution is implemented by Samsung 7LPP process and leverages stacking technology using TSV (through silicon via) to achieve high bandwidth and low latency interface between logic die and SRAM die, and small form factor. Implementation results present read/write memory latency of 7.2/2.6 ns and 24.3 GB/s memory bandwidth with 0.156W average power consumption per channel at 760MHz frequency. The memory bandwidth per power of SAINT-S is 6.2x higher than GDDR6's and 2.2x higher than HBM2e.

[Motivation]

- **AI devices require higher memory bandwidth and low latency with small form factor.**





Higher B/W -> Increase I/O Density

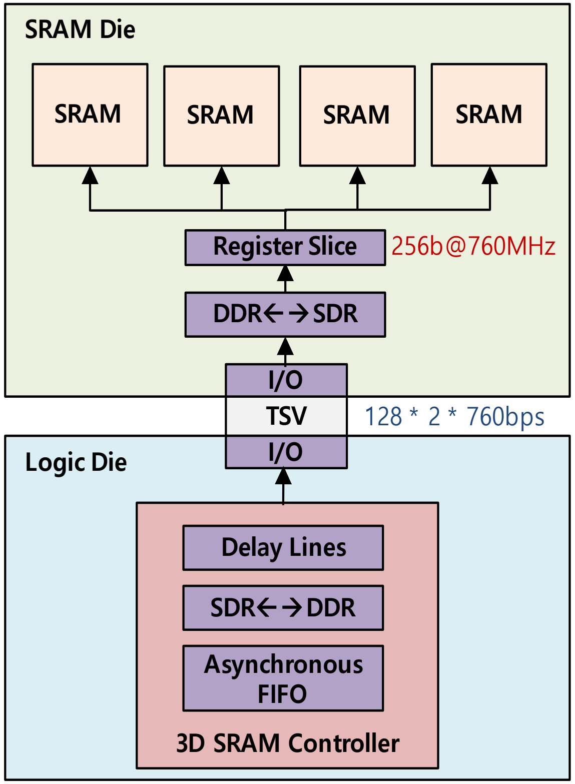
Low latency -> Embedded SRAM

Need 3D SRAM solution

- **To fulfill the memory requirements, Samsung Foundry proposed a three-dimension (3D) static random access memory (SRAM) stacking solution (SAINT-S, Samsung Advanced INterconnection Technology with SRAM).**

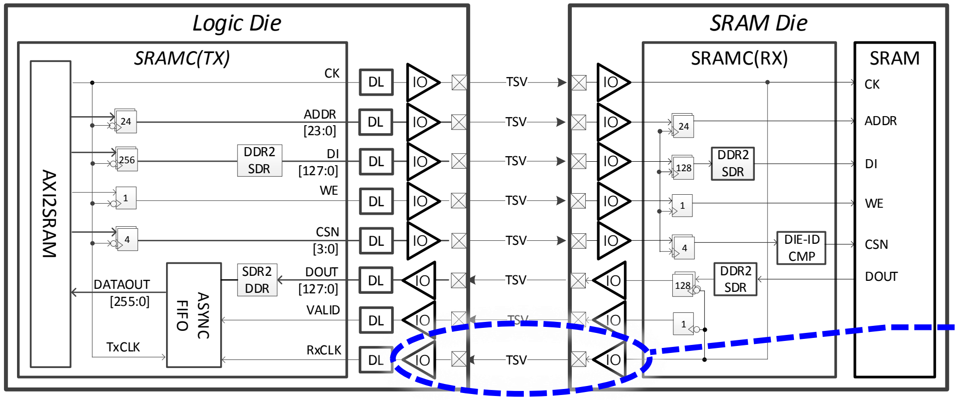
[Architecture]

- **Logic die and SRAM die are combined using TSV**
- **Logic Die includes custom 3D SRAM controllers as well as CPU and DMA to verify performance and power**
- **3D SRAM Controller**
  - Access to SRAM (64Mbits) on SRAM die through TSV
    - 256bit @ 760MHz per channel
  - Source synchronous interface and asynchronous FIFO
  - Double data rate (DDR) conversion to reduce the number of signals
  - Controllable delay lines to compensate the clock to data and data to data skew

