



Advancements in HBM Process Technology Boosting Bandwidth and Stacking Layers for the Future

Presenter: Tom Hsu/Senior Analyst



HBM Front-end Process Migration to Boost Bandwidth and Mono Die Density

- Increase of I/O requires more TSV numbers (front-end process), and also bumps/pad pitches (backend process).
- Increase of mono die density requires front-end process migration.
- Die size continue to become larger in HBM by around 15%, leading to higher cost per Gb.

HBM Bandwidth Roadmap

		<i>HBM2e</i>	<i>HBM3</i>	<i>HBM3e</i>	<i>HBM4</i>	<i>HBM4e</i>	<i>HBM5</i>	<i>DDR5</i>
<i>Bandwidth per Stack (GB/s)</i>		410-461	717-819	1024-1254	2048-2560	>3072	TBD	2.4-3.2
<i>Speed per Pin (Gbps)</i>		3.2-3.6	5.6-6.4	8.0-9.8	8.0-10.0	>12	TBD	4.8-6.4
<i>I/O Bit per Stack (bits)</i>		1024	1024	1024	2048	2048	4096	4
<i>Channel per Stack</i>		8	16	16	32	32	TBD	1
<i>Bit per Channel (bits)</i>		128	64	64	64	64	TBD	4
<i>Mono Die Density</i>		16Gb	16Gb	24Gb	24Gb	32Gb	TBD	8-32Gb
<i>Process Node</i>	<i>SK hynix</i>	1y nm	1z nm	1b nm	1b nm	1c nm	TBD	
	<i>Samsung</i>	1y nm	1z nm	1a nm	1c nm	1c nm	TBD	
	<i>Micron</i>	1z nm		1beta nm	1beta nm	1gamma nm	TBD	

HBM Backend Process Technology Shows Migration to Boost Number of Stacking Layer

- SK hynix maintains its leading position in production yield and product performance by developing MR-MUF process for HBM2e, which was upgraded to Advanced MR-MUF process for 12hi products with adoption of TC bonding.
- For 16hi, it remains uncertain if suppliers will adopt hybrid bonding process or continue using TC-NCF and MR-MUF. Flux-less TC bonding is also considered.
- For HBM4/4e, with 775µm limit in total height, joint gap height of 16hi becomes around 5µm, with core die height of 20-25µm.
- Expansion of production scale and process complexity of HBM will boost equipment demand.

