Expanding the Boundaries of the AI Revolution:

An In-depth Study of High Bandwidth Memory

Changyong Ahn & Nayoung Lee | March 2019

Outline

- SK hynix Overview
- Memory challenges of Deep Learning
- HBM Overview
- HBM Deep Dive
- Future HBM solution

Memory challenges of Deep Learning

Machine Learning/Deep Learning Use Cases

Security

Data analysis

Financial Service

Law

Machine Learning

Autonomous Driving

Gaming

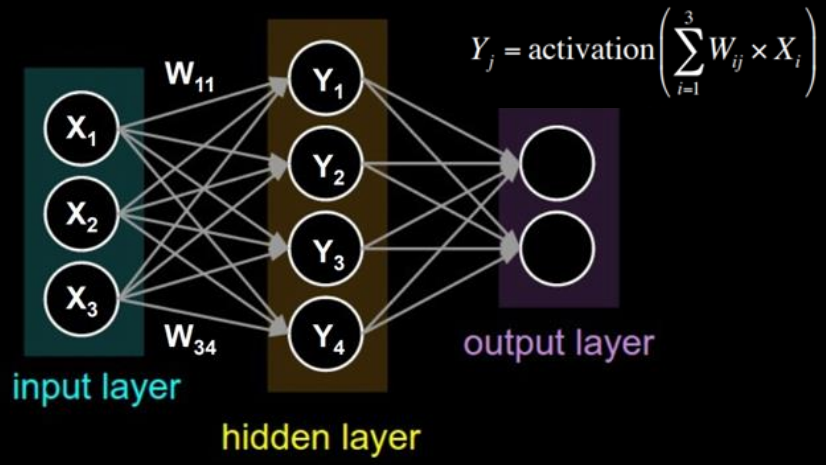
Medical Diagnostics

Smart home

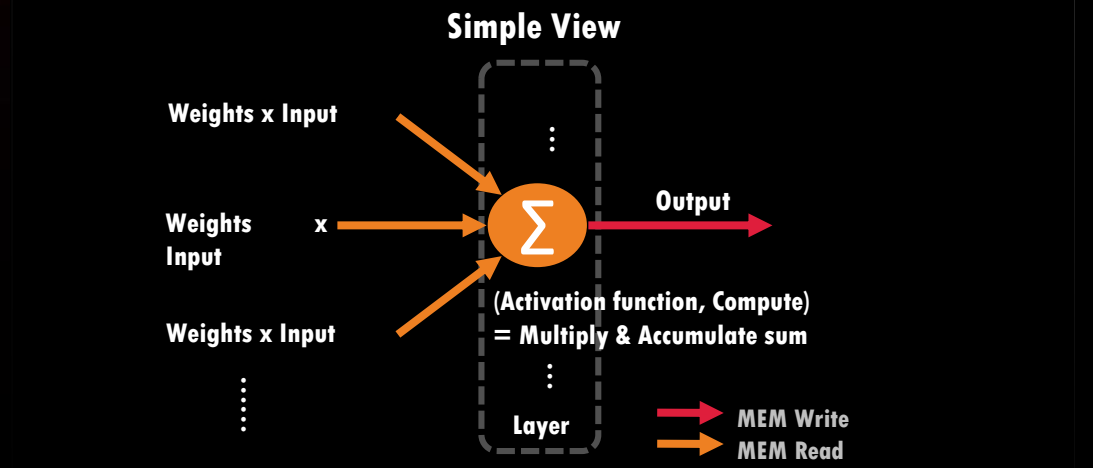
Memory Challenges of Deep Learning

Deep Neural Network Fundamental Concepts

Deep Neural Network



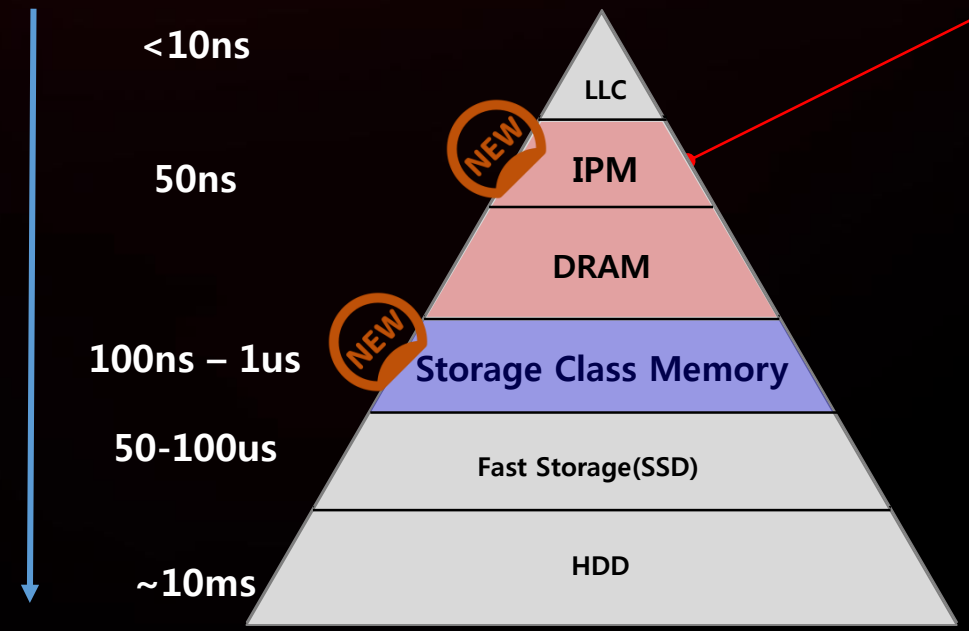
Source: Standford



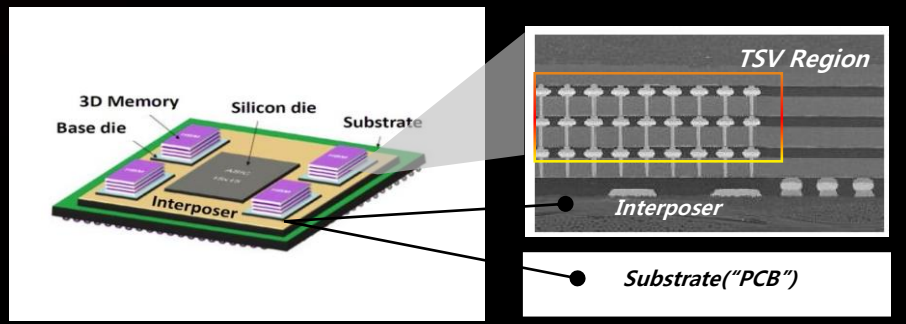
Year	CNN	# of layers	# of Parameters	Memory size (MB)	Top5 Error Rate