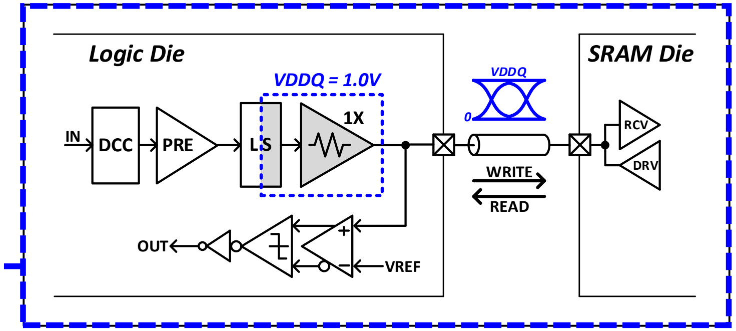


[Horizontal Design View]

Source Synchronous DDR Interface



TSV IO



Low Power Features for TSV IO

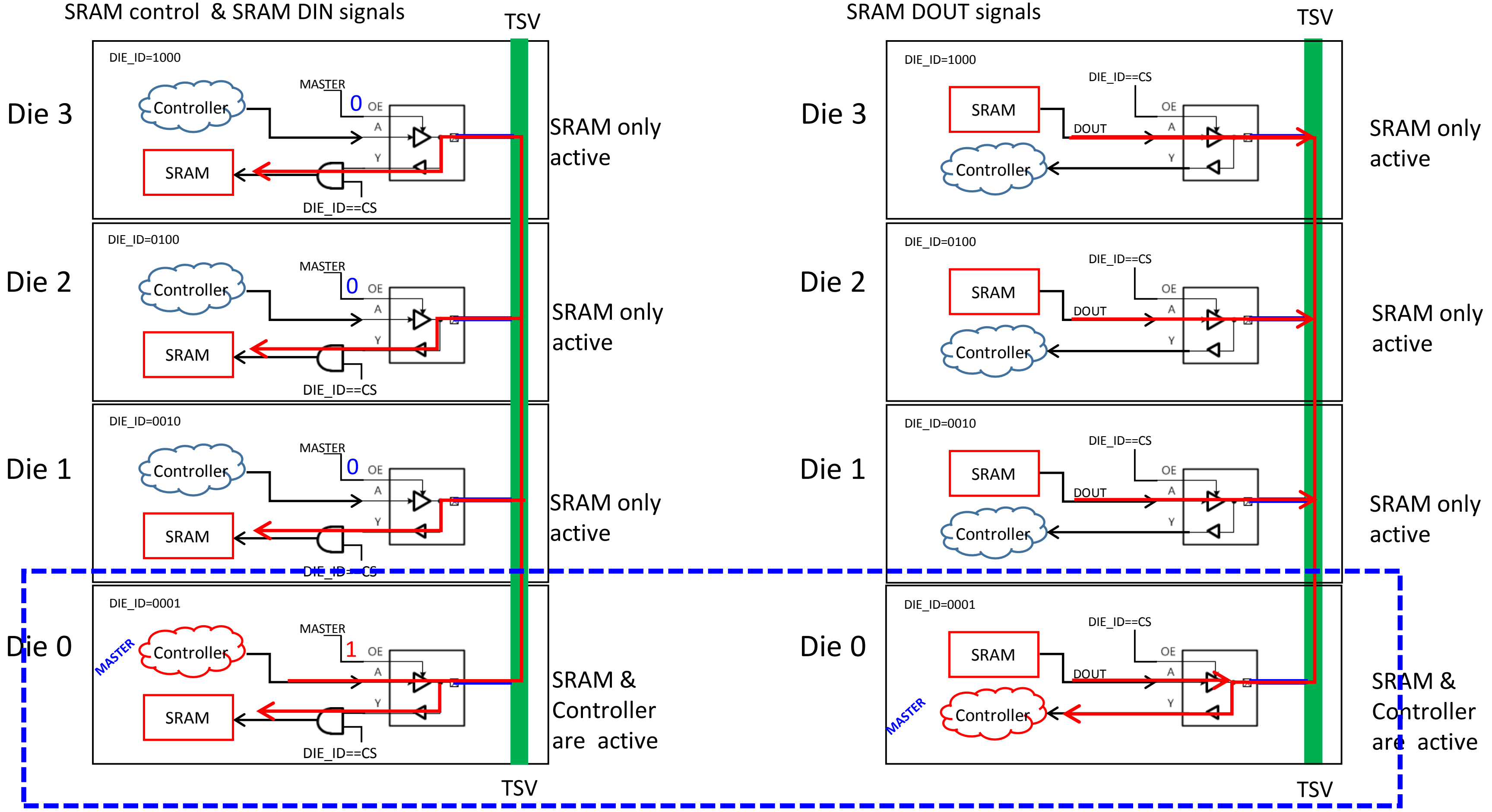
	Current work	Future work
Driver	0.07pj/bit	0.05pJ/bit
	• Eliminating the on-die termination (ODT) • Minimizing the # of blocks using VDDQ power • Using source termination in main driver to improve signal integrity	• Un-terminated low voltage swing signaling
Receiver	0.03pJ/bit	0.02pJ/bit
	• Using the only core logic power (VDD) and it is designed to minimize the # of stages • No level-shifter is needed in receiver path	• CMOS base receiver scheme instead of single-ended type receiver

