

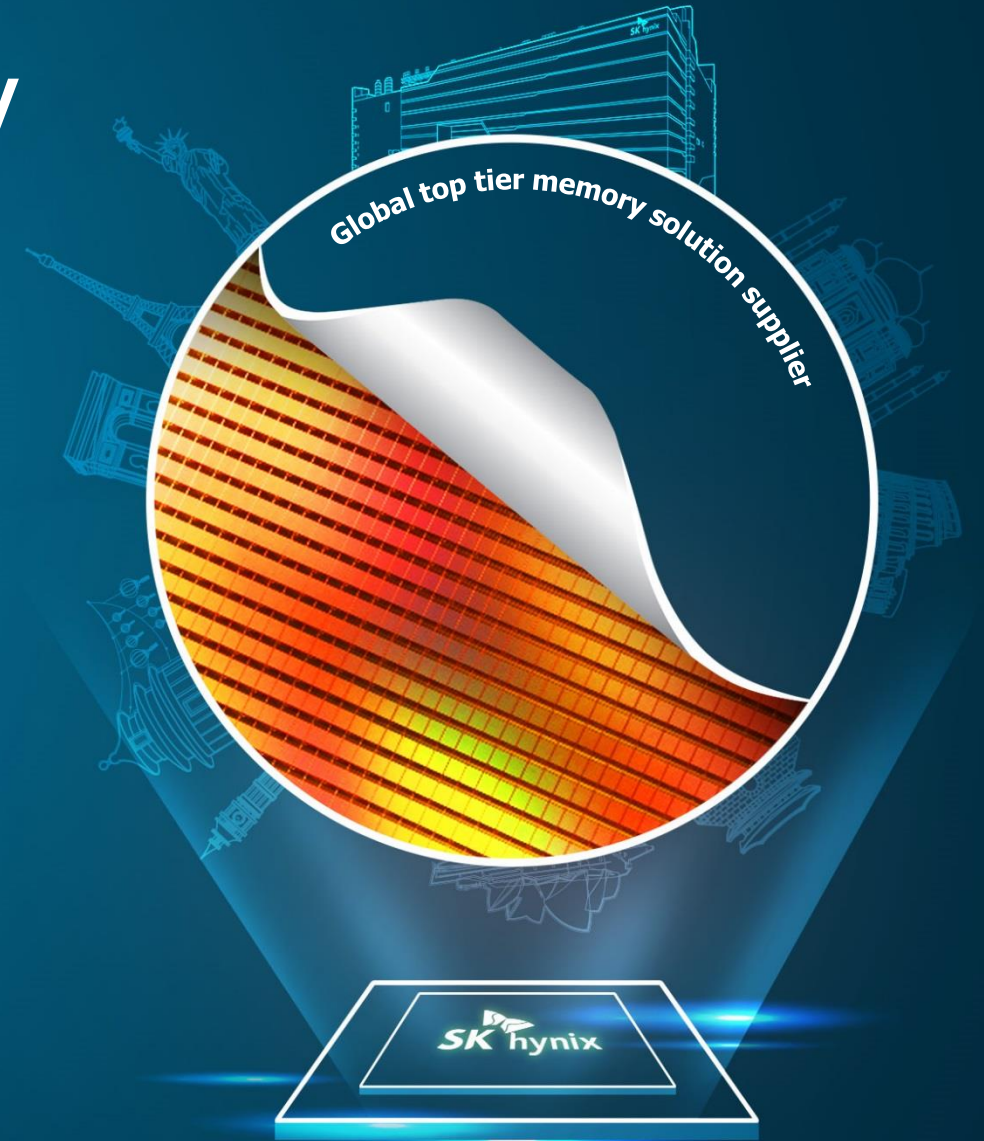
Advanced Packaging Technology for Beyond Memory

Nov. 22nd, 2023

—
Ho-Young Son, Ph.D



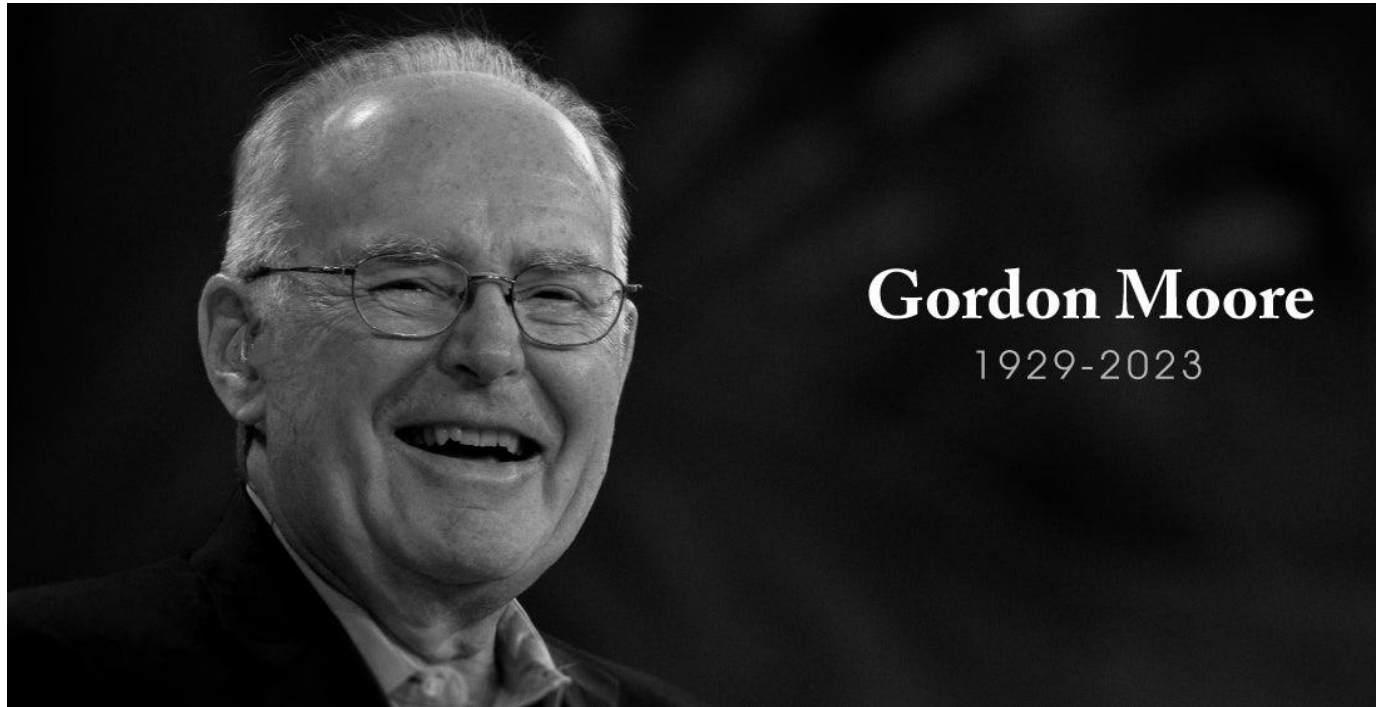
© SK hynix Inc.
This material is proprietary of SK hynix Inc. and subject to change without notice.



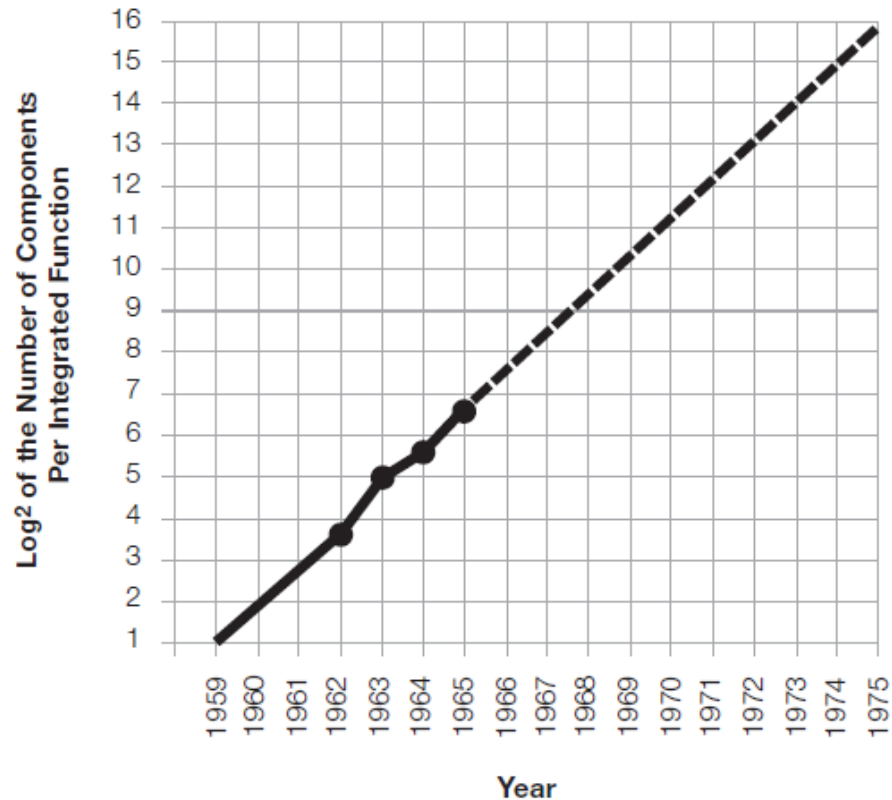
CONTENTS

Advanced Packaging Technology for Beyond Memory

- I ○ Introduction
- II ○ What is HBM?
- III ○ Package Technology for HBM
- IV ○ Potential Application for Beyond Memory
- V ○ Summary



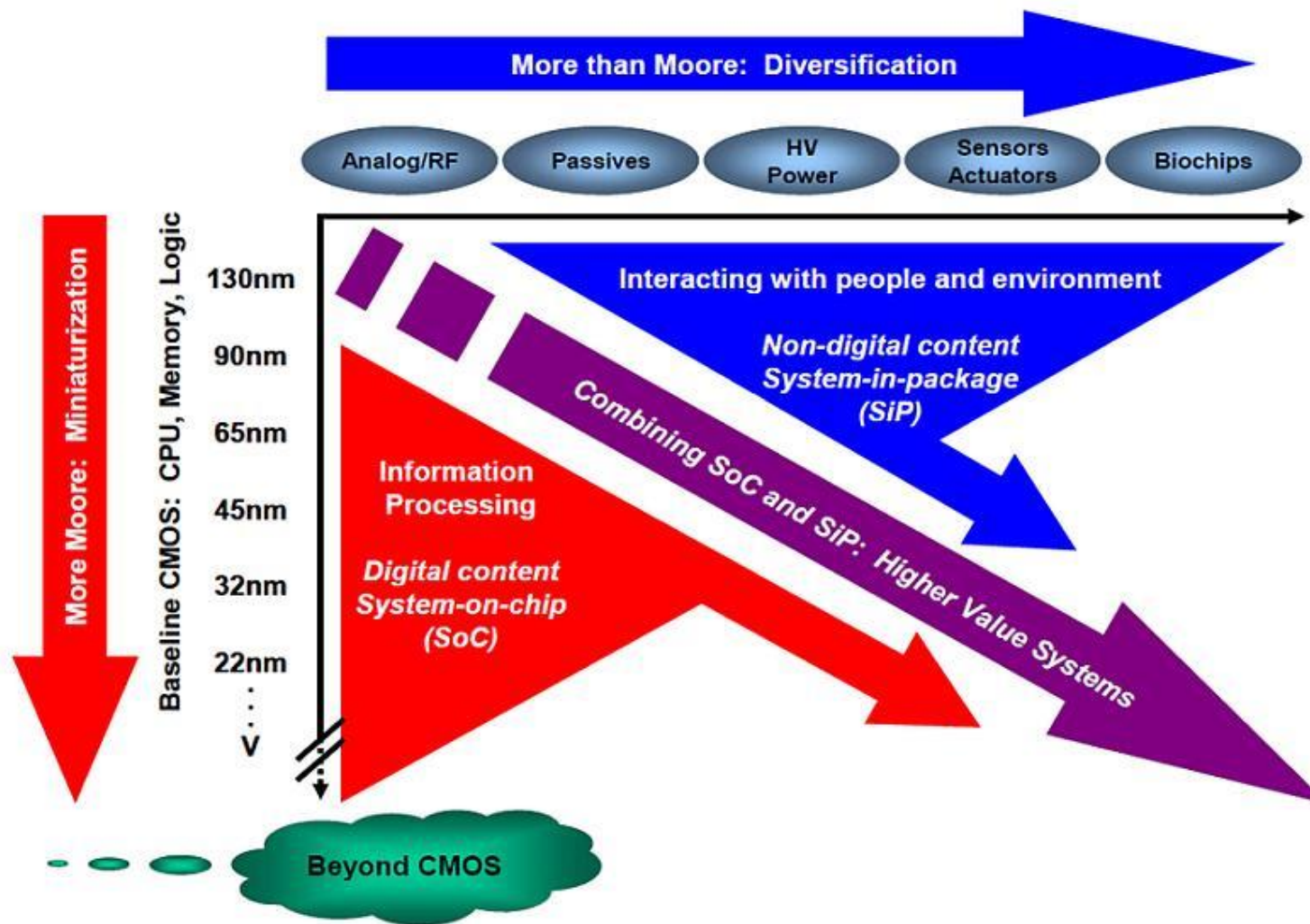
- ✓ Last and Key member among Traitorous Eight
- ✓ Founder of Intel
- ✓ Moore's Law
- ✓ 우리가 몰랐거나 오해하고 있는 사실?



Electronics, v38, n8, 1965

- ✓ 1965년에 발표한 Gordon Moore의 주장
- ✓ Empirical Data (by past 6 years) → Law (after 10 years) → Philosophy (for over next 30 years)
- 반도체 분업화

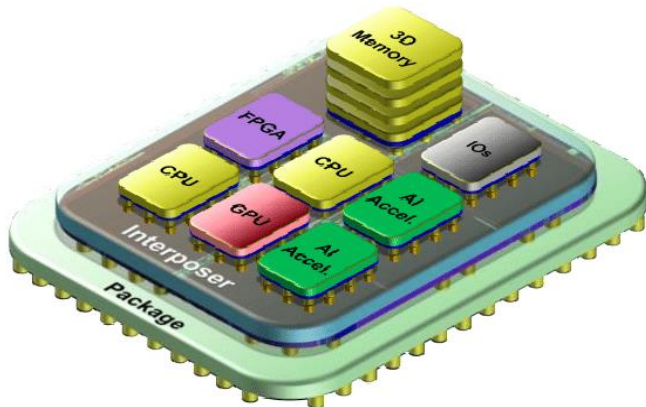
More Moore, More than Moore



w.r.t 'More Moore and More than Moore'

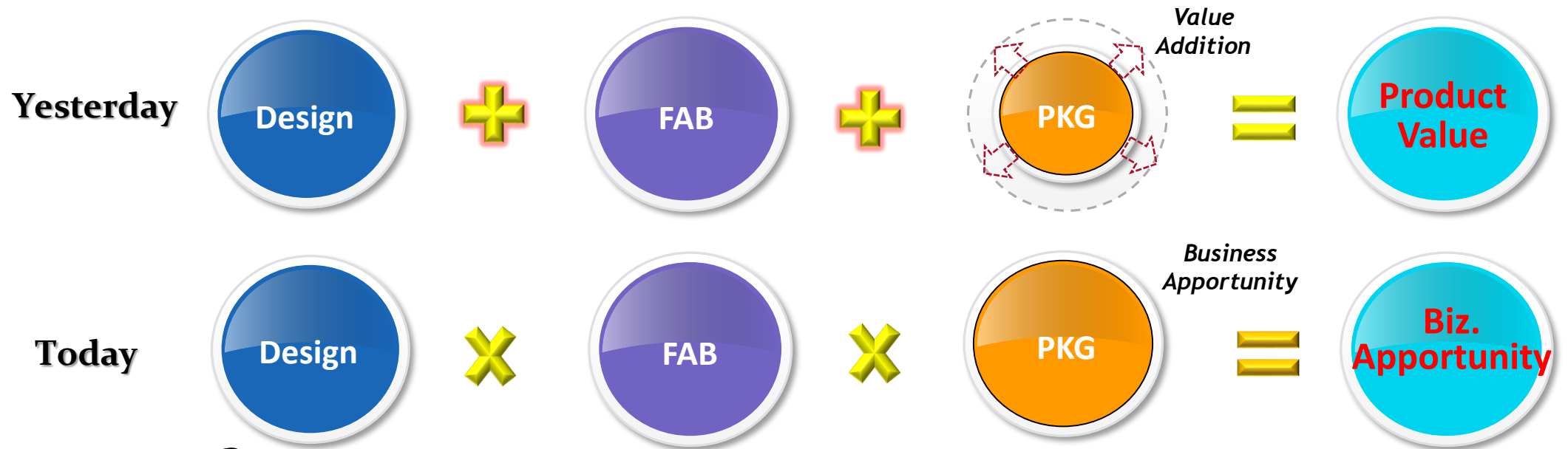
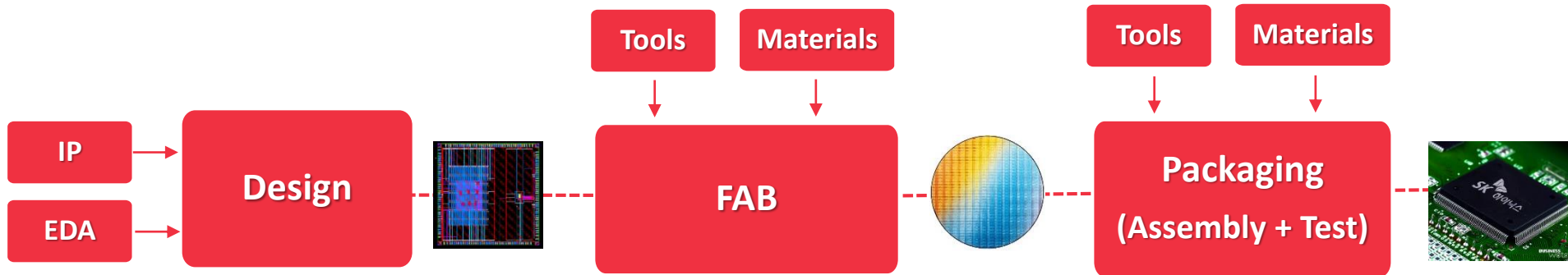
- ✓ It may prove to be more economical to build large systems out of **smaller functions, which are separately packaged and interconnected**. The availability of large functions, **combined with functional design and construction**, should allow the manufacturer of large systems to design and construct a considerable variety of equipment both rapidly and economically