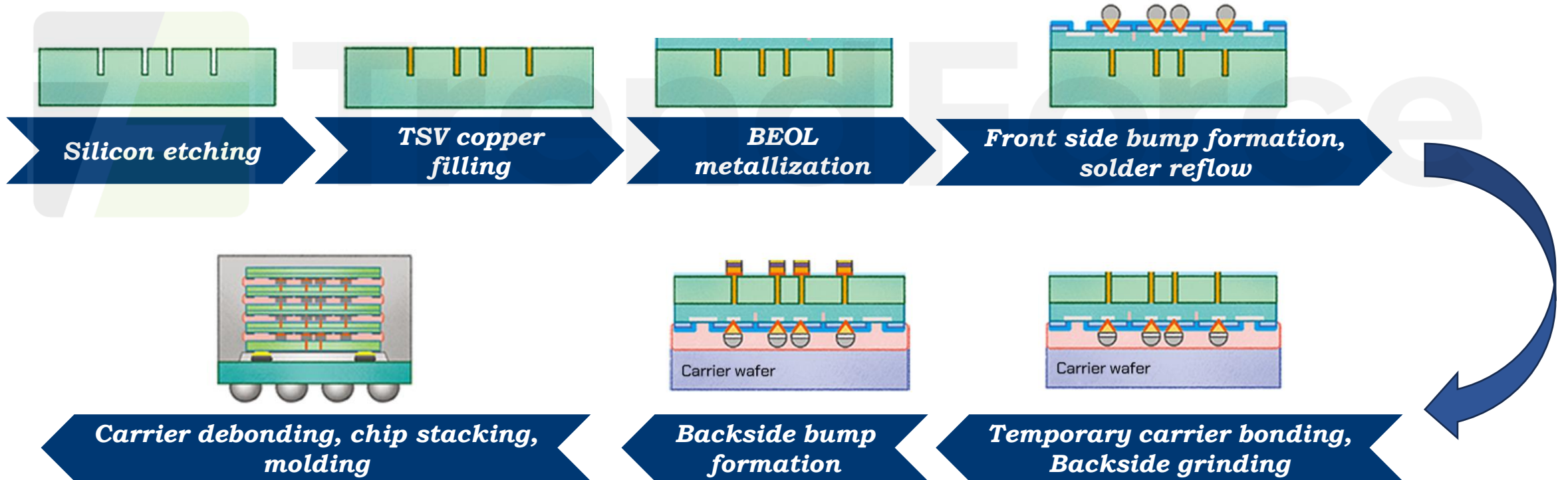
HBM Stacking Process Technology Roadmap

		HBM2e		HBM3		HBM3e			HBM4			HBM4e			HBM5	
Stacking		4hi	8hi	8hi	12hi	8hi	12hi	16hi	8hi	12hi	16hi	8hi	12hi	16hi	16hi	>20hi
NVIDIA AI GPU		Ampere		Hopper		Blackwell			Rubin			Rubin Ultra			TBD	
NVIDIA CoWoS HBM		5		5/6		8			8			8/12			3D stacking	
Stacking Technology	SK hynix	MR-MUF	MR-MUF	Adv. MR-MUF	MR-MUF	Adv. MR-MUF		MR-MUF	Adv. MR-MUF		MR-MUF	Adv. MR-MUF	TBD	TBD	TBD	TCB
	Samsung	TC-NCF	TC-NCF		TC-NCF		N.A.	TC-NCF		TBD	TC-NCF		TCB	TCB		
	Micron	TC-NCF	TC-NCF		TC-NCF		N.A.	TBD	TC-NCF	TBD	TBD			TBD	TCB	

HBM TSV and Backend Process Flow

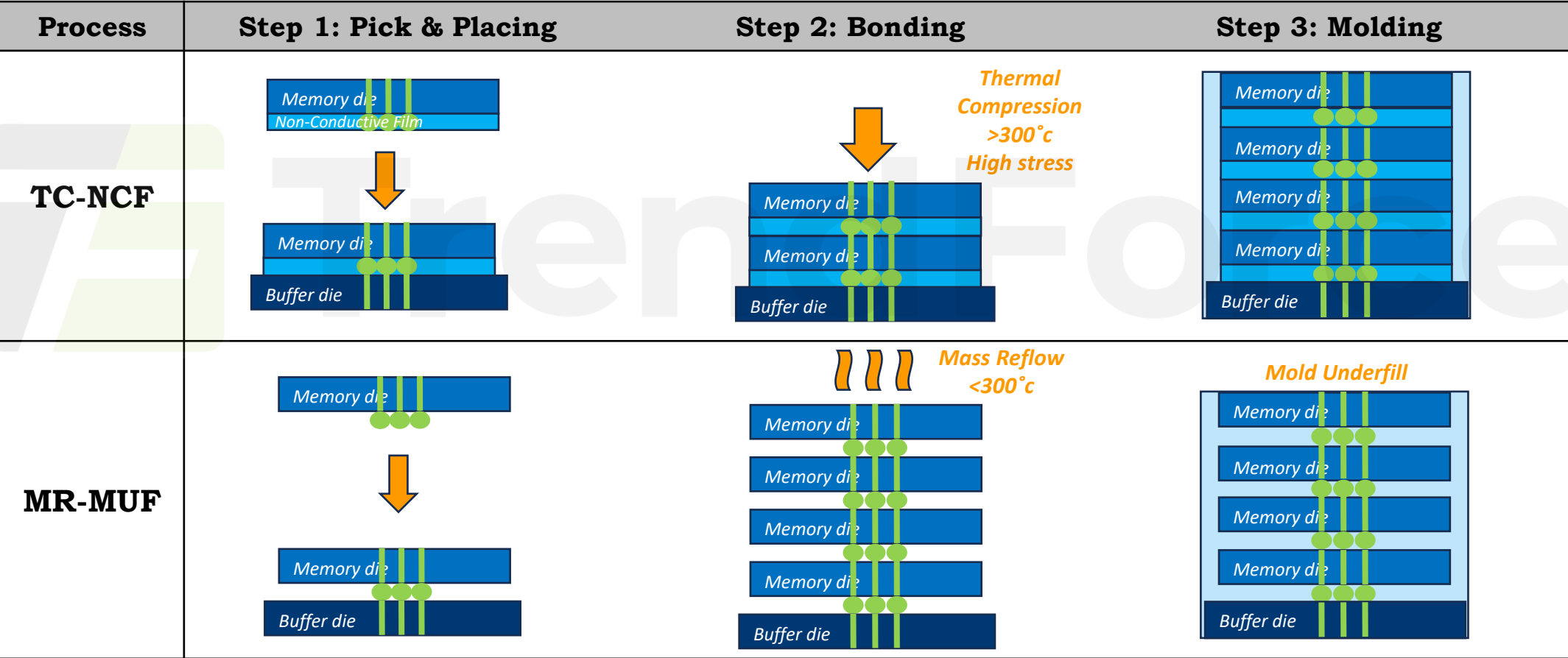
- HBM suppliers all adopt Via-middle TSV process.
- Key TSV front-end process steps include silicon etching, TSV copper filling and BEOL metallization.
- Key TSV backend process steps include temporary carrier bonding/debonding and backside grinding.
- Key stacking process steps include mass reflow, thermal compression bonding and compression molding.