1998	LeNet	8	60K		
2012	AlexNet	7	60 million	240	15.3%
2014	GoogleNet	19	4 million		6.67%
2014	VGG Net	16	138 million	574	7.3%
2015	ResNet	50/152		519	3.6%

Memory Solution for ML/DL Systems

Memory Sub system hierarchy change



"In Package Memory"



	Conventional DRAM	IPM
Target Market/Price	Broad & Cheap	Specific & high **Premium**
Standardization	JEDEC	**Semi Custom**
Qualification Period	Relatively short	**Relatively long**
Key factors	Price Competitiveness	**Reliability / Performance**

* Source : SK hynix

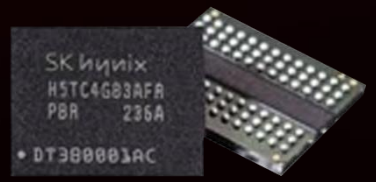
1) In-Package Memory 2) SCM("Storage Class Memory") : 3DXP, PCRAM

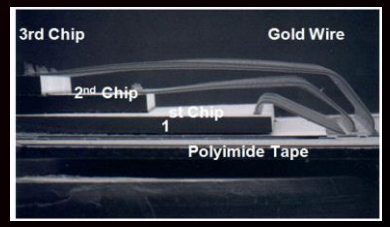
HBM Overview

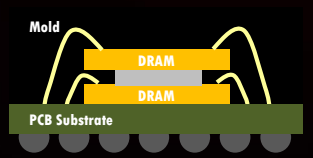
HBM, What's the difference?