[Vertical Design View]

- **Any die on the stack can be a master by MASTER_SEL**

MASTER = Die 0 OUTPUT
SRAM control & SRAM DIN signals

MASTER = Die 0 INPUT
SRAM DOUT signals



SRAM only active

SRAM only active

SRAM only active

SRAM & Controller are active

SRAM only active

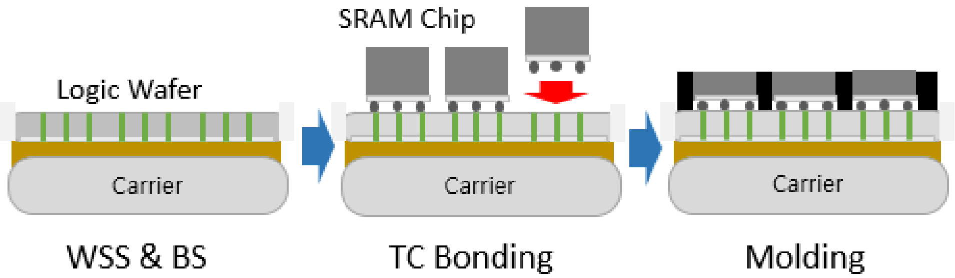
SRAM only active

SRAM only active

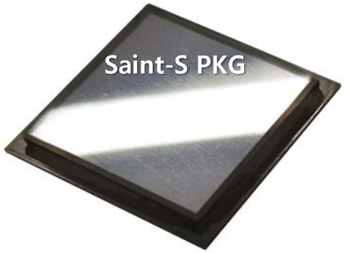
SRAM & Controller are active

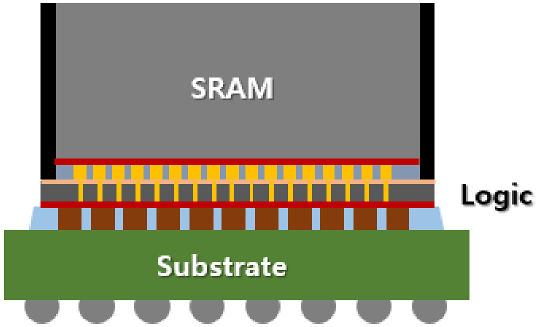
[Package Structure]