HBM TSV and Backend Process Flow



HBM Stacking Process Technology

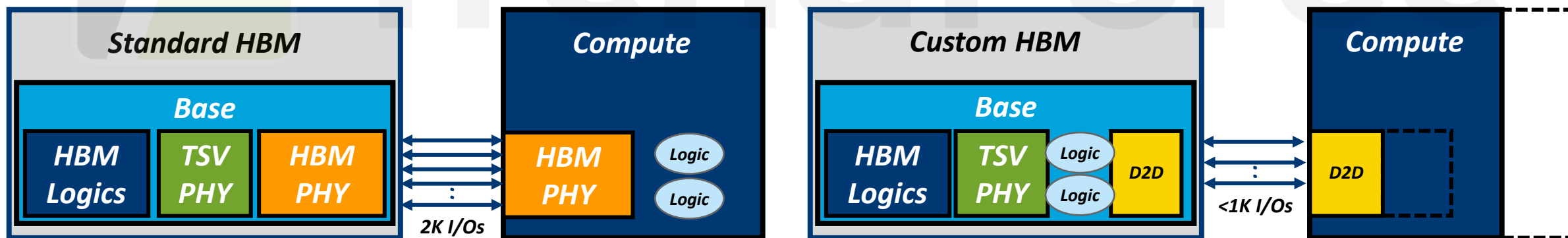
- TC-NCF process adopts thermal compression with high temperature and stress when stacking each layer, leading to longer cycle time.
- MR-MUF process only adopts mass reflow with low temperature when completing the stacking of all layers.
- Advanced MR-MUF process adopts thermal compression with low temperature and stress when stacking each layer, and implements mass reflow when completing the stacking of all layers.



cHBM Enables Better System-level Performance and TCO

- In standardized HBM (sHBM) structure, HBM PHY (physical layer) blocks are put in base die and GPU/ASIC as a physical interface.
- Compared to sHBM, customized HBM (cHBM) offers the customization of base die, which equipped with more advanced process node. The logic computing functions which have higher demand for instant connection with memory can be put in base die.
- Better system-level performance and TCO (Total cost of ownership) can be achieved by cHBM structure by:
 - less PHY area in HBM base die, saving die size;
 - less PHY area in GPU/ASIC, saving die size, which can be used for computing functions which requires more advanced process node;
 - less data flow, IOs and channel length (from 6mm to 2mm) between HBM and GPU/ASIC are required, saving power.