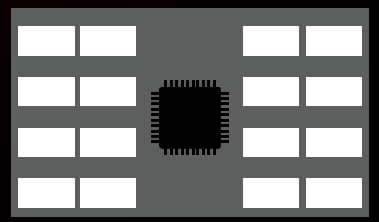
GDDR/DDR/LPDDR

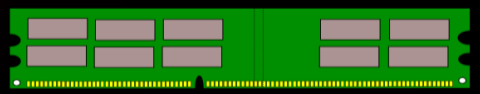
➢ FBGA







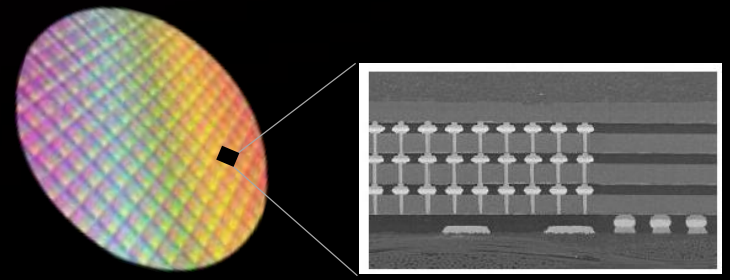




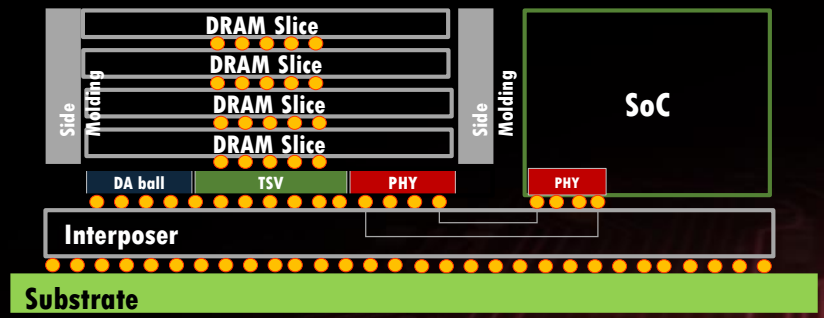
Directly soldered on PCB or used as a DI MM

HBM

➢ KGSD



➢ HBM in 2.5D SiP



HBM Advantages

More Bandwidth

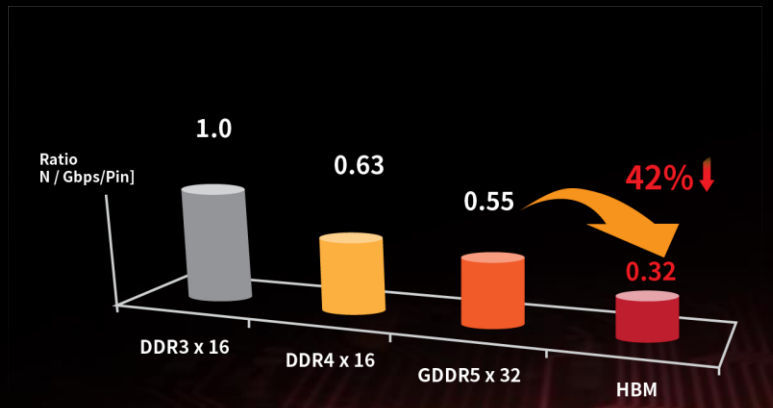
High Power Efficiency

Small Form Factor

To Achieve 1TB Bandwidth .....

40ea of DDR4-3200 Module | **160ea of DDR4-3200** | **4ea HBM2 in a single 50mm x 50mm Sip**

	DDR4	LPDDR4(X)	GDDR6	HBM2	HBM2E (JEDEC)	HBM3 (TBD)
Data rate	3200Mbps	3200Mbps (up to 4266 Mbps)	14Gbps (up to 16Gb ps)	2.4Gbps	2.8Gbps	>3.2Gbps (TBD)
Pin count	x4/x8/x16	x16/ch (2ch per die)	x16/x32	x1024	x1024	x1024
Bandwidth	5.4GB/s	12.8(17)GB/s	56GB/s	307GB/s	358GB/s	>500GB/s
Density (per package)	4Gb/8Gb	8Gb/16Gb/2 4Gb/32Gb	8Gb/16Gb	4GB/8GB	8GB/16GB	8GB/16GB/ 24GB (TBD)