Electronics, v38, n8, 1965



- ✓ Moore는 Transistor scaling 뿐만이 아니라 Heterogeneous Integration에 대해 예견하고 있음
- ✓ More Than Moore 또는 Beyond Moore는 과연 적절한 표현일까?
- ✓ Dr. G. Moore의 선견지명

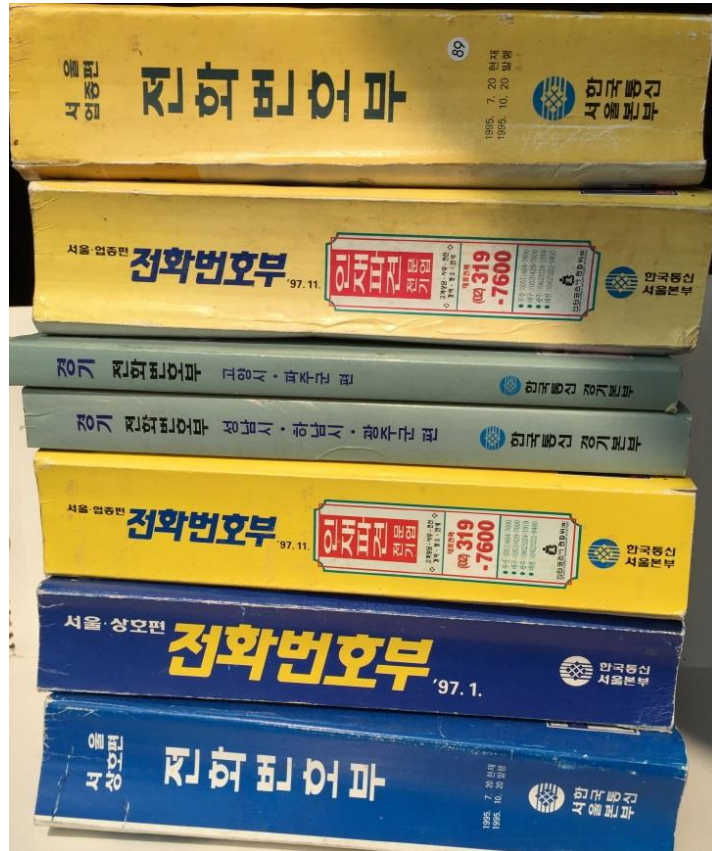
제품의 가치를 높임 (Value Addition) → 새로운 사업화를 좌우(Biz. Creation) → 기업의 경쟁력에 직결

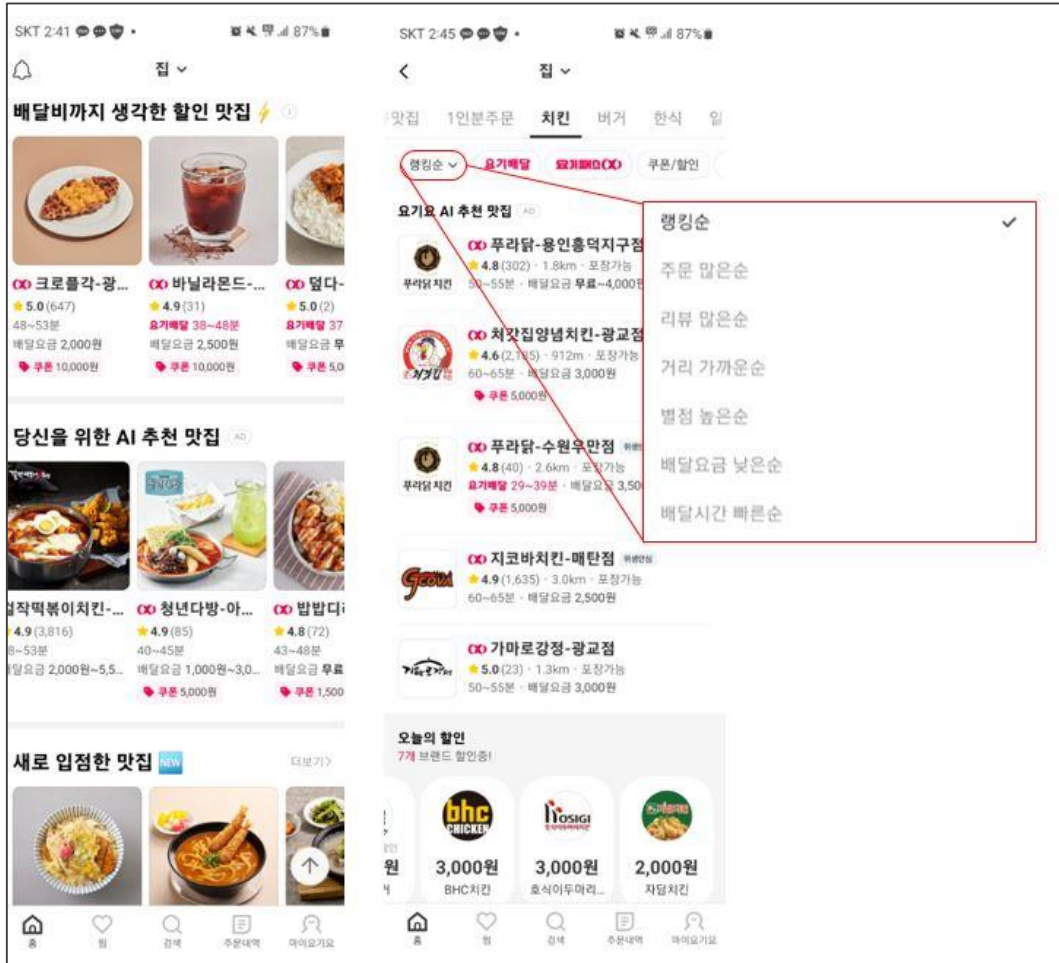


Tomorrow?

Source : K.W.Lee, 첨단 반도체 패키징 국회포럼, 2022

30년 전 치킨 주문하기





치킨을 어디서 주문할 것인지?

어떤 치킨 종류를 택할 것인지?

어떤 치킨 브랜드, 영업점에서 주문할 것인지?

- 소비자의 수고를 최소화하도록 지원하는 AI Platform

(= 귀찮은 것은 Memory가 알아서 해주는 System)

Data Traffic Growth induces High Memory Adoption

ICT System에서의 Data Usage의 급격한 증가 (Data Storage, Data Processing)

Data Traffic Growth

(Unit: ZB)

200

150

100

50

Total Data Volume



Number of Mobile devices exceed PC

Connected devices: 14B

Google launches AlphaGo Zero

5G Commercialized

Connected devices: 50B
Number of SSD exceed HDD

Open AI launches ChatGPT and has crossed 1 million users within one week of its launch.

2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023F 2024F 2025F

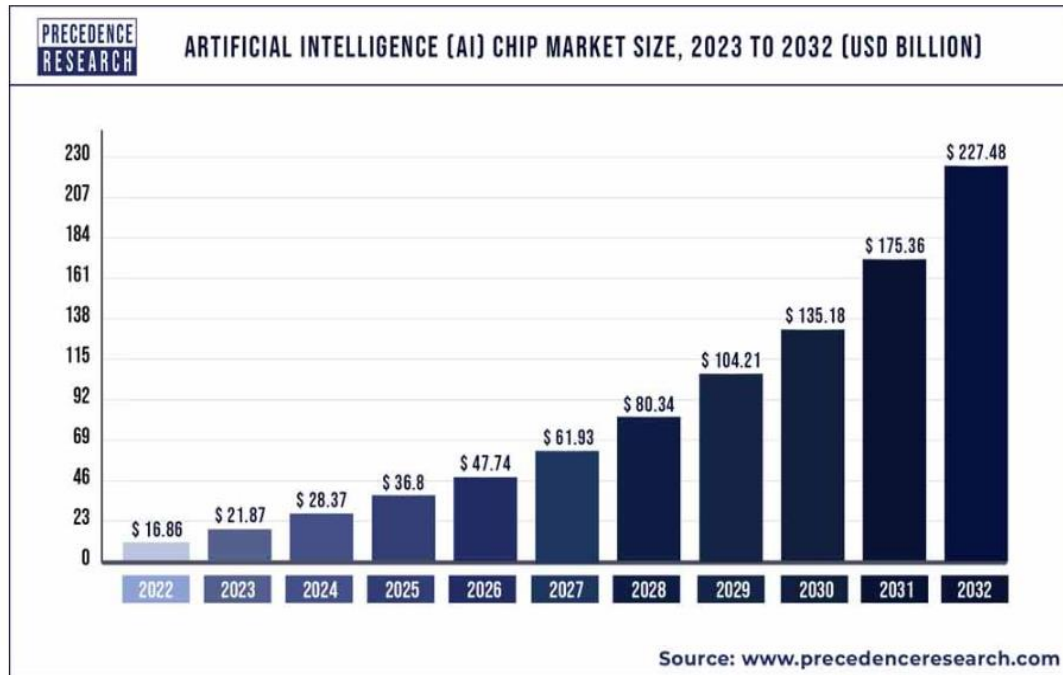
Source: IDC, Merrill Lynch, Morgan Stanley data

© SK hynix Inc.

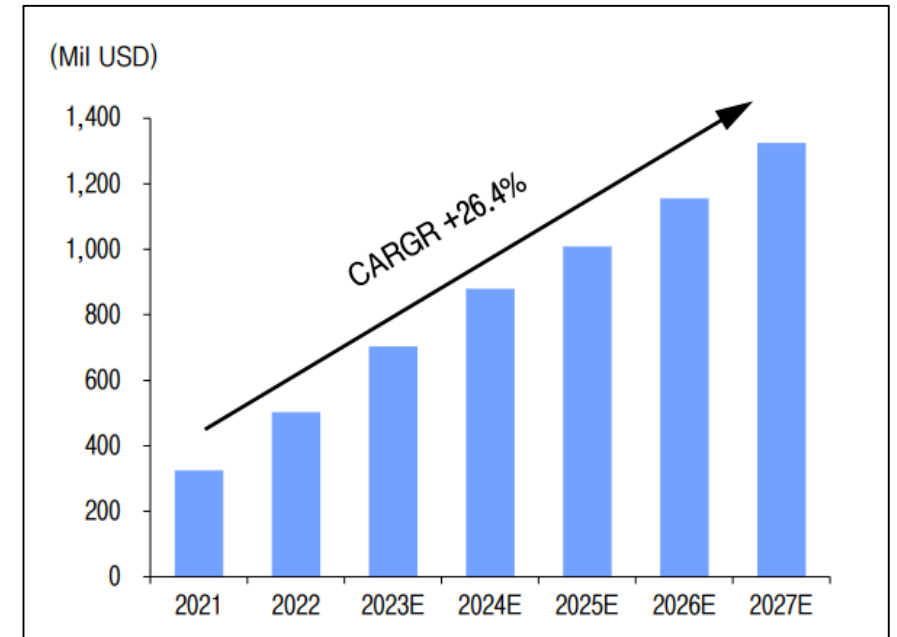
This material is proprietary of SK hynix Inc. and subject to change without notice./ Confidential

AI 시장 성장에 맞춰 HBM Market 도 급성장 중

AI Semiconductor Market Size



HBM Market Growth

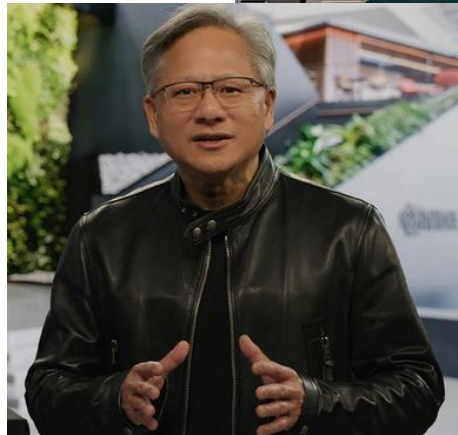


Source : Trendforce, 이베스트리서치센터 (23.5.31)

nVIDIA's H100 AI Accelerator



ChatGPT H/W
(8 X nV H100 Powering in Tool)

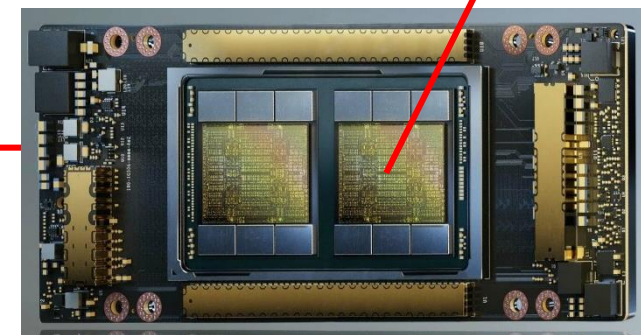
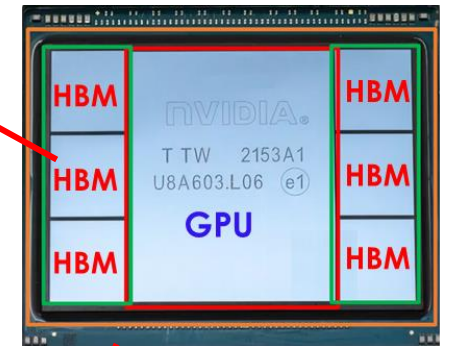