CoW Process

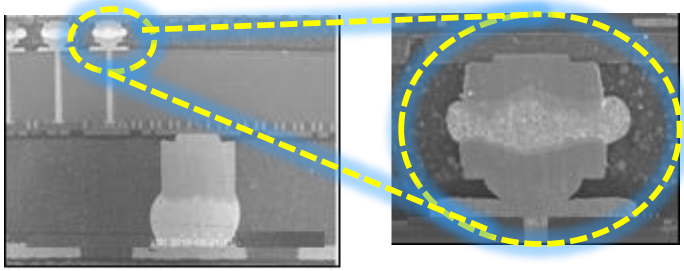


Package Image



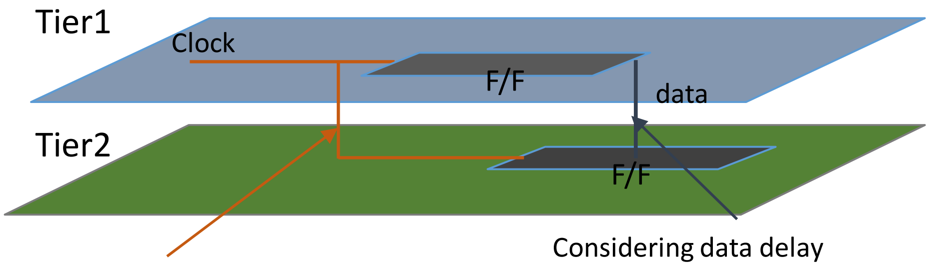


- Chip on wafer (CoW) Bonding
- Based on Mass Manufacturing Infra with >1.33 CpK
- Passed JEDEC Package Reliability Standard
- Package Flexibility; Multi-memory Stacking (≥ 2H)
  Memory Side by Side Bonding
  Face to Face Bonding

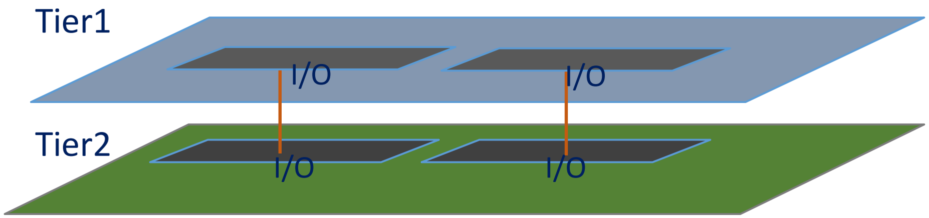


[Implementation & Design Methodology]
3D IC Interface STA Method

- **Budget-based Flow**
  - [Constraint Budgeting] Tier1: set delay constraints on register- to-IO input paths
  - [Constraint Budgeting] TSV: IO-to-IO Path
  - [Constraint Budgeting] Tier2: set delay constraints IO-to-register paths
  - Add the constraints and run STA
  - Run SPICE simulation using TSV paths and check the constraints
- **By running STA of each die concurrently, TAT and resource can be optimized as in a conventional design with maintaining SPICE-level accuracy of jitter/DCD in TSVs**



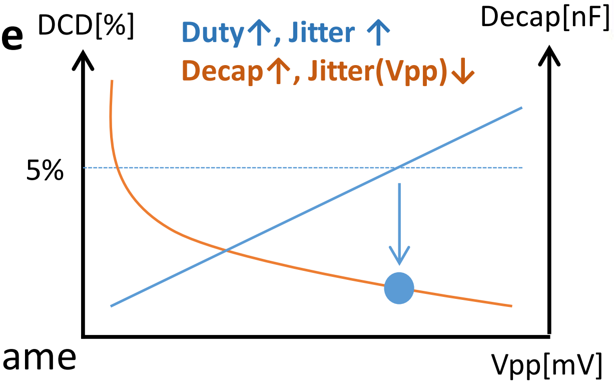
Considering clock skew of both Tier1 and Tier2 at a time



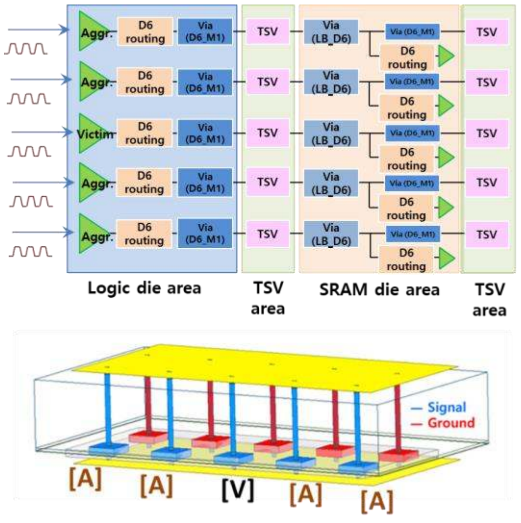
Designers have better to assign other tier's context(clock latency, data delay and TSV delay) at one tier' in SDC