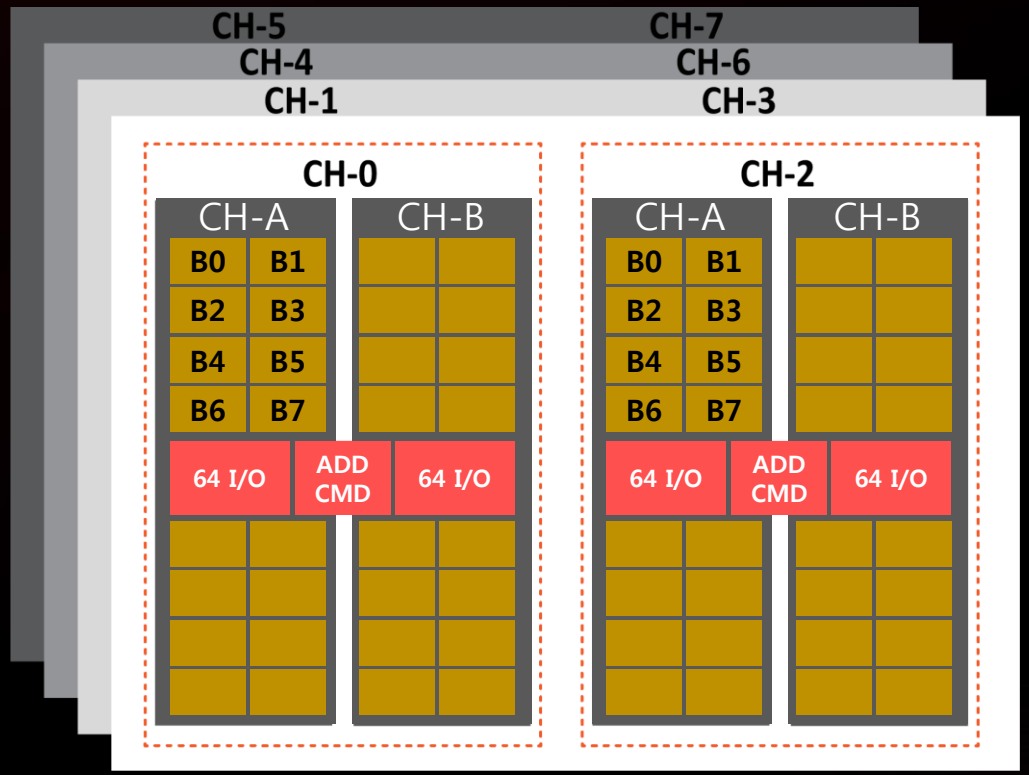


HBM Deep Dive

HBM Architecture

HBM2 core die supports 4 pseudo channels or 2 channels

Each channel consists of 2 Pseudo Channels. Only BL4 is supported

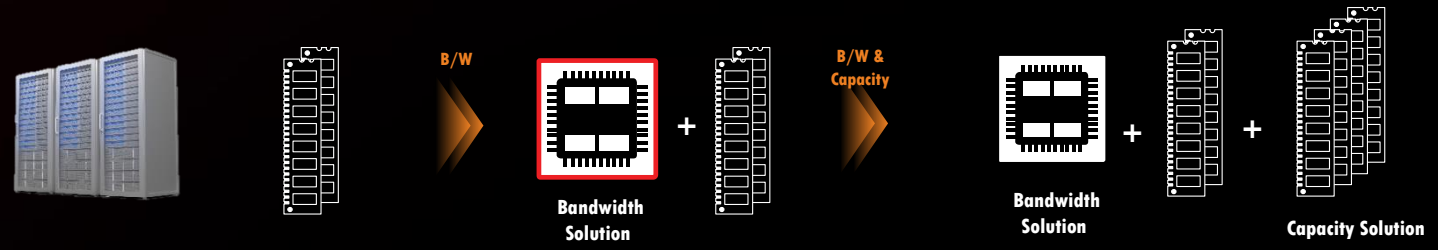


Items	Target
# of Stack	4/8(Core) + 1(Base)
Ch./Slice	2
Total Ch. for KGSD	8/16 (8ch based operation)
IO/Ch.	128
Total IO/KGSD	1024(=128 x 8)
Address/CMD	Dual CMD
Data Rate	DDR

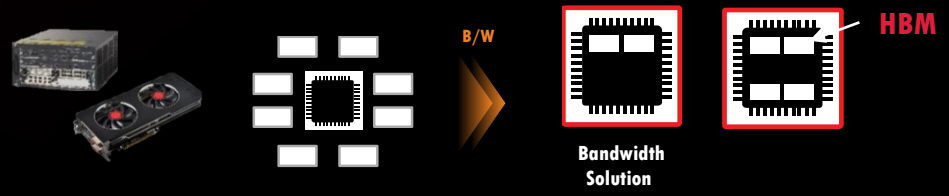
Next-Gen. System Architecture Leveraging HBM

HBM and 2.5D SiP integration unlock new system architecture

HPC & Server
(B/W & Capacity)