SK Hynix HBM3



1st Supplier and Market Leader

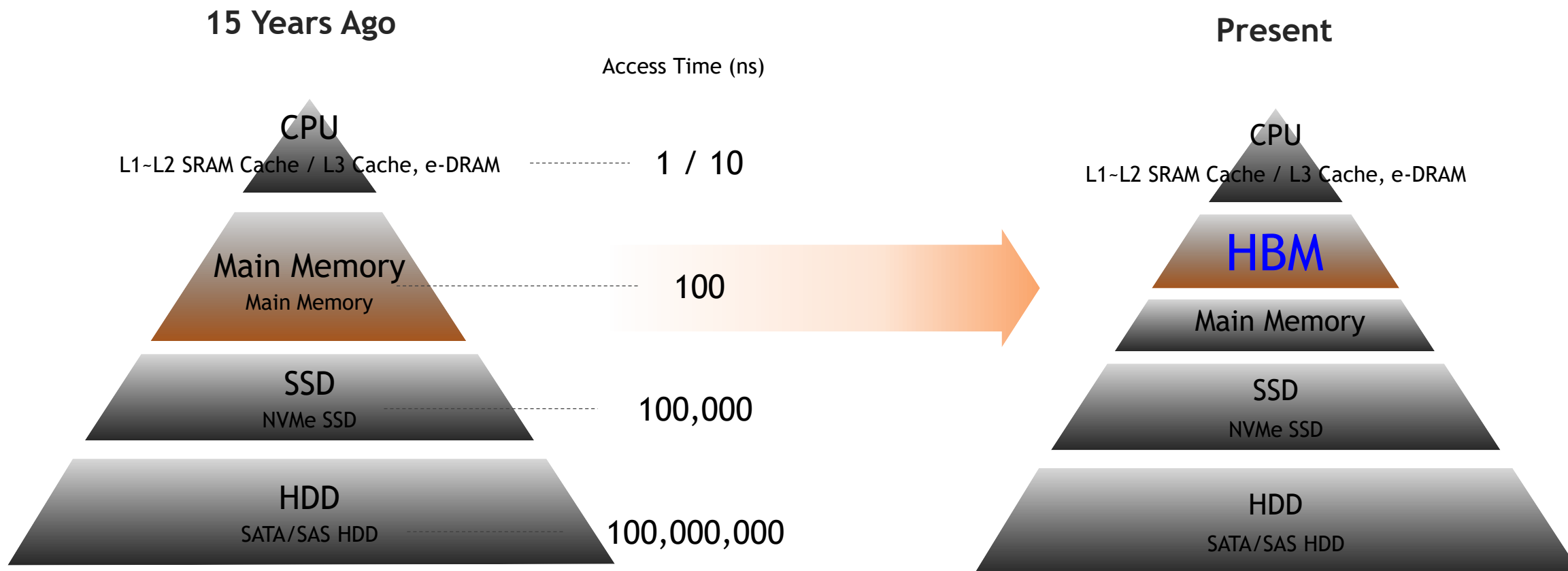
2.5D SiP
(1GPU + 6HBM)



nVidia H100 Tensor Core GPU

HBM in Memory Architecture

HBM은 xPU와 Main Memory 사이의 Latency를 가지는 메모리



○ High Bandwidth Memory

- Provided Highest Bandwidth among existing memory and Highest Capacity per unit area

○ **Industry Standard In-Package Memory for High Perf. Graphics and HPC/AI Computing**

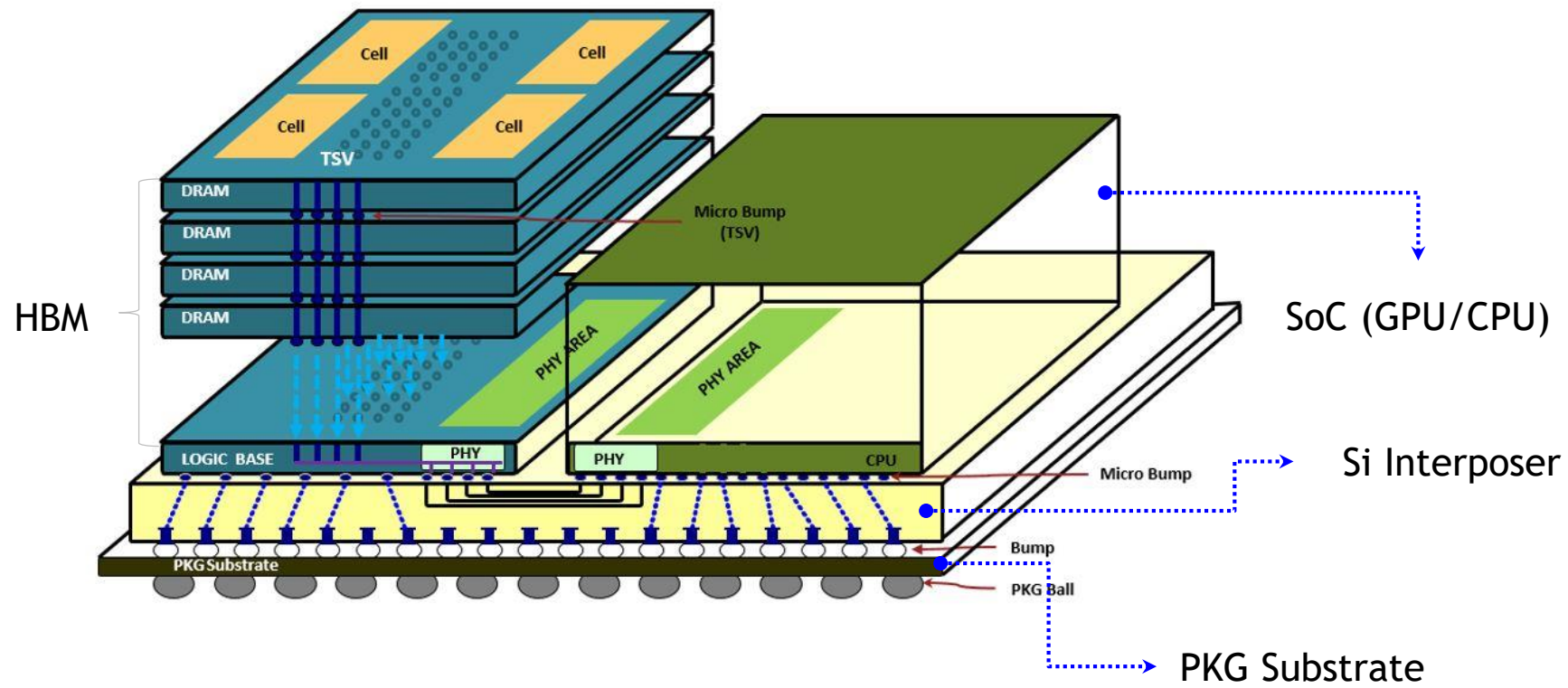
- Proposed by AMD & SK Hynix
- JEDEC Test Group Started in 2011 : Architecture (JC11), Mechanical Spec (JC42)
- World FIRST HBM1 released by SK Hynix (2013)
- World BEST HBM2E/HBM3/HBM3E succeeded by SK Hynix (2019/2022/2023)

○ **State-of-the-art Packaging Technology Adopted**

- 2.5D silicon interposer with micro-bumps
- 3D TSV and micro-bumps for DRAM multi-die stack (HBM)

2.5D SiP (System-in-Package) with HBM

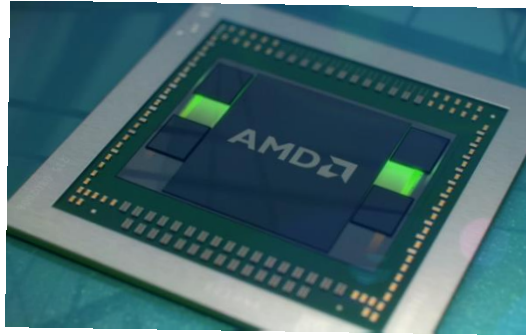
- 1) External : SoC PHY to HBM PHY interconnect across Si interposer
- 2) Internal (Memory) : Logic Buffer (Base) + Multi DRAM Die (Core) using TSV/Micro-Bumps



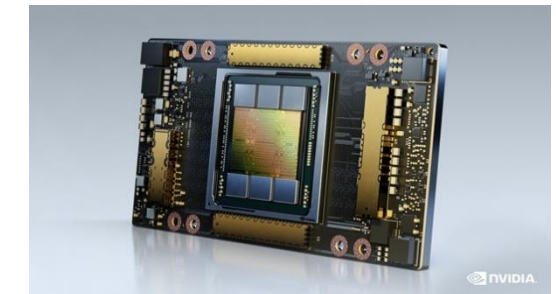
High End Graphics



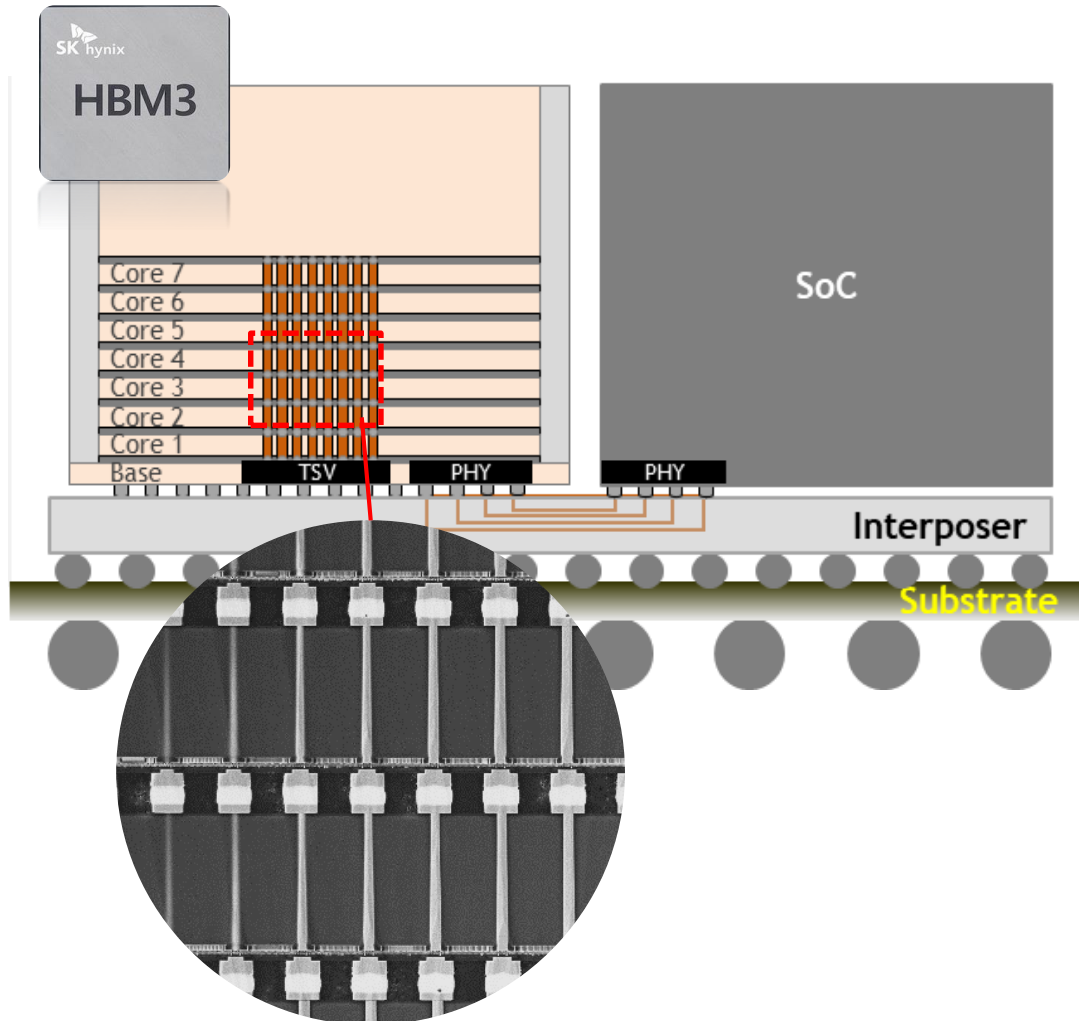
High Performance Computing (HPC)



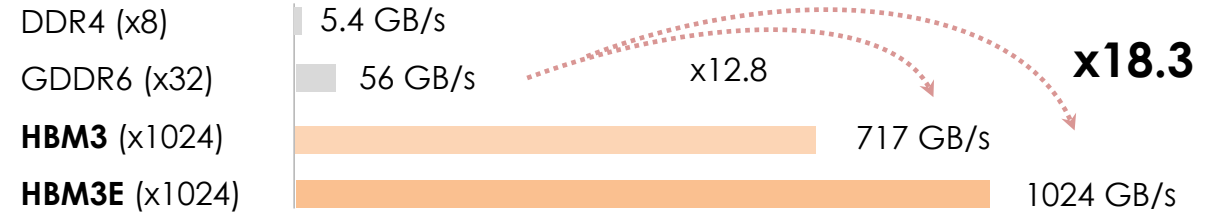
Artificial Intelligence (AI)



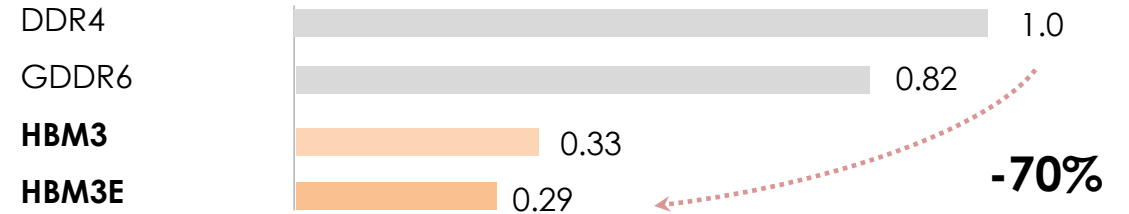
Bandwidth, Capacity, Power Efficiency, Thermal Dissipation 측면 현존 최고 성능 메모리 제품



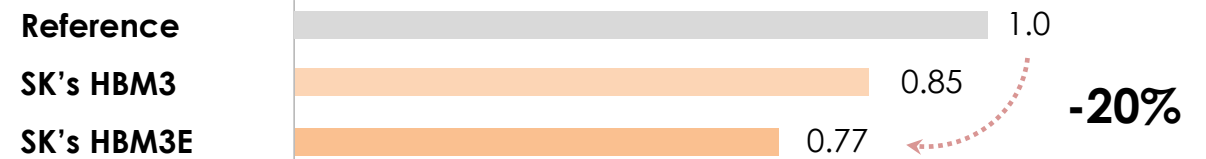
Highest Memory Bandwidth



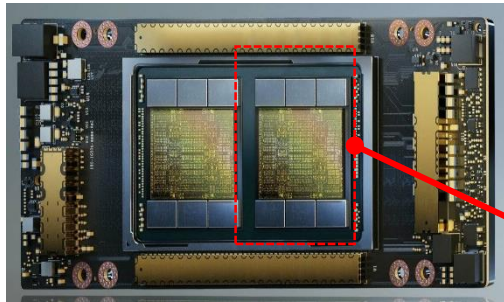
Highest Power Efficiency



Better Heat Dissipation (approx.)



HBM Ecosystem

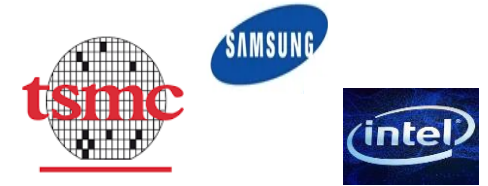


HBM (Memory Maker)

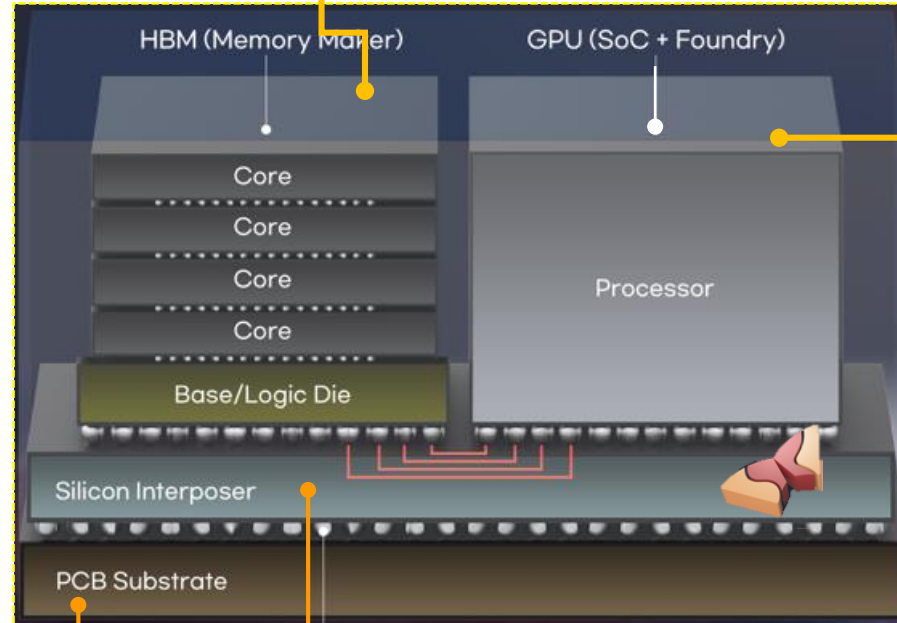


Tool/Material Suppliers

SoC (Foundry)



Fabless/Hyperscaler



2.5D SiP (Packaging & Test @ OSAT)