[Implementation & Design Methodology] PSI Analysis on TSV Interconnects

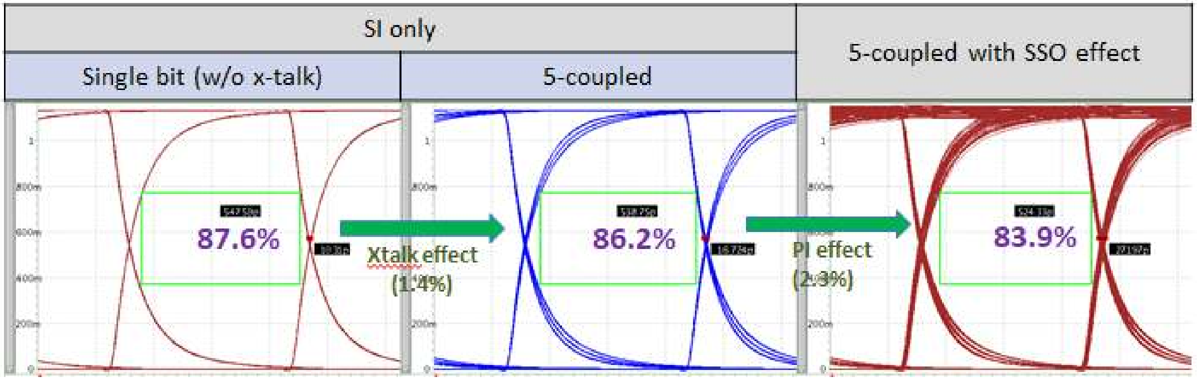
- **IO decap for Power Noise design guideline at pre-layout stage**
  - Criteria: below 5% of Duty Cycle
  - Insertion guideline: 30nF for enabling 760MHz interface



- **PSI Analysis for 1-stack SRAM at 800MHz**
  - Focusing on only 5-coupled lanes (Total IO: 256ea) means the same effect to consider all SSO noise from other 251 IO lanes



< Channel simulation & TSV modeling >



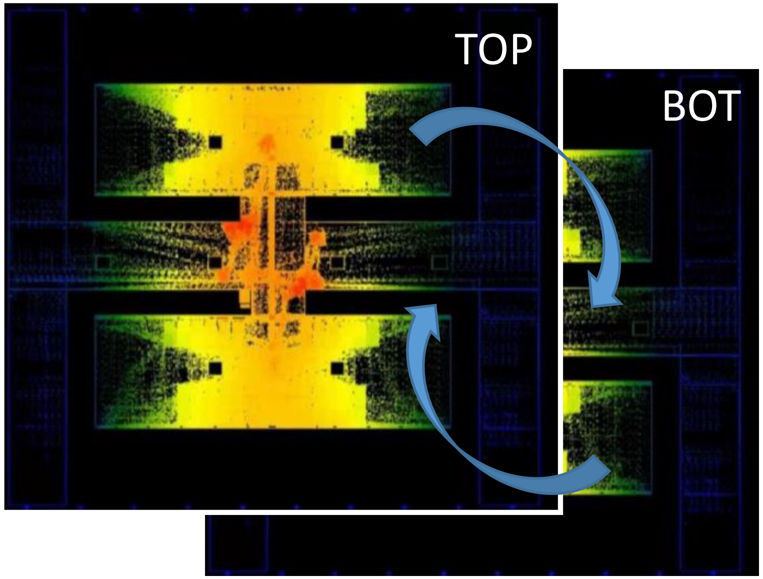
[Implementation & Design Methodology] 3D IC IREM Analysis Flow