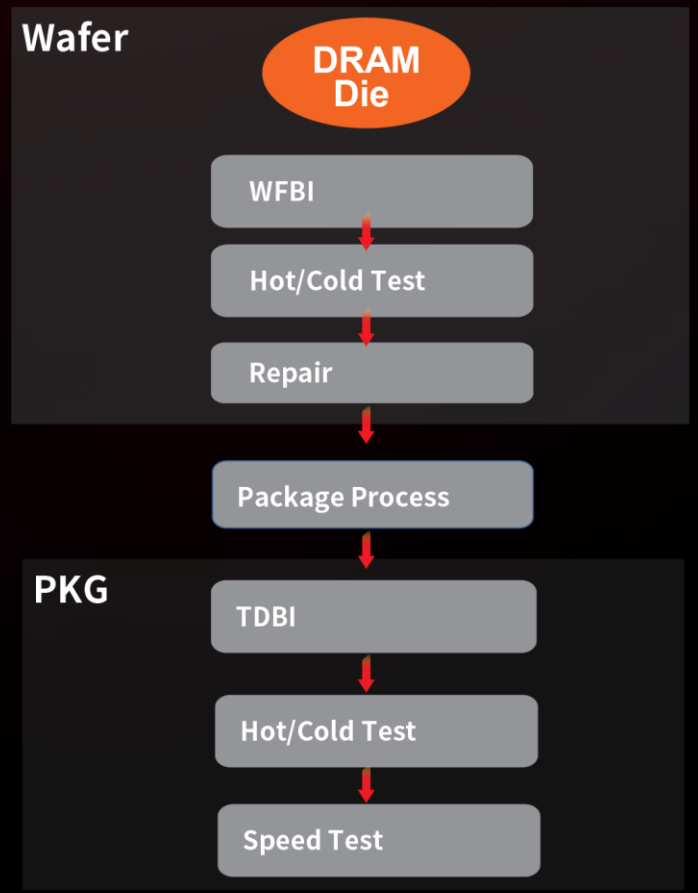
Network & Graphics
(B/W)



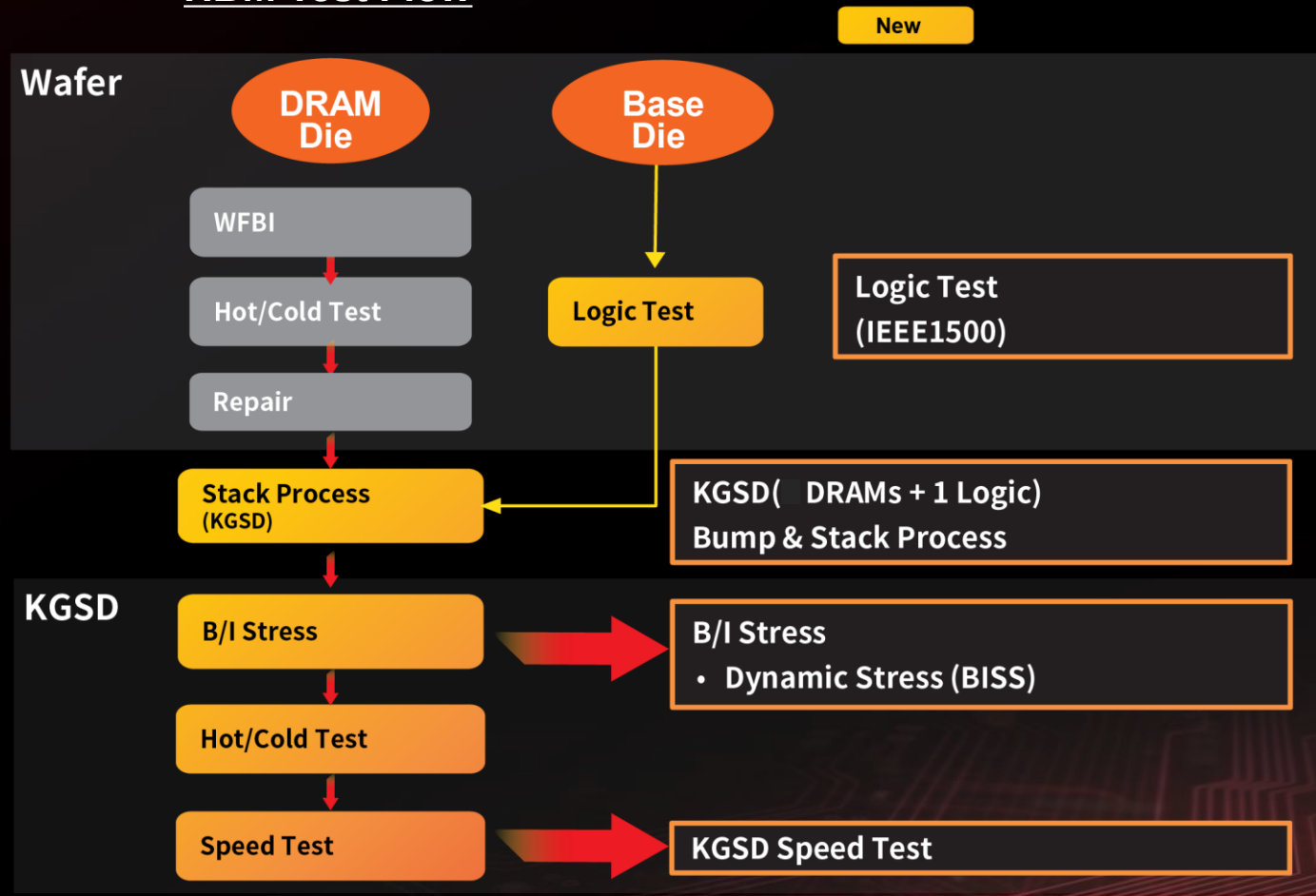
Client-DT & NB
(B/W & Cost)