Substrate

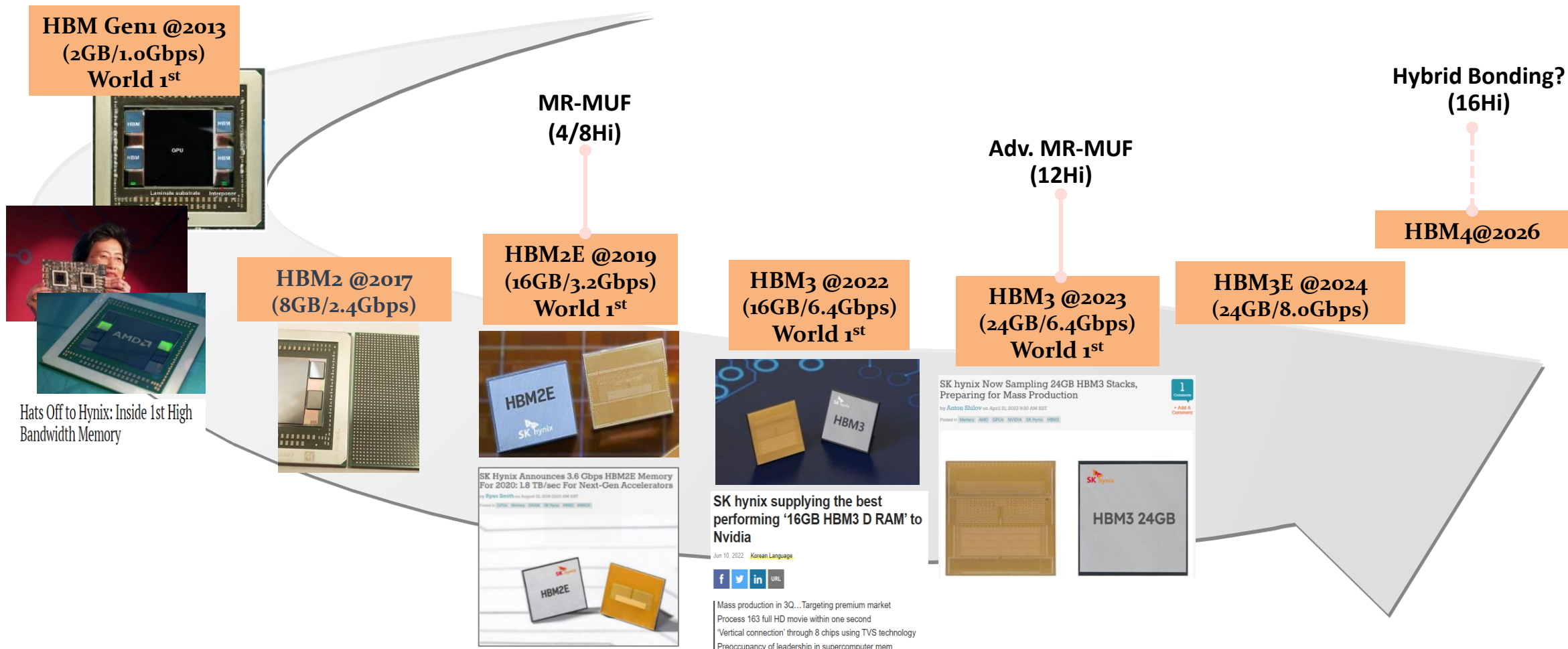


Si Interposer (Foundry)



SK Hynix HBM's Footprint

Industry 최초, 최고 사양의 HBM을 연달아 출시 중
 지속적인 기술 혁신을 통해 HBM4 이후 Technology Leadership 유지 예상



HBM - High Bandwidth?

교통 체증 (Power Efficiency), 도로폭/ 단위 시간 당 이동 차량 대수 (Bandwidth), 사람/화물 (Capacity)



In Highway

4 Lanes X 256 km/h



1024 Lanes X 1 km/h



In Memory

4 IO X 256 Gbps/IO

8 IO X 128 Gbps/IO

16 IO X 64 Gbps/IO



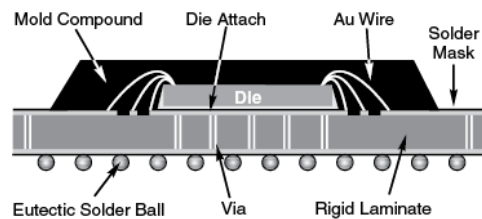
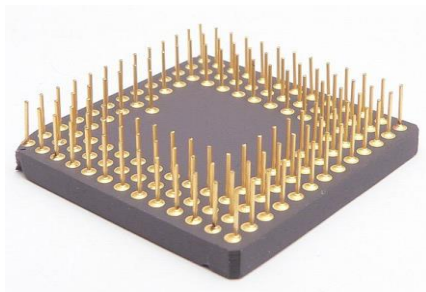
1024 IO X 1 Gbps/IO

HBM I/Os vs. Conventional Package

HBM은 현존하는 Memory Package 중 가장 많은 수의 내/외부 I/O 수를 가지며 그에 상응하는 Advanced Package 기술을 필요로 함

PGA (Pin Grid Array)

'70s~
>1mm

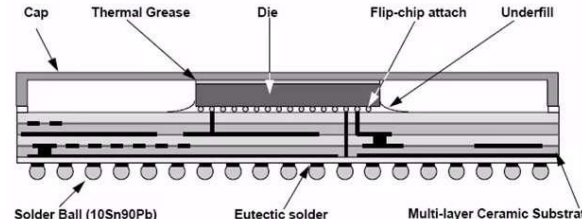
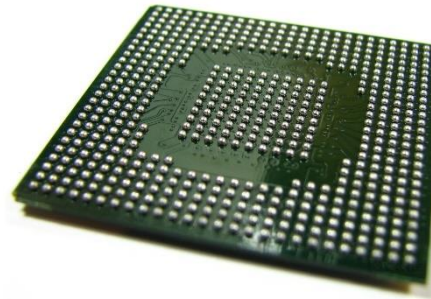


Wire bonding

Die Attach Film, Molding Compound, Thru-Via, Solder Ball

BGA (Ball Grid Array)

'90s~
0.4~0.8mm

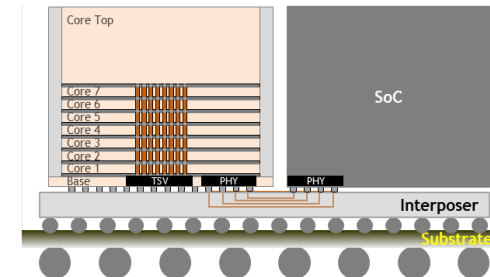
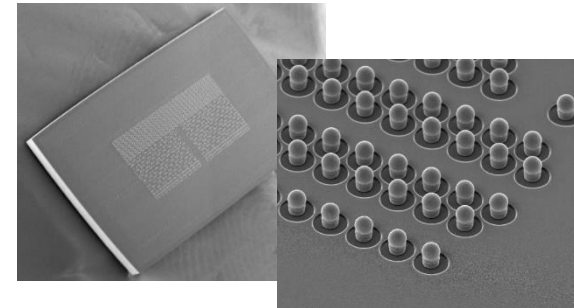


Flip Chip

Flip Chip Bump (C4, Cu-pillar), Underfill, Multi-layer Substrate

MPGA (Micro-Pillar Grid Array) for HBM

2010s~
<0.1mm



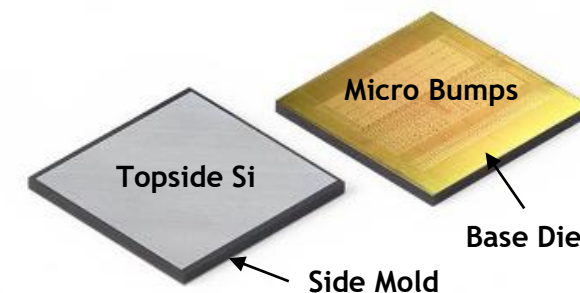
2.5D SiP (CoWoS)

TSV, Micro-Bump, Interposer, Thin Wafer Process, C2W Bonding, Wafer Mold

HBM은 KGSD (known-good stacked die) 형태로 제공되며 >5K TSV, >165K Micro-bump로 구성



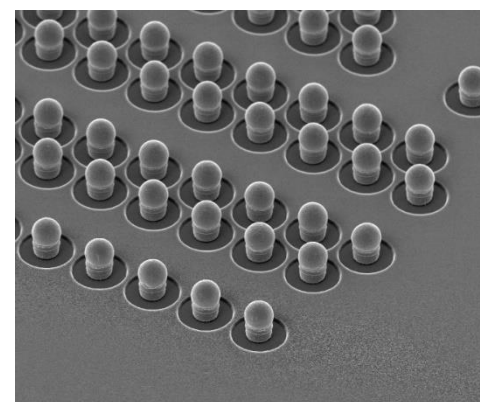
HBM Cubes



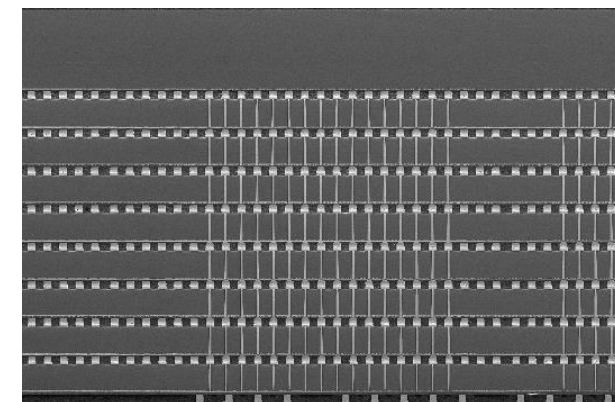
Top/Bottom View

HBM3

- 720um Z-Height, 11x11mm²
- 7775 Base Micro-Bumps
- >5.6K TSVs
- >165K Core Micro-Bumps (Incl. Dummy Bumps)
- Exposed Si of a top die for thermal dissipation
- Side mold encapsulated surrounding DRAM core dies
- 8Hi/12Hi in Production (Top die height is only different)

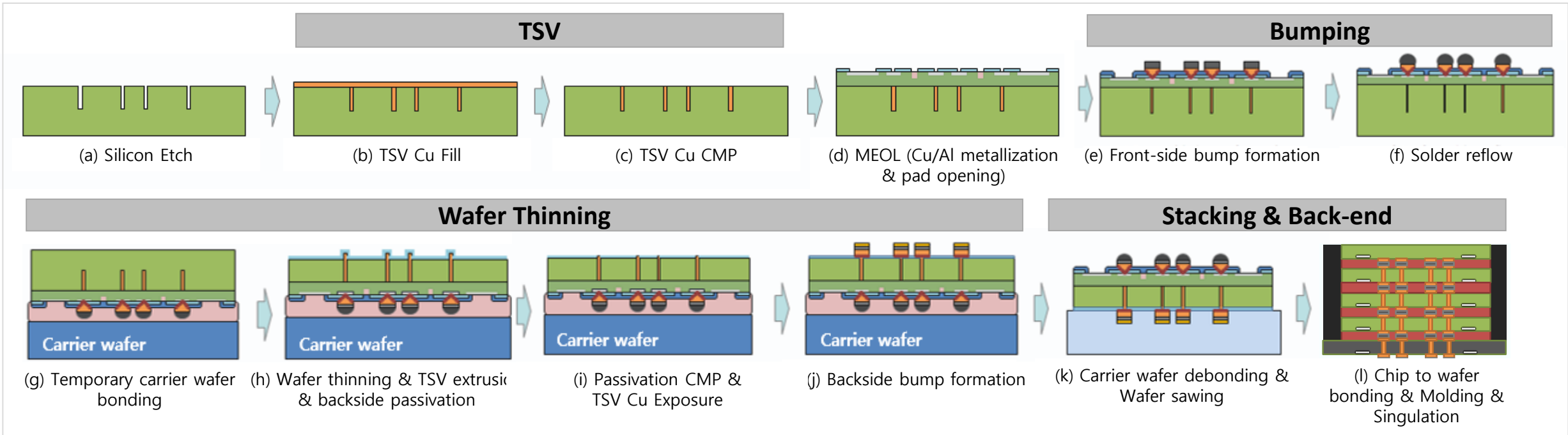


Base Die Bump

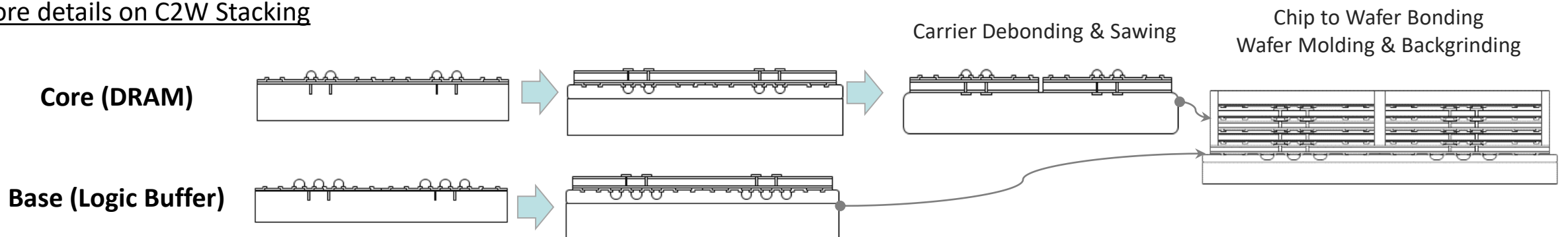


TSV/Bump Joints

HBM Package Process Flow



More details on C2W Stacking

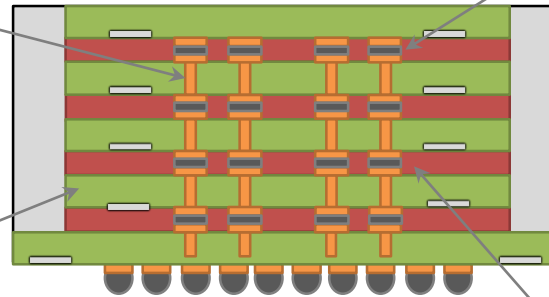


Wafer Support System과 Chip Stack 기술이 핵심이며, 특히 Chip Stack 기술이 HBM 경쟁력을 좌우



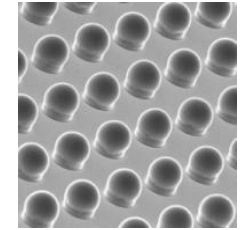
■ Via (TSV) Formation

- Matured technology
- High Stack 채용 따른 TSV Depth 완화 추세



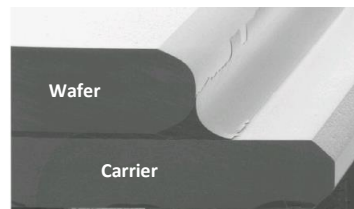
■ Bump Formation

- Matured technology but Challenging
- Pitch 감소/High Stack 채용 따라 Bump Size 미세화
→ 접합 수율, 신뢰성/품질에 직접적 영향



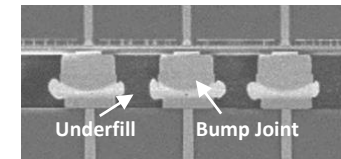
■ Wafer Support System

- Critical and High Challenging Process
- Thin Wafer 두께 감소 (Stress/Warping 증가), Stack 수 증가 (Bump 변형, Delamination/Warping Handling) 따른 Issue 극복 필요
- Vendor Specific Process



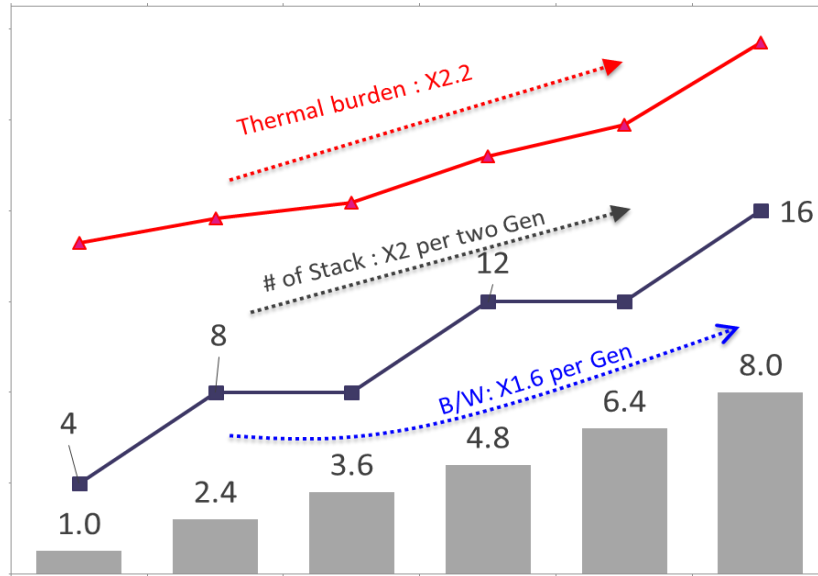
■ Chip Stack/Underfill

- Critical and High Challenging Process (Quality, Cost/Productivity)
- 대표 방식
 - 1) TC+NCF (Thermal Compression Bonding + Non-Conductive Film)
 - 2) MR+MUF (Mass Reflow Bonding + Molded Underfill)
 - 3) Hybrid Bonding (Cu-Cu & Inorganic Hybrid Bonding)