Comparison Between Structure of sHBM and cHBM



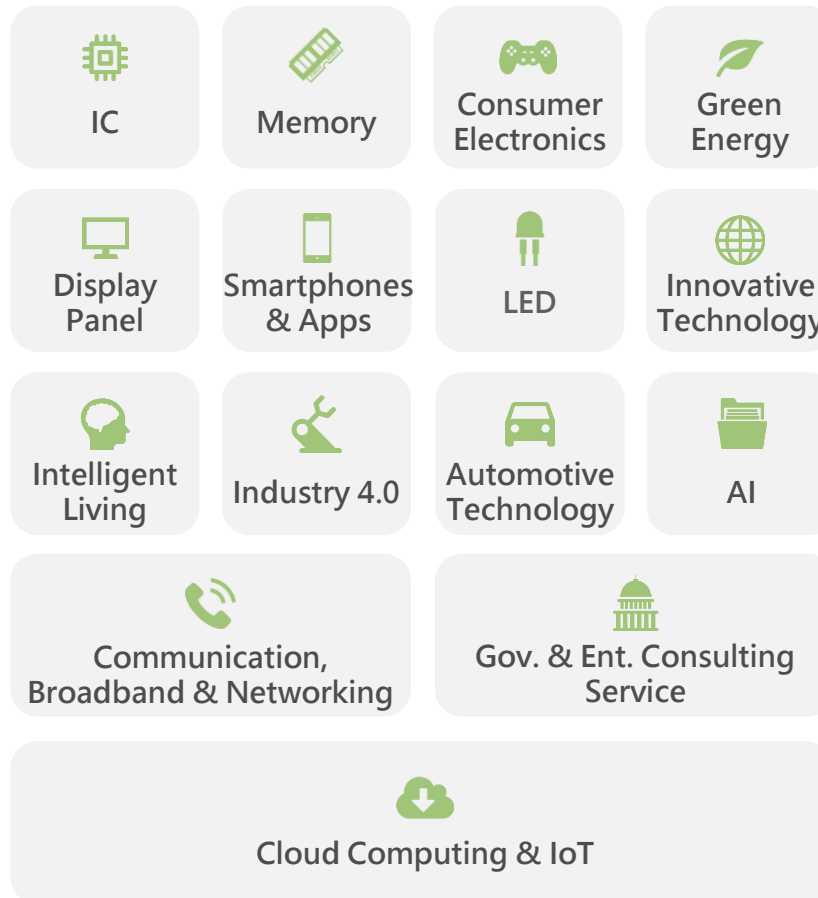
HBM Base Die Manufacturing to Rely on TSMC

- With enhancing computing power, FinFET and logic process will become a MUST for HBM base die from HBM4e generation. SK hynix and Micron, without inhouse FinFET technology, become dependent on external foundry services.
- cHBM base die requires more integration with GPU/ASIC. TSMC, having dominated the GPU/ASIC foundry market, is the best candidate to provide foundry services for cHBM base die.
- HBM suppliers tend to outsource base die bumping process to base die foundries and CoWoS packaging service providers.

HBM Base Die Foundry Supply Overview

		SAMSUNG	SK hynix	micron
HBM4	sHBM	4nm FinFET by Samsung System LSI	12nm FinFET by TSMC	Planar by Hiroshima Fab 15
	cHBM	FinFET by TSMC or Samsung System LSI	N.A.	N.A.
HBM4e	sHBM	FinFET by Samsung System LSI	12nm FinFET by TSMC	4nm FinFET by TSMC
	cHBM	FinFET by TSMC or Samsung System LSI	3nm FinFET by TSMC	FinFET by TSMC

TrendForce & TRI Research Areas



Sales & Services

Semiconductor Research

DRAM, NAND Flash, Foundry

SR_MI

SR_MI@TrendForce.com

Green Energy Research

Solar PV

GER_MI

GER_MI@TrendForce.com

Optoelectronics Research

Micro LED, Mini LED, VCSEL, UV, Video Wall, Lighting

OR_MI

OR_MI@TrendForce.com

Display Research

TFT-LCD , OLED , Smartphone , Tablet , NB , Monitor/AIO , TV

DR_MI

DR_MI@TrendForce.com

ICT Application Research

Communication & Broadband, Consumer Electronics, Innovative Technological Applications, Automotive, Industry 4.0, Gov. & Ent.

TRI_MI

TRI_MI@TrendForce.com

Thank You

To find out more about TrendForce, scan the QRCode below.

Or visit the URL for further information.



English
www.trendforce.com



Traditional Chinese
www.trendforce.com.tw



Simplified Chinese
www.trendforce.cn