HBM Test Flow

General DRAM Test Flow

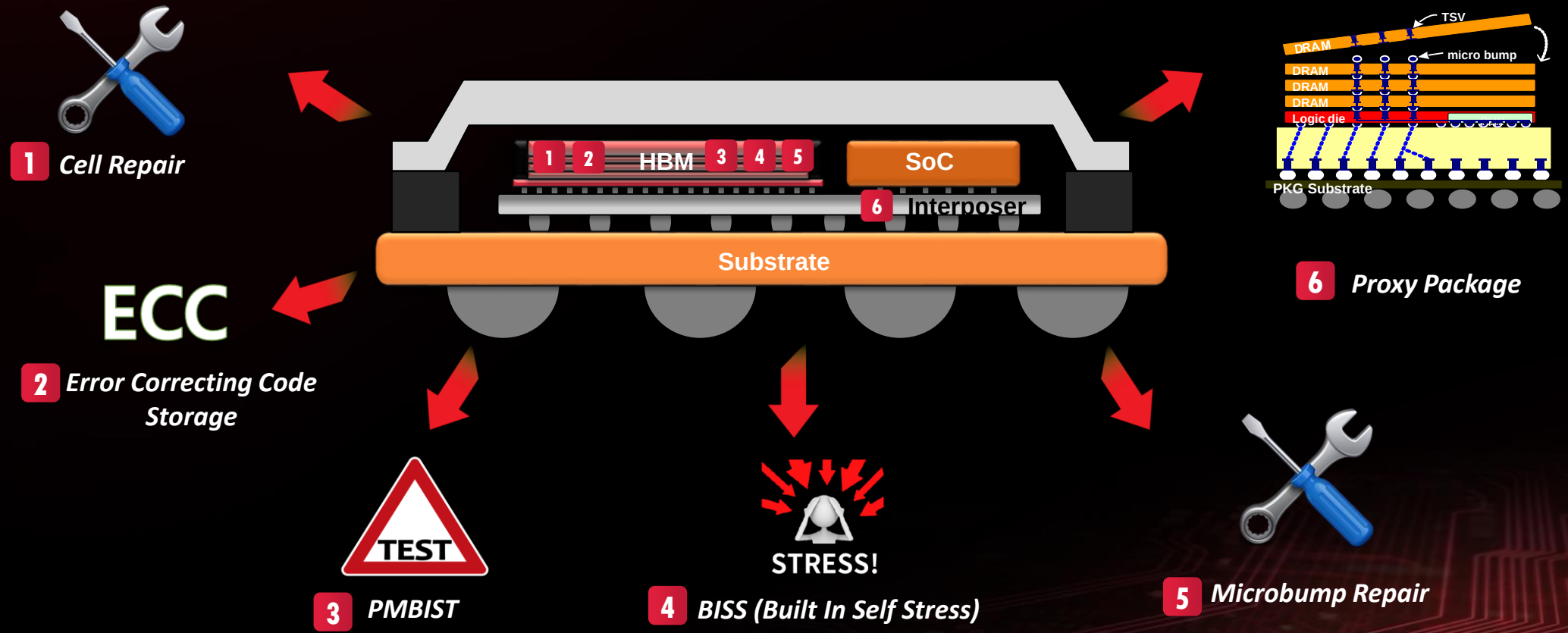


HBM Test Flow



Quality and Reliability Features

HBM Features enable high quality and reliability at post 2.5D assembly