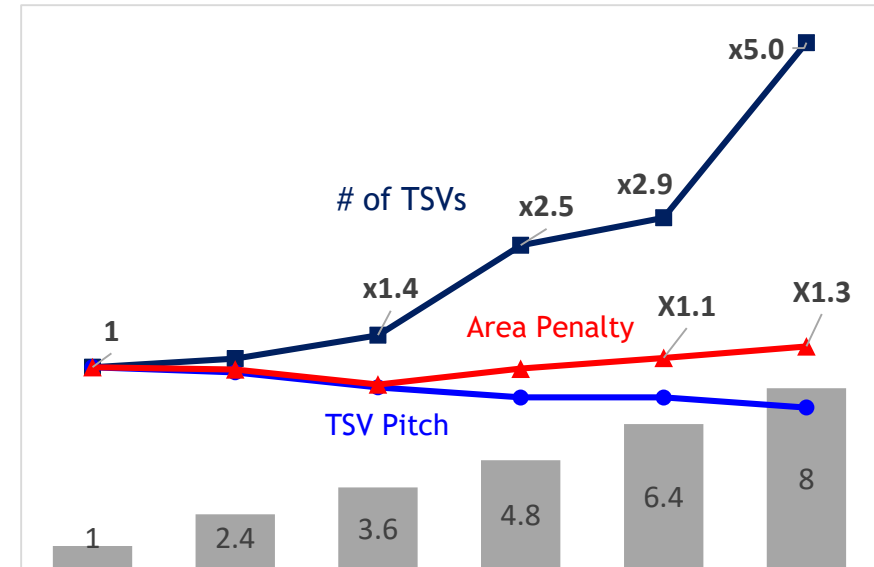


지속적인 HBM Power & Bandwidth 증가에 따라 Stack 수/TSV 개수 증가 및 Thermal Issue 심화

High Stack & Thermal (품질/기술)



TSV Area Penalty (원가/사업)



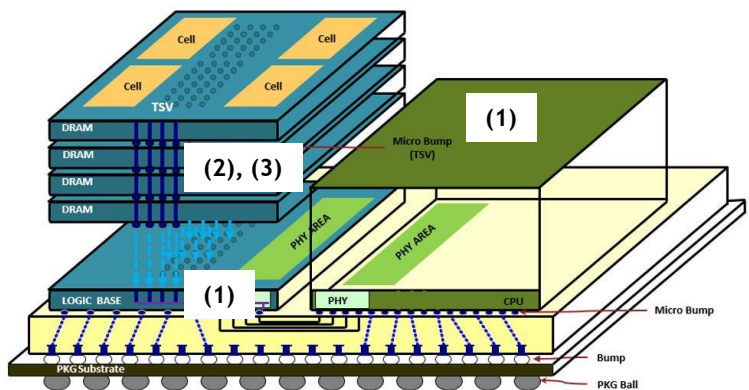
극복
방안

- **High Stack Solution**
 - Thin Die Handling, Narrow Gap Filling
 - Hybrid Bonding
- **HBM Thermal Dissipation**
 - Underfill 열전도도 향상, Thermal Dummy Bump 삽입, Si 두께 증가/Gap Fill 두께 최소화

- **Fine TSV/Bump Pitch**
 - Fine Pitch Bump 접합 수율/신뢰성 확보
 - Small Gap Fill 특성 확보
- **Cost Competitiveness**

Why thermal issues in HBM?

○ HBM의 Thermal Issue가 중요한 이유



(1) Head Source : SoC, Base Die (Power increase as Bandwidth Increase)

(2) HBM Package 자체의 열방출 방해 요소 : High stack, Polymer Underfill

(3) 열에 취약한 Memory Device

: temp-dependent degradation (refresh)

○ How can we mitigate thermal risk?

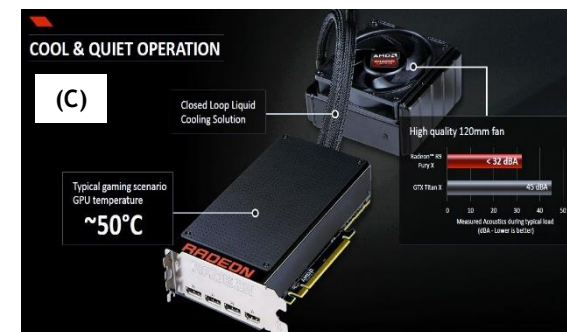
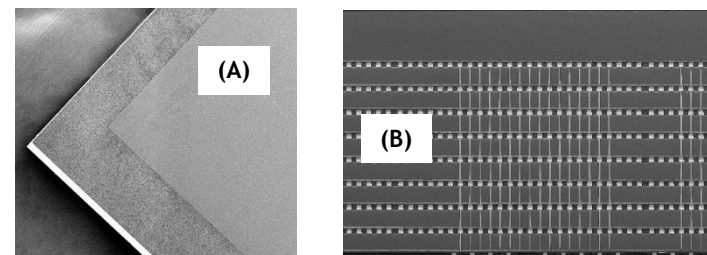
(A) Thermally enhanced HBM package (Standard) : Exposed Si

(B) HBM Package Optimization (Supplier specific)

: Thermal dummy bumps, High conductive gap-fill material

(C) System Level Thermal Design

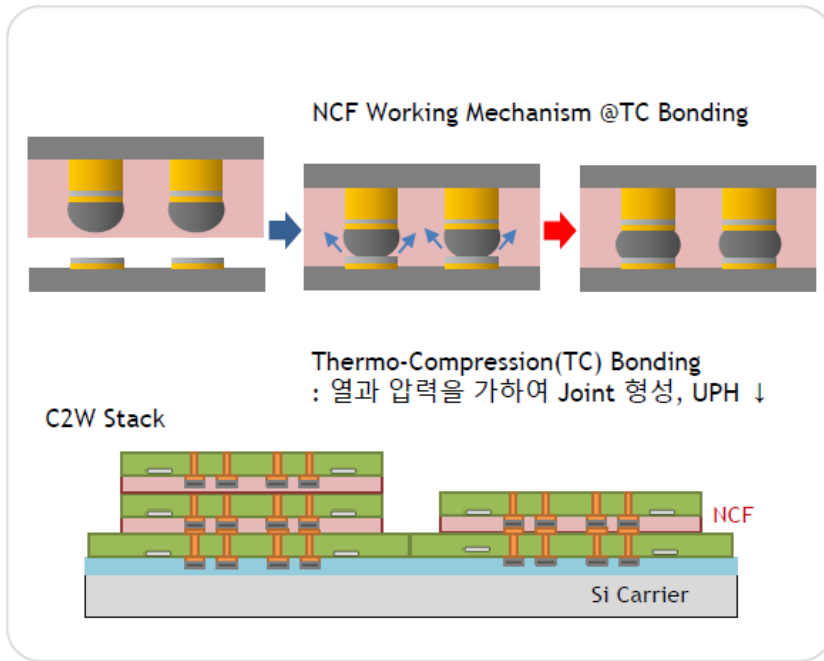
: Heat Sink / TIM Selection



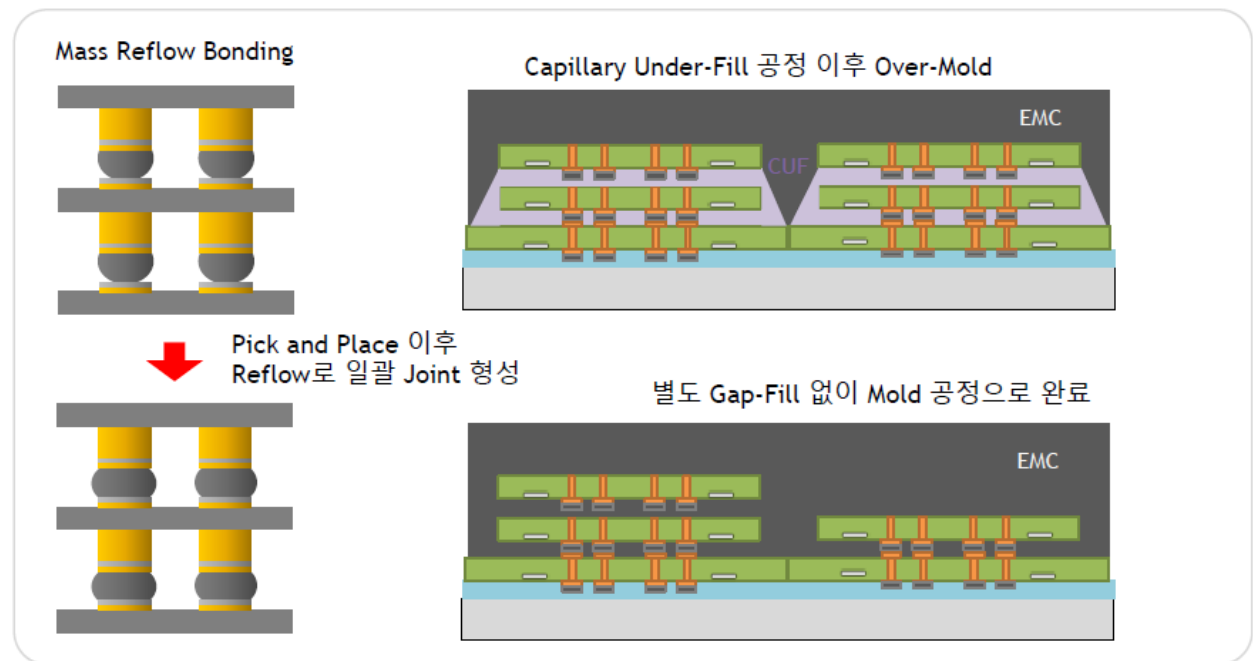
Source : AMD

MR+MUF는 높은 생산성과 Thermal Bump 채용으로 인한 Thermal Dissipation 특성 극대화 가능

TC+NCF



MR+CUF/MUF

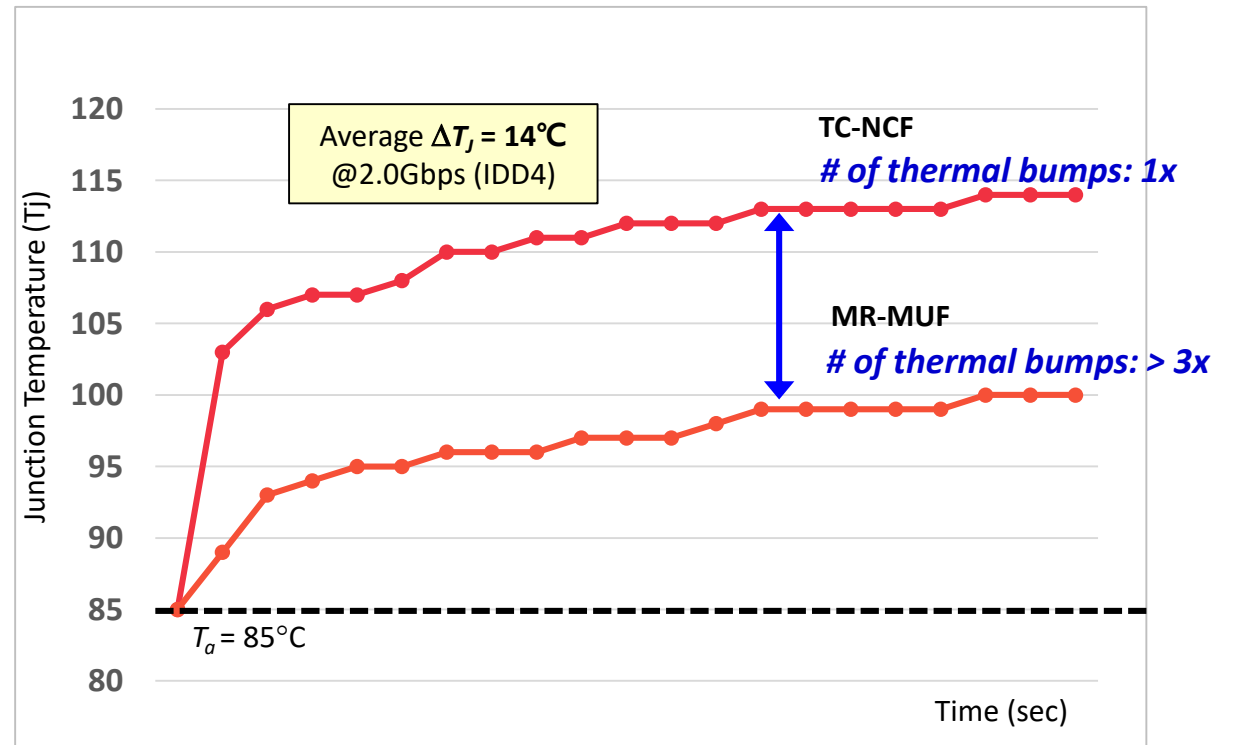
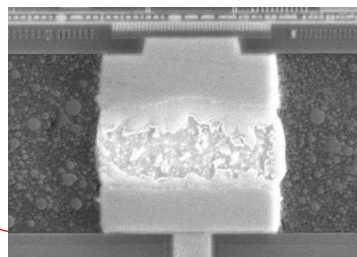
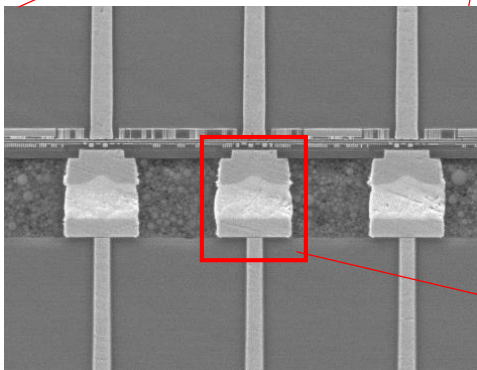
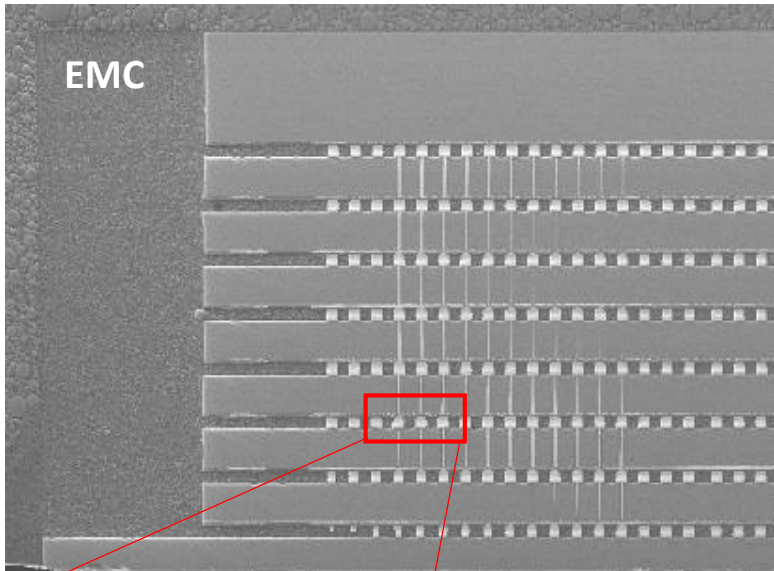


TC+NCF (Thermo-Compression + Non-Conductive Film)

MR+CUF/MUF (Mass Reflow + Capillary/Molded Underfill)

Advantage of MR+MUF Technology

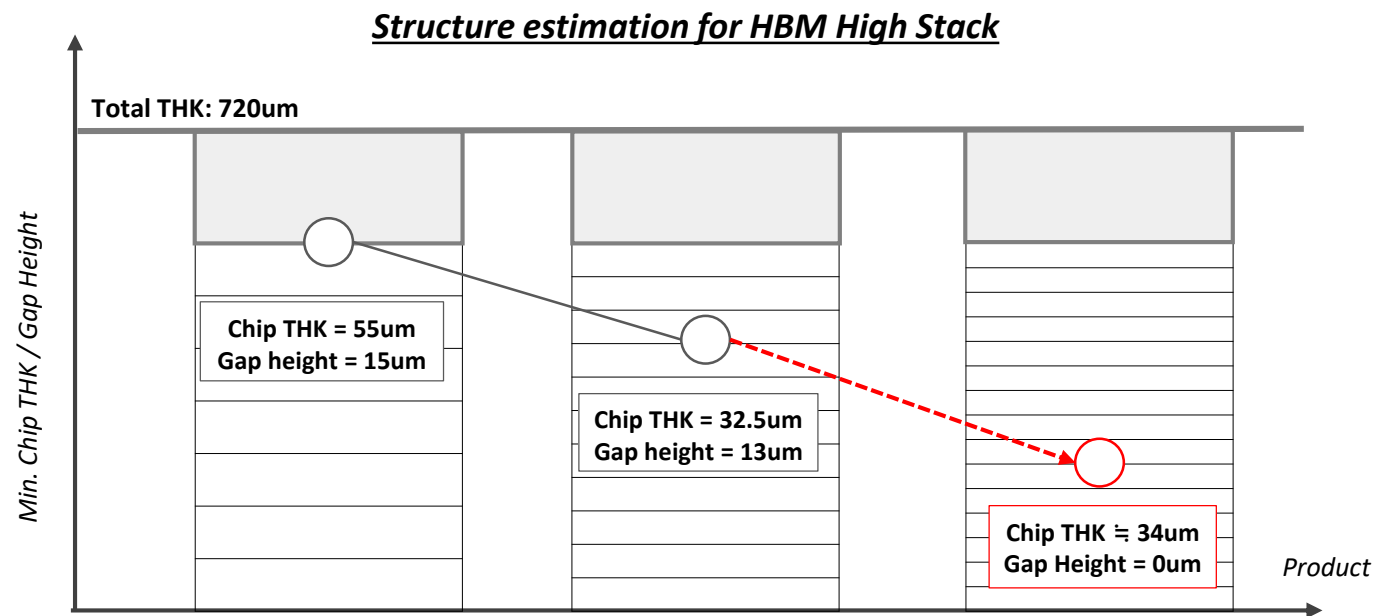
TC+NCF 대비 더 많은 수의 Thermal Dummy Bump 채용, MUF의 높은 열전도도에 의해 HBM의 열방출 특성 극대화, System이 HBM을 보다 효과적으로 사용 가능