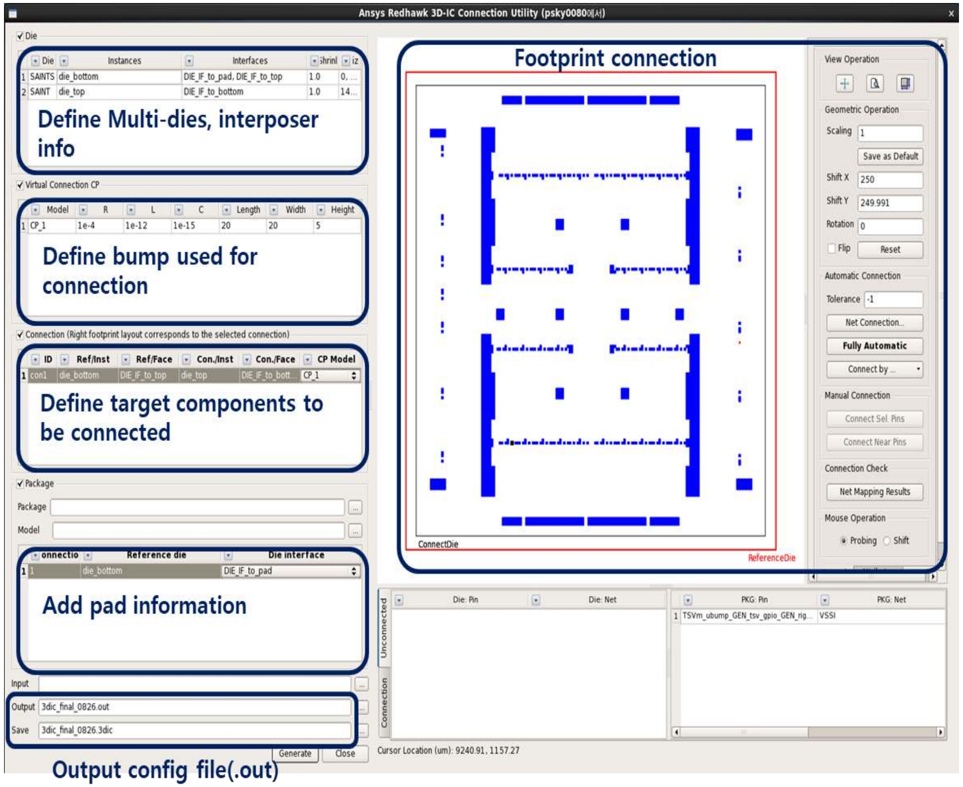
Concurrent multi-die analysis flow

- Showing how hot spots have an effect on the other die concurrently
- According to this result, the power meshes of both logic and SRAM die are reinforced

Difference from single chip IREM

- Chip to chip connection
- TSV modeling





Ansys Redhawk 3D-IC Connection Utility (psky0080에서)

☑ Die

	Die	Instances	Interfaces	Shrinl	iz
1	SAINTS	die_bottom	DIE_IF_to_pad, DIE_IF_to_top	1.0	0, ...
2	SAINT	die_top	DIE_IF_to_bottom	1.0	14...

Define Multi-dies, interposer info

☑ Virtual Connection CP

	Model	R	L	C	Length	Width	Height
1	CP_1	1e-4	1e-12	1e-15	20	20	5

Define bump used for connection

☑ Connection (Right footprint layout corresponds to the selected connection)

	ID	Ref/Inst	Ref/Face	Con./Inst	Con./Face	CP Model
1	con1	die_bottom	DIE_IF_to_top	die_top	DIE_IF_to_bott...	CP_1

Define target components to be connected

☑ Package

Package:

Model:

	onnectio	Reference die	Die interface
1	1	die_bottom	DIE_IF_to_pad

Add pad information

Input:

Output: 3dic_final_0826.out

Save: 3dic_final_0826.3dic

Output config file(.out)

Footprint connection

View Operation

Geometric Operation

Scaling: 1

Save as Default

Shift X: 250

Shift Y: 249.991

Rotation: 0

□ Flip

Reset

Automatic Connection

Tolerance: -1

Manual Connection

Connection Check

Mouse Operation

● Probing ○ Shift

Die: Pin	Die: Net

	PKG: Pin	PKG: Net
1	TSVm_ubump_GEN_tsv_gpio_GEN_rig...	VSSI

Cursor Location (um): 9240.91, 1157.27

Automatic 3DIC connection utility

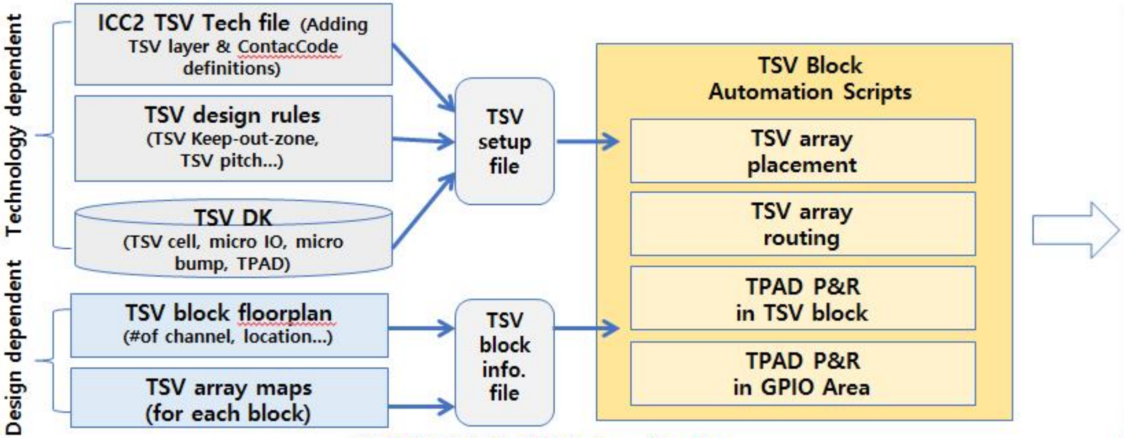
[Implementation & Design Methodology] 3D-IC Auto P&R Flow