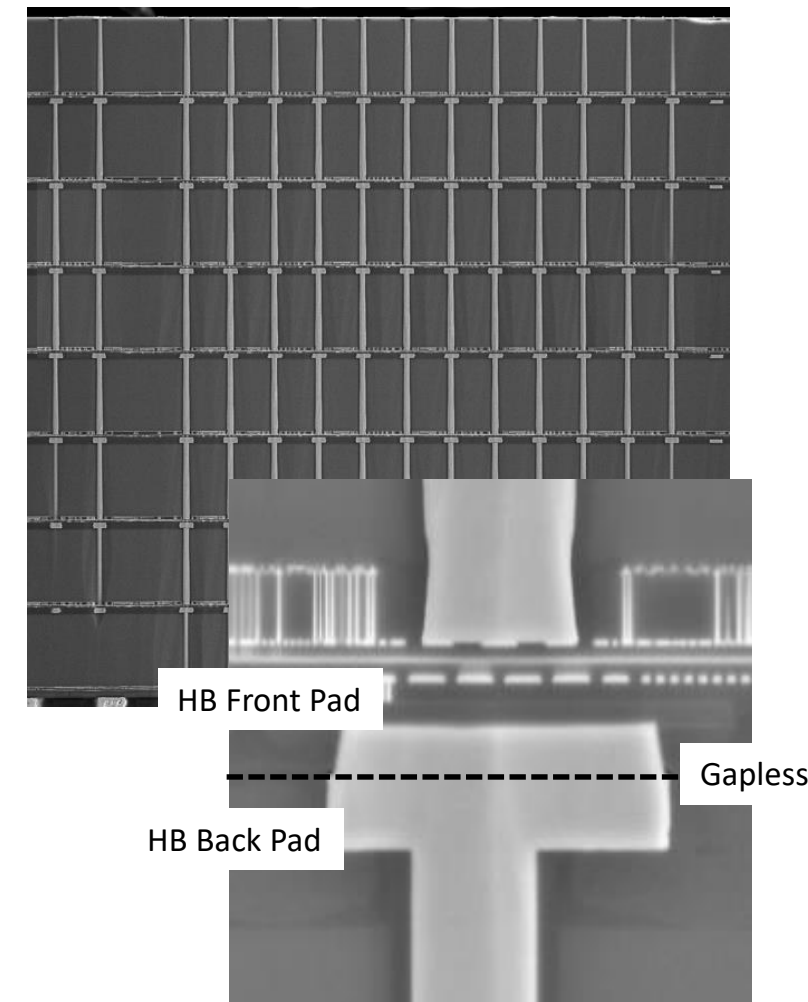
Test condition: Pin speed 2.0Gbps, VDD=1.26V, IDD4, Full Channel

Lower Junction Temp. (T_j) by Improved Thermal Resistance (> 30%↓)

16Hi 이상 HBM Solution으로 Gap-less Hybrid Bonding 기술 개발 중으로 HBM4 이후 채용 예상

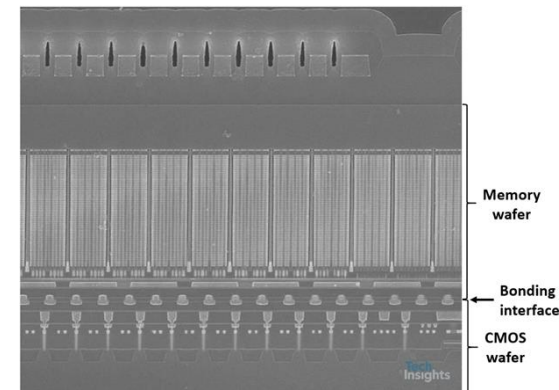
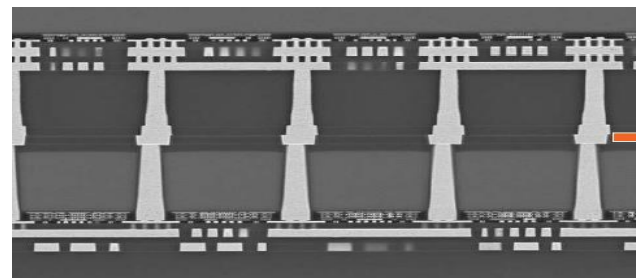


Gap-fill method	MR-MUF	MR-MUF	Hybrid bonding
Stack Height	8Hi (HBM3,HBM3E)	12Hi (HBM3,HBM3E)	16Hi (HBM4)
Bump Pitch	25um	22um	<20um



Other Applications of Hybrid Bonding

Logic-Logic, Logic-Memory 등 이중 Die 적층, CMOS-Cell Wafer 등 기존 Wafer를 분할, 재구성하여 적층하는 경우 Hybrid Bonding 기술 적용 가능



	3D Logic-Memory Stack	Core (Cell) - Peri Stack
Die 구성	이중 Die	Die 분할 이중 Wafer
대표 Application/기술	TSMC SoIC, Intel Foveros, Samsung X-Cube	YMTC Xstacking
I/O Pitch	수 um	Sub Micron
적층 형태	C2W	W2W
접합 방식	Micro Bump → Cu-Cu Hybrid Bonding	Cu-Cu Hybrid Bonding

Source : TSMC, Tech Insights (YMTC X-stacking)

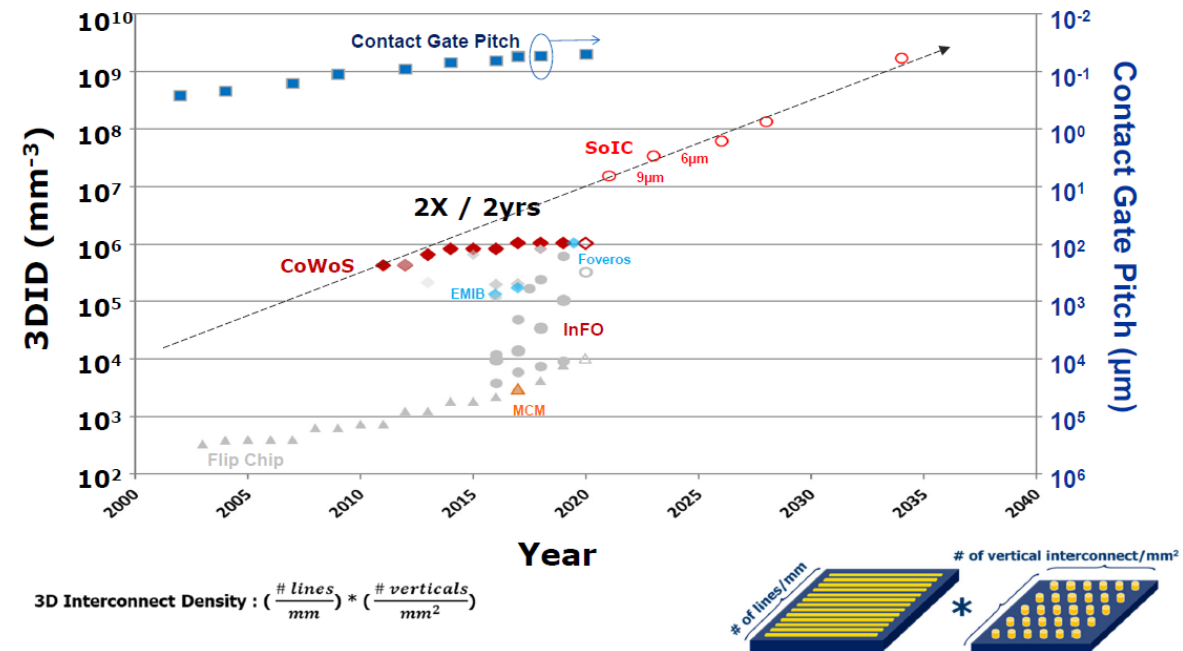
Why hybrid Bonding in LSI?

Micro-Bump 대비 Interconnect Density의 획기적인 증가, Latency/Power Consumption 감소 가능

Technology	3D-IC F2B	3D-IC F2F	SoIC™ F2B	SoIC™ F2F
Structure (Cross-Section)				
Bond Density*	1.0X	1.0X	16.0X	16.0X
Speed**	0.1X	1.0X	3.7X	11.9X
Bandwidth Density***	0.1X	1.0X	59.7X	191.0X
Power Consumption (Energy/bit)	3.7X	1.0X	0.6X	0.05X

* 3D-IC bump pitch: 36 μm; SoIC bond pitch: 9 μm

** Speed: 1/total wire delay; SoIC bond on top metal; *** Bandwidth Density: Bump Density*Speed

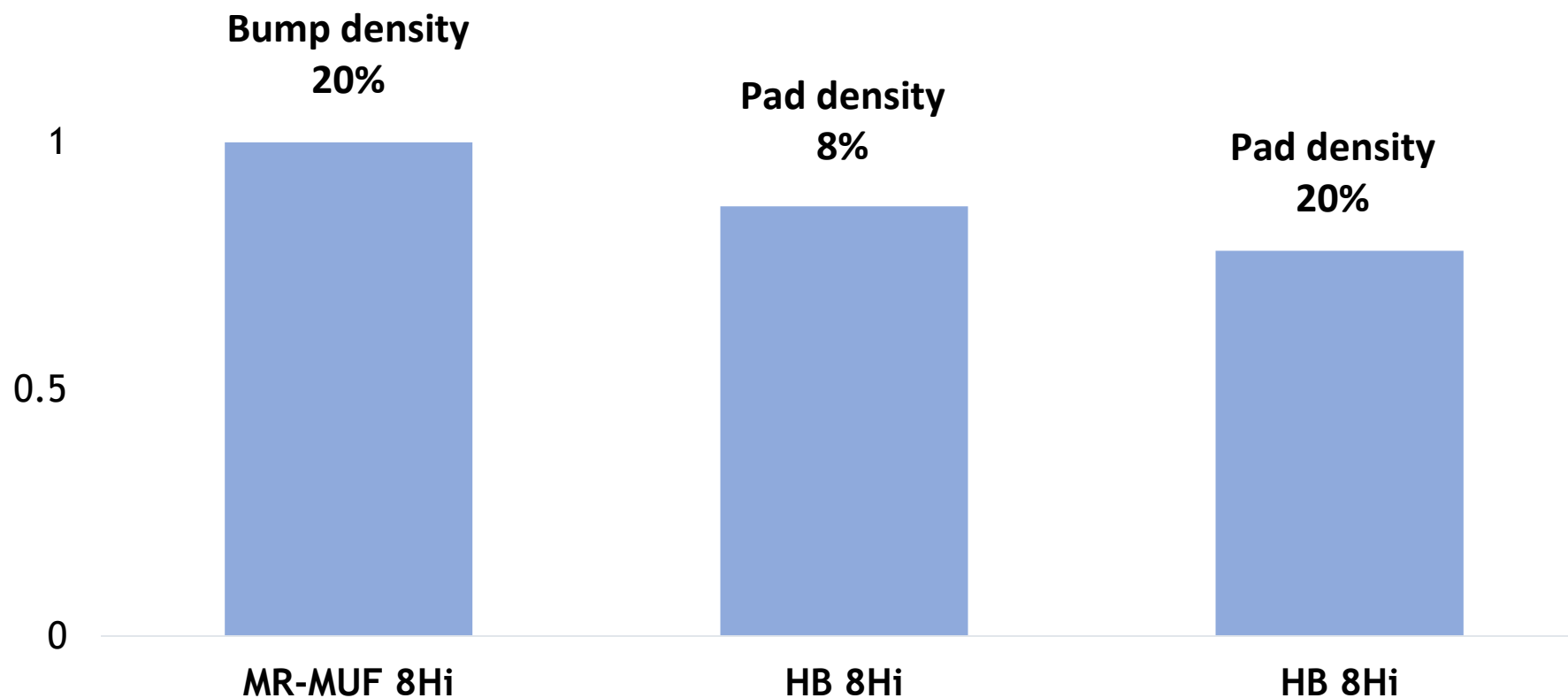


Source : Douglas Yu (TSMC), ISSCC2021

Thermal Resistance of HBM with Hybrid Bonding

MUF 등 Polymer 소재 미사용으로 낮은 Bump/Pad Density에서도 낮은 열저항을 보임

Thermal Conductivity : Nit 1.7, Ox 1.4, EMC 0.8~1.2 W/mK



Thin & Multi-Die Stacking에 Hybrid Bonding 적용 시 많은 기술적 Challenge들이 있음

Roughness
Dielectric
Cu
Dishing
Rounding

평탄도, 접합 강도

Bonding Pad 구조 및 접합 수율

- ✓ CMP WIW/WID Uniformity (Passivation)
- ✓ Inter-stack topology
- ✓ Stack Void (Topology oriented)

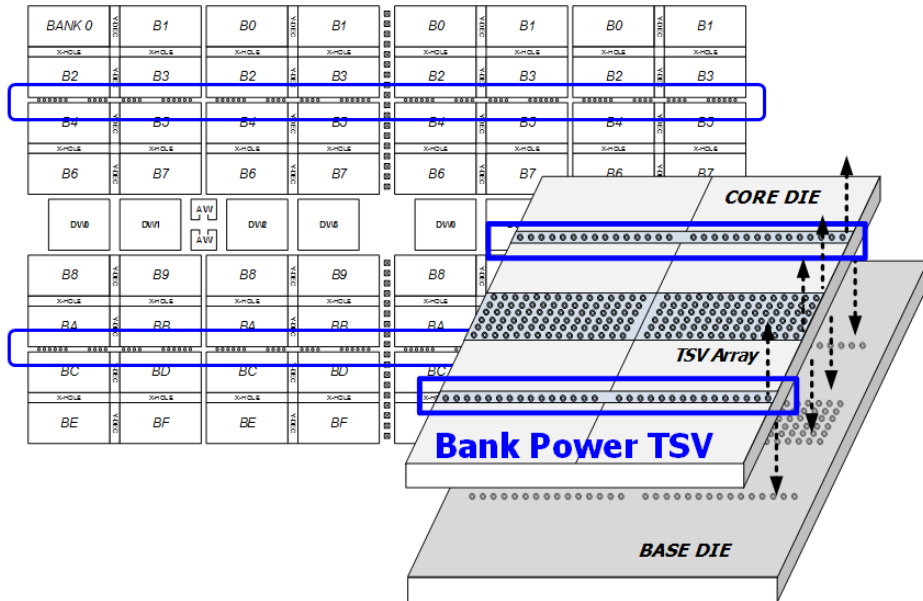
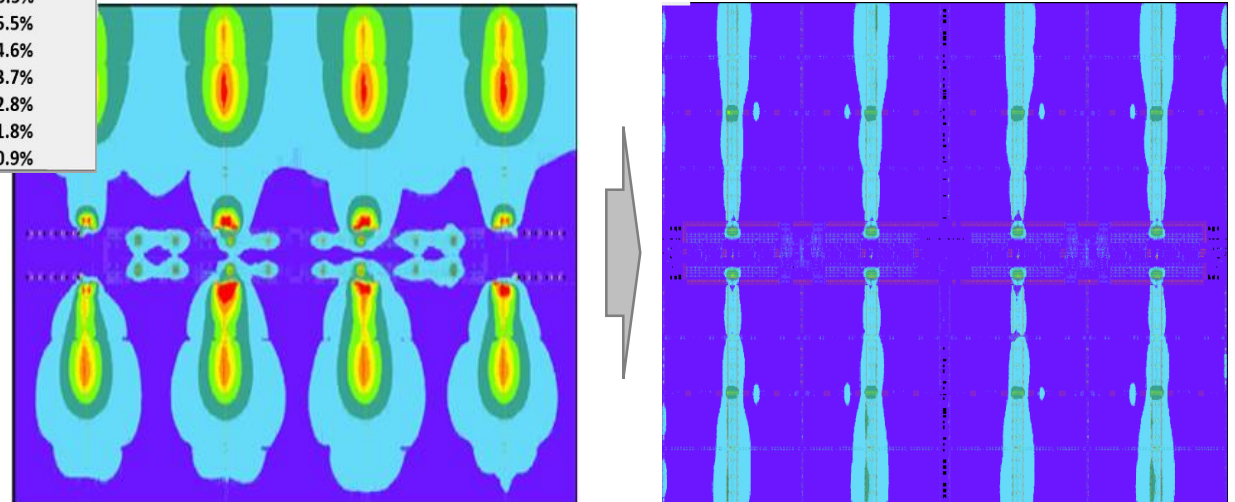
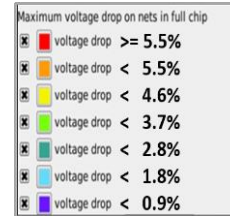
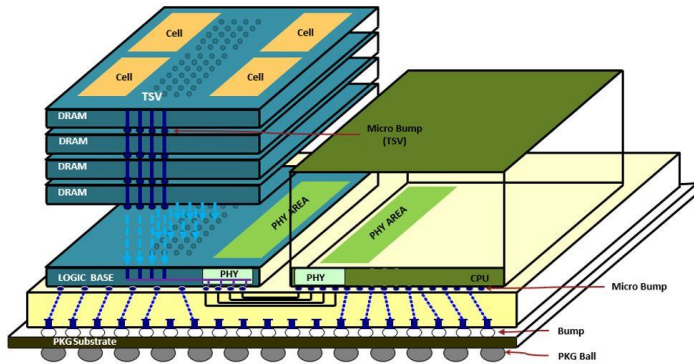
- ✓ CMP Cu Recess / Dishing Control
- ✓ Surface Activation & Plasma Treatment

Particle Control

- ✓ Advanced Dicing Solution
- ✓ Particle Cleaning and Inspection

Power Delivery in HBM stacked chips

Stack 수 증가 등에 의해 HBM 내의 PDN Issue는 갈수록 심화, Bank Power TSV 배치 시 개선

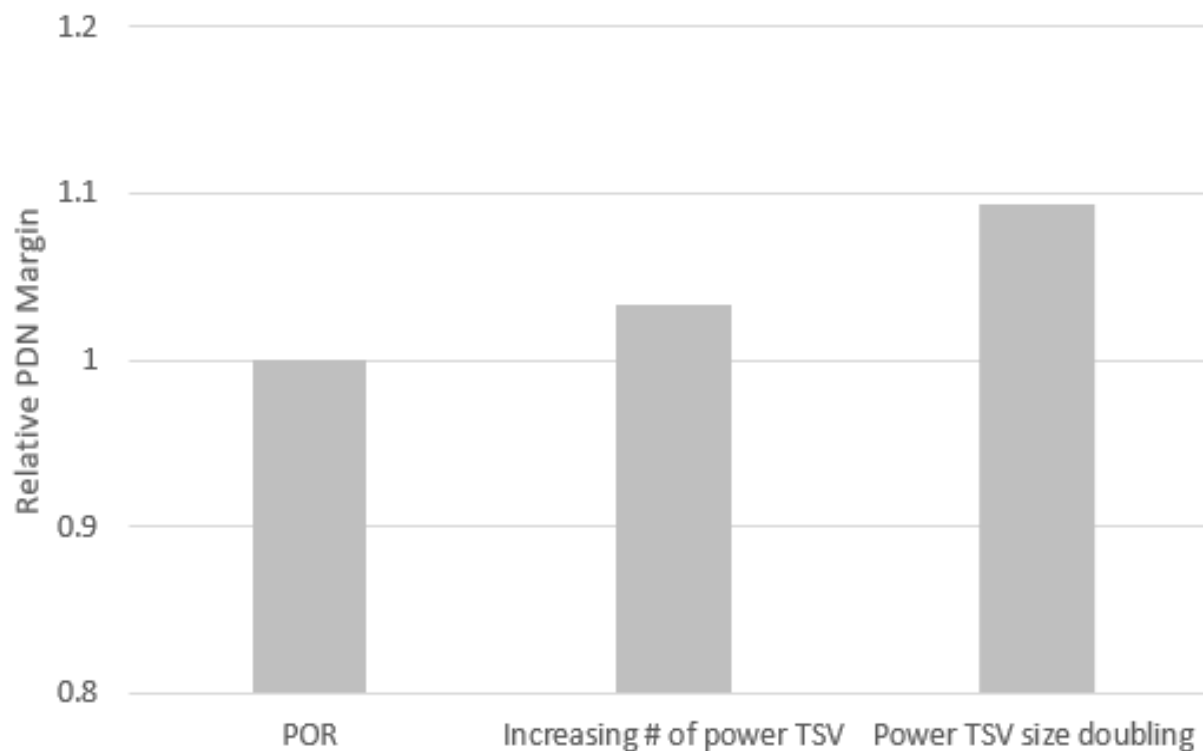


- ✓ IR Drop causes tCCDL degradation
- ✓ Due to Bank Power TSV, More than 50% of IR drops are diminished in IDD4W (worst pattern for Core IR Drop)

Source : J. H. Kim (SK Hynix), ISSCC2018

Power Delivery in HBM stacked chips

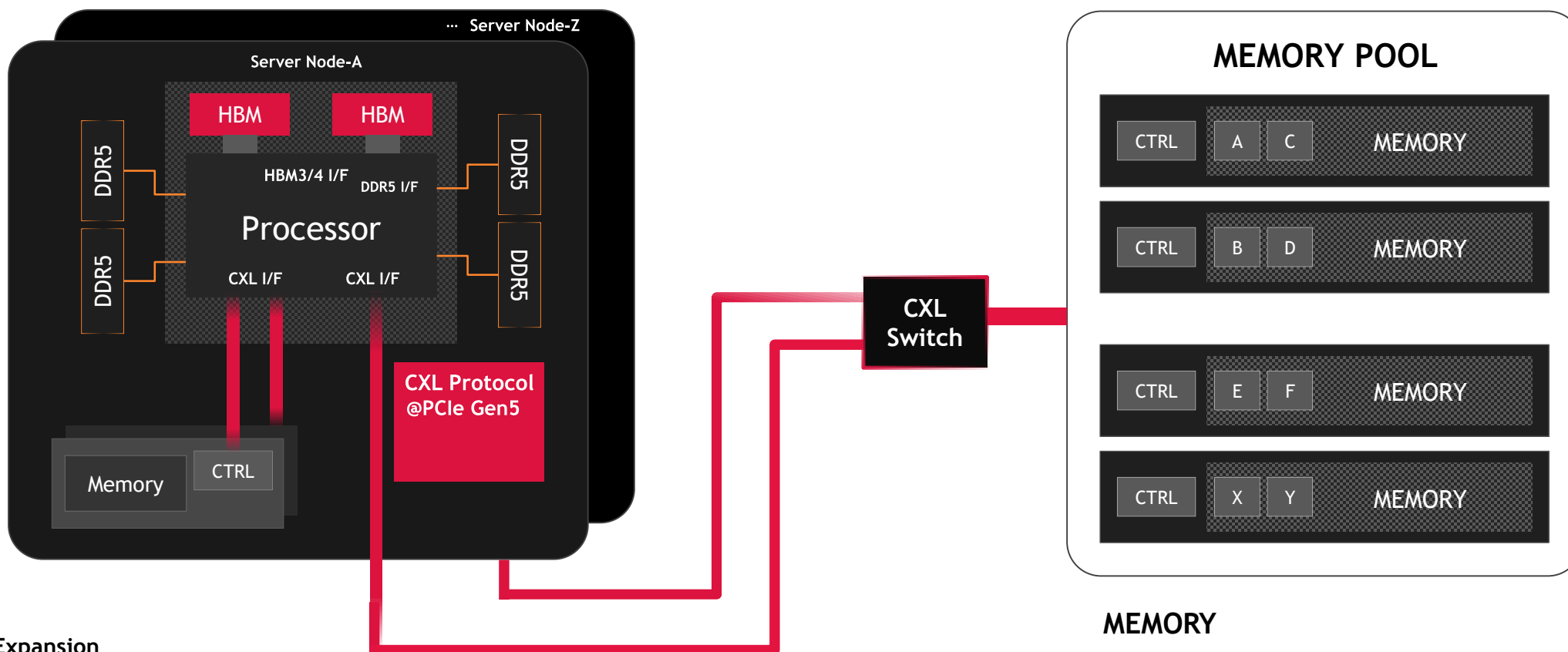
Power TSV의 개수를 늘리거나 Size를 늘릴 경우 PDN 개선 가능
Process 측면 Side Effect (Cost, Yield)에 따라 적용 여부 결정



CXL as Pooled Memory Solution

DRAM Channel 한계, Power 극복을 위한 CXL (Compute Express Link) 메모리 활용 예상

Memory Controller + 다양한 형태의 Memory가 결합된 형태



MEMORY USAGES

- Memory Bandwidth Expansion
- Memory Capacity Expansion
- Storage Class Memory

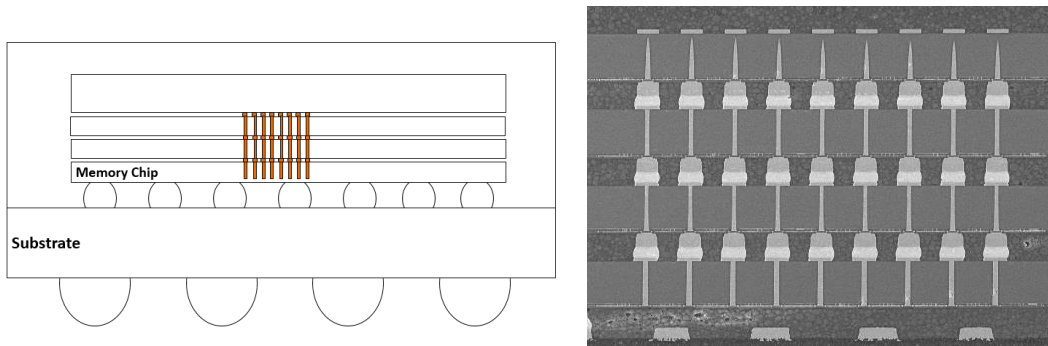
MEMORY

DDR4/5, MDS, PCM ...Etc.

Low Cost 고용량 메모리 Package 등 CXL 활용에 대비한 Package 기술 수요 예상

Bandwidth Solution

- Flip Chip (SDP), 3D/TSV (DDP/QDP) 활용
- Parasitic Loading 최소화를 통해 SI 특성 확보

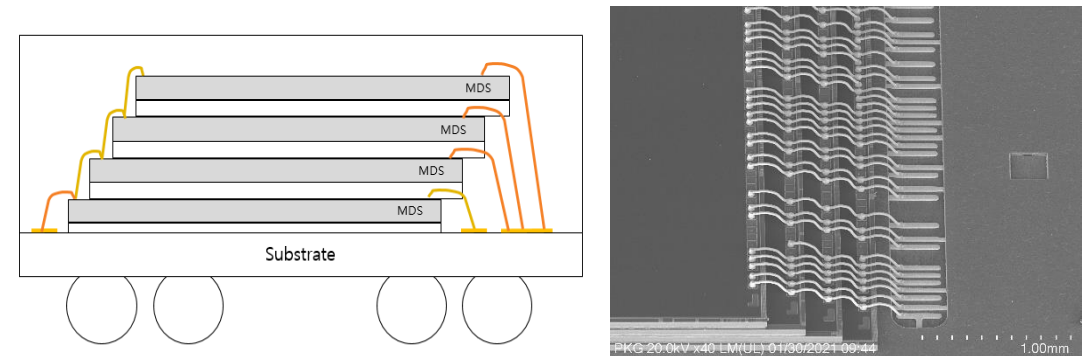


Packaging Challenges

- Cost Effective 3D/TSV 적층 기술
- DDP/QDP 용 (Dual/Quad Die Package) 대안 구조

Capacity Solution

- MDS (Managed DRAM Solution) 활용 저가, 고용량 구현
- Wire Bonding을 이용한 다단 적층

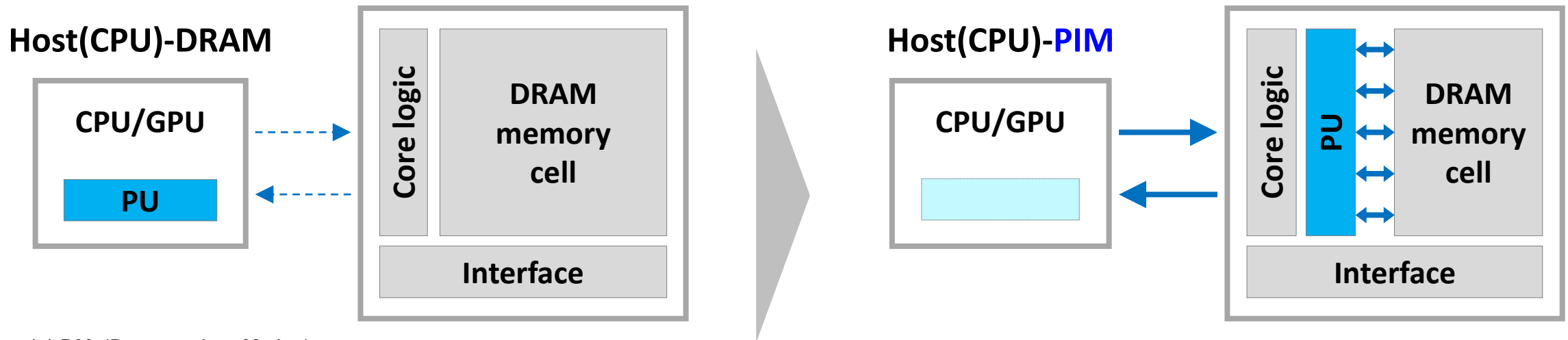


Packaging Challenges

- Long wire 기인 특성 저하 해결
- Cost-Performance Co-Optimization

PIM (AiM) Application

System Load를 줄이기 위해 Memory Intensive Operation을 Memory 혹은 Memory 근처로 이동
 Process Unit + DRAM Cell 이종 결합 고려



※ PU (Processing Units)

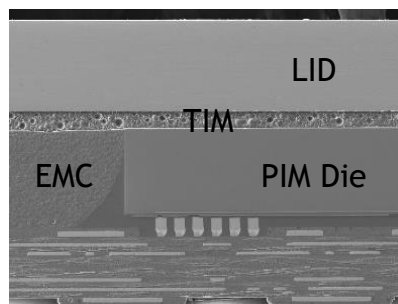
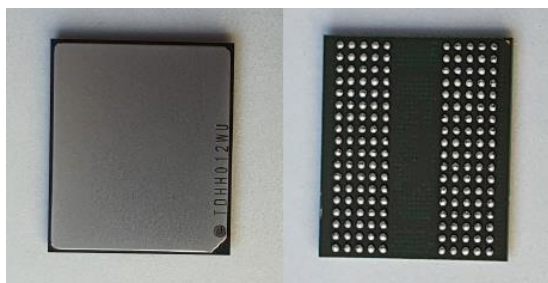
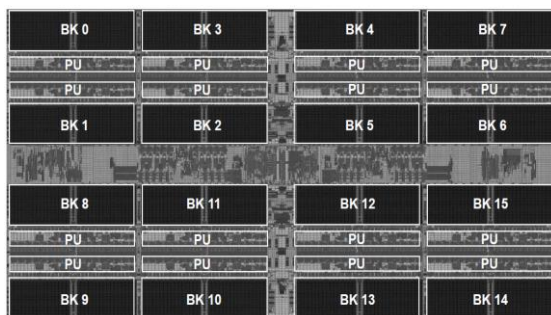
- I/O Speed, 외부 연결의 Bandwidth 동일
- Power Efficiency 및 성능 향상

SK Hynix's GDDR6 AiM with Lidded Package

GDDR6에 PU 탑재 PIM 구현, 기존 대비 X8 전력 효율, X10 특성 향상된 성능 구현
 Power 증가로 새로운 PKG Thermal Solution 필요, 메모리 최초 Lidded PKG 도입

GDDR6 Based AiM (PIM)

“A 1nm 1.25V 8Gb, 16Gb/s/pin GDDR6-based Accelerator-in-Memory supporting 1TFLOPS MAC operation and various activation functions for Deep Learning Applications” (ISSCC 2022)