3D-IC P&R Challenges

- New TSV-related rules restrict floor-planning and placement work
- The number of TSVs is more than 1000ea
- Efficient TSV signal/power placement architecture is necessary to minimize design overhead

P&R Solutions

- Custom scripts for automated placement & routing



Technology dependent

ICC2 TSV Tech file (Adding TSV layer & ContacCode definitions)

TSV design rules (TSV Keep-out-zone, TSV pitch…)

TSV DK (TSV cell, micro IO, micro bump, TPAD)

Design dependent

TSV block floorplan (#of channel, location…)

TSV array maps (for each block)

TSV setup file

TSV block info. file

TSV Block Automation Scripts

TSV array placement

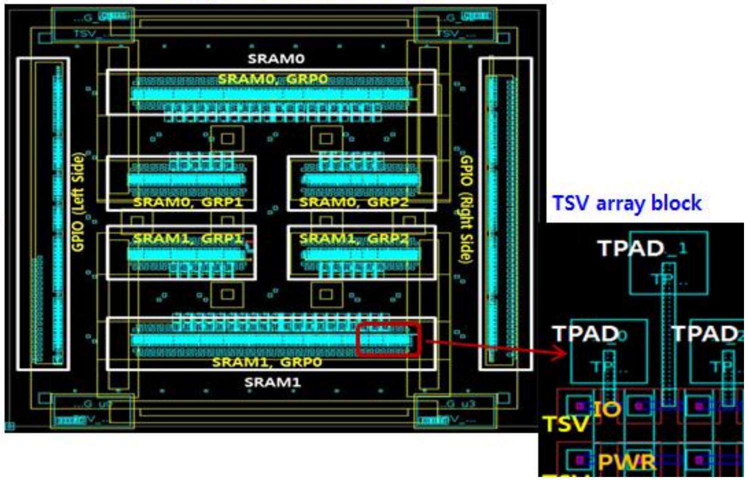
TSV array routing

TPAD P&R in TSV block

TPAD P&R in GPIO Area

ICC2 3D-IC Auto P&R Automation Flow

Test Design (6 TSV Block/2 GPIO)



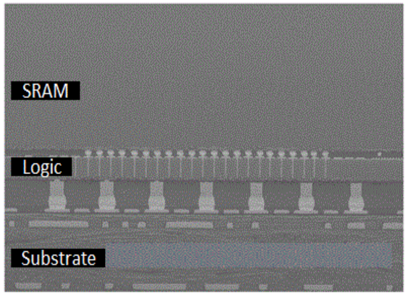
[Results & Future work]

Chip Summary

	Specification
Process	7nm CMOS
Die Size	9 x 9 mm^2 (SRAM die) , 9.5 x 9.5 mm^2 (Logic die)
Package Size	12 x 12 mm^2
SRAM Capacity	64Mbit per die
Supply Voltage	0.85V (Logic/SRAM), 1.0V (TSVIO), 1.8V (GPIO)
Frequency	~760MHz
# of channel	128bit x 2ch
Power consumption	0.156W (SRAM+TSVIO)
Bandwidth	24.3GB/s per channel
Latency	7.2 ns(read), 2.6 ns (write)

Performance Comparison

	Memory Bandwidth per Power (GB/second/Watt)	Memory Latency (ns)
GDDR6	25	45 (70% column hit)
HBM2e	70	
SAINT-S	156.1	Read 7.2/Write 2.6



SAINT-S PKG SEM picture

SAINT-S Test board

Thank You

samsung foundry

SAMSUNG