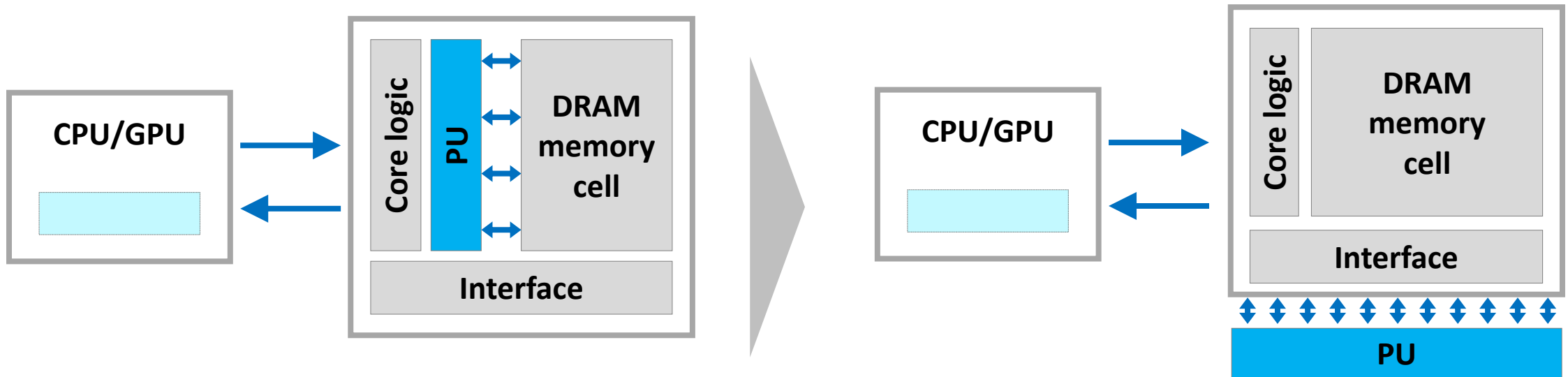


	Over Mold Type	Lidded PKG
Structure		
Sim.		
Speed	16Gbps	20Gbps
θ_{JC}	1.6 °C/W	0.5 °C/W
Measured T _j (@ATE)	100 °C	94 °C
Remark	EMC (k=3.0W/m·K)	t (LID 0.3mm, TIM 0.08mm)

Source : Seungju Lee (SK Hynix), ISSCC2022
 Ho-Young Son (SK Hynix), VLSI Symposium 2022

PIM에서의 Potential Package Architecture

Expanded Memory-Intensive Operation 구현을 위한 미래 PIM에서 고려해볼 수 있는 한 형태 (Example)



Power Efficiency **X8**
Performance **X10**
Inference Only

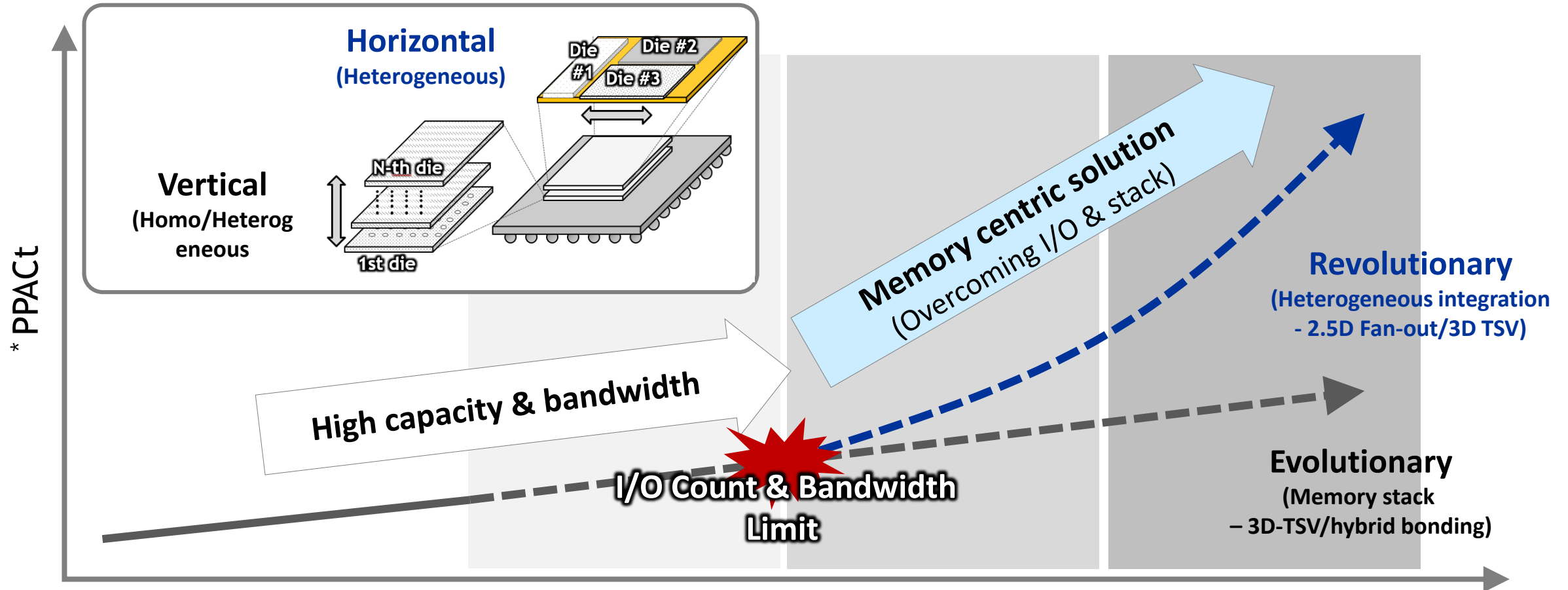
Flip Chip PKG 기술 활용

Power Efficiency **>X15**
Performance **>X10**
특성 (ex. Training)

DRAM - PU 간 Interconnect 수 급증 예상되며 새로운 형태의 Interconnect, Packaging 기술 필요할 것임

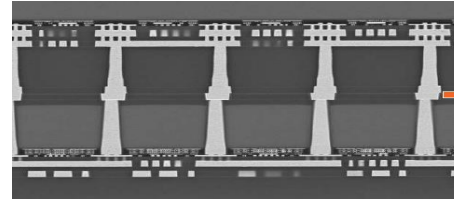
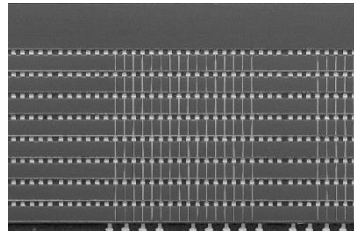
Heterogeneous Integration에서의 Advanced Package 기술의 역할

가까운 미래에 I/O Count와 Bandwidth 한계에 봉착할 것



* PPACT(Performance, Power, Area, Cost, Time to market)

기술 요구 수준에 따라 Architecture (이종 Die 혹은 Device 분할), 적층 방식 (C2W/W2W) 등 구분

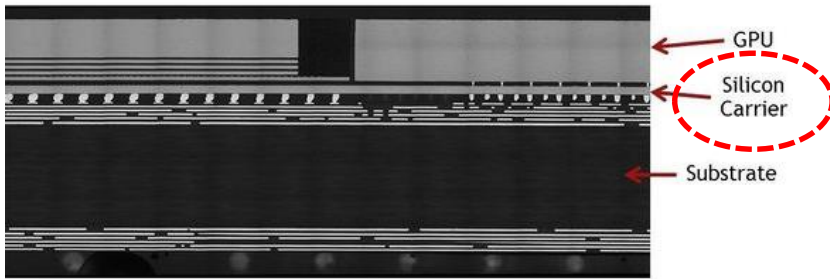


	3D Memory Stack	3D Logic-Memory Stack	Core (Cell) - Peri Stack
Die 구성	동종/이종 die	이종 Die	Die 분할 이종 Wafer
대표 Application/기술	HBM	TSMC SoIC, Intel Foveros, SS X-Cube	Nand ISS, 차세대 DRAM
I/O Pitch	Min. 22um	Min. 6um	Sub Micron~수 um
적층 형태	C2W (HBM Base-Core) W2W (HBM Core)	C2W	W2W
접합 Material	Micro-Bump (20~50um), Cu-Cu HyB	Micro-Bump (25~30um) Cu-Cu HyB (6~9um)	Cu-Cu HyB
비고	Pitch는 HBM3-A 적용 기준	Foundry 중심 전개	DRAM Speed 한계 대비

2.5D Integration 기술

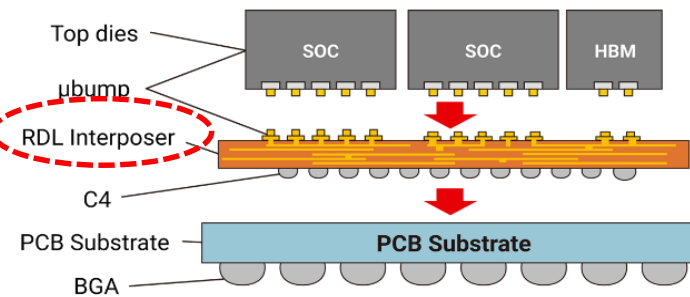
수평 방향의 Die-to-Die 연결 기술로 Si Interposer, RDL Interposer, Si Bridge 등 활용
HBM-xPU (SoC), SoC-SoC Integration에 활용

Si Interposer Based



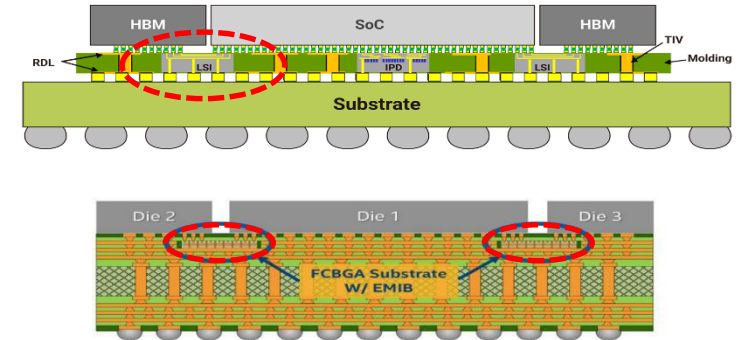
- (+) Matured Technology
(Widely Used)
- (-) High Interposer Cost
- (-) RC Delay by Si Sub
- (-) CoW Warpage

RDL Interposer Based



- (+) Lower Cost and Eco-Friendly
- (+) Enhanced SI Performance
- (-) Looser Design Rule than Si I/P
- (-) Technology not matured

RDL/Advanced Sub + Si Bridge



- (+) Enhanced SI Performance
- (-) Highly Difficult in Process Integration

주요 업체의 2.5D/3D Integration 기술

Back-End Integration

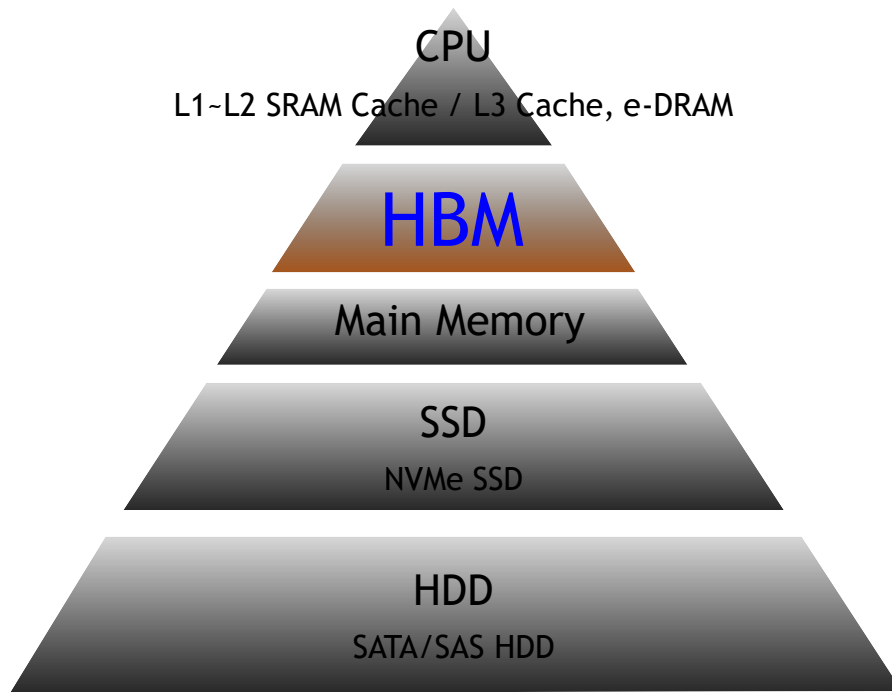
Front-End

2D (AP Only, AP+PMIC)		<ul style="list-style-type: none"> InFO (Integrated Fan-Out) : InFO, -M, -oS, -LSI, -SoIS 	<ul style="list-style-type: none"> ePLP (Panel Level Package) 	
	2.5D (Logic + HBM)	<ul style="list-style-type: none"> Si Interposer 	<ul style="list-style-type: none"> I-Cube S, H-Cube 	
	<ul style="list-style-type: none"> RDL Interposer / or Substrate 	<ul style="list-style-type: none"> CoWoS-R, CoWoS-L 	<ul style="list-style-type: none"> I-Cube E 	<ul style="list-style-type: none"> EMIB, Co-EMIB
3D (SoC+SoC, SoC+SRAM)		<ul style="list-style-type: none"> SoIC : SoIC-P, SoIC-X, CoWoS + SoIC 	<ul style="list-style-type: none"> X-Cube 	<ul style="list-style-type: none"> FOVEROS, : Foveros, -Direct, Foveros-Omni

Source : TSMC, Samsung, Intel Website

HBM과 같은 Memory가 새롭게 등장한 것처럼, 가까운 미래에도 현재의 한계를 극복하기 위한 Solution이 등장할 것이며, Advanced Packaging 기술은 이를 구현하는 필수 요소가 될 것임

Present



Future

1. Scaling Limitation of **SRAM**

- SRAM과 HBM 사이의 Gap Filling Solution

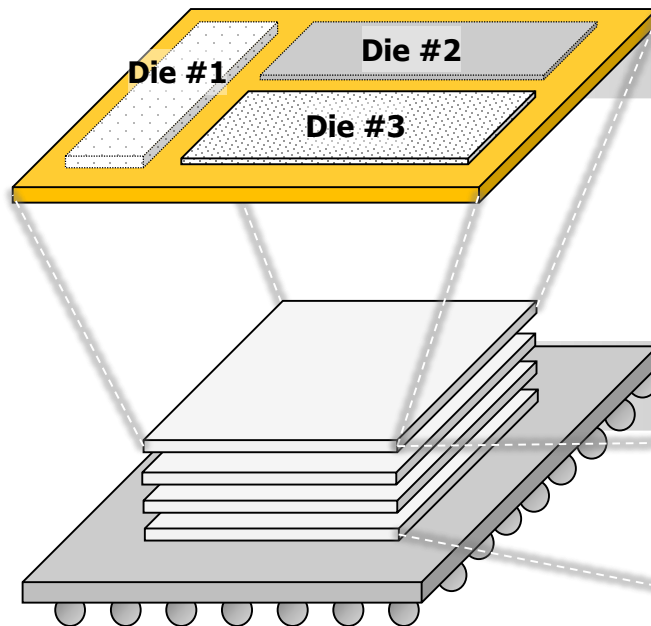
2. Opportunities of **CXL, PiM**

- Memory offloading, emerging interface 필요성

3. **Memory Controllers**의 중요성 증가

- Application의 증가 (Design resource)
- Leading Tech Node 적용에 따른 Cost Adder

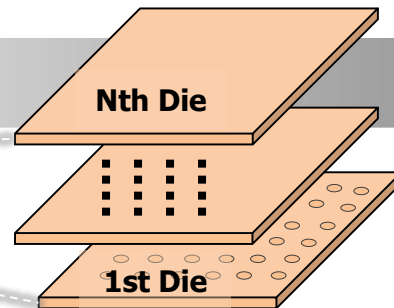
Memory를 넘어 2.5D Fan-out Packaging, 3D Hybrid Bonding 기반 Heterogeneous Integration 기술로의 확장



이종칩 집적 기술 응용

Fan-out 기반 수평 연결 – Chiplet, SiP 기술

- Memory Centric 융복합 Solution
- High-End 공정 및 Memory Device 한계 극복



수직 적층 기술 고도화

TSV 기반 Memory 적층 - Hybrid Bonding 기술

- Bandwidth, Capacity Scaling 극대화
- Thermal Performance, Power Efficiency 극대화



Thank you

Growing together

for better tomorrow

For further questions: hoyoung.son@sk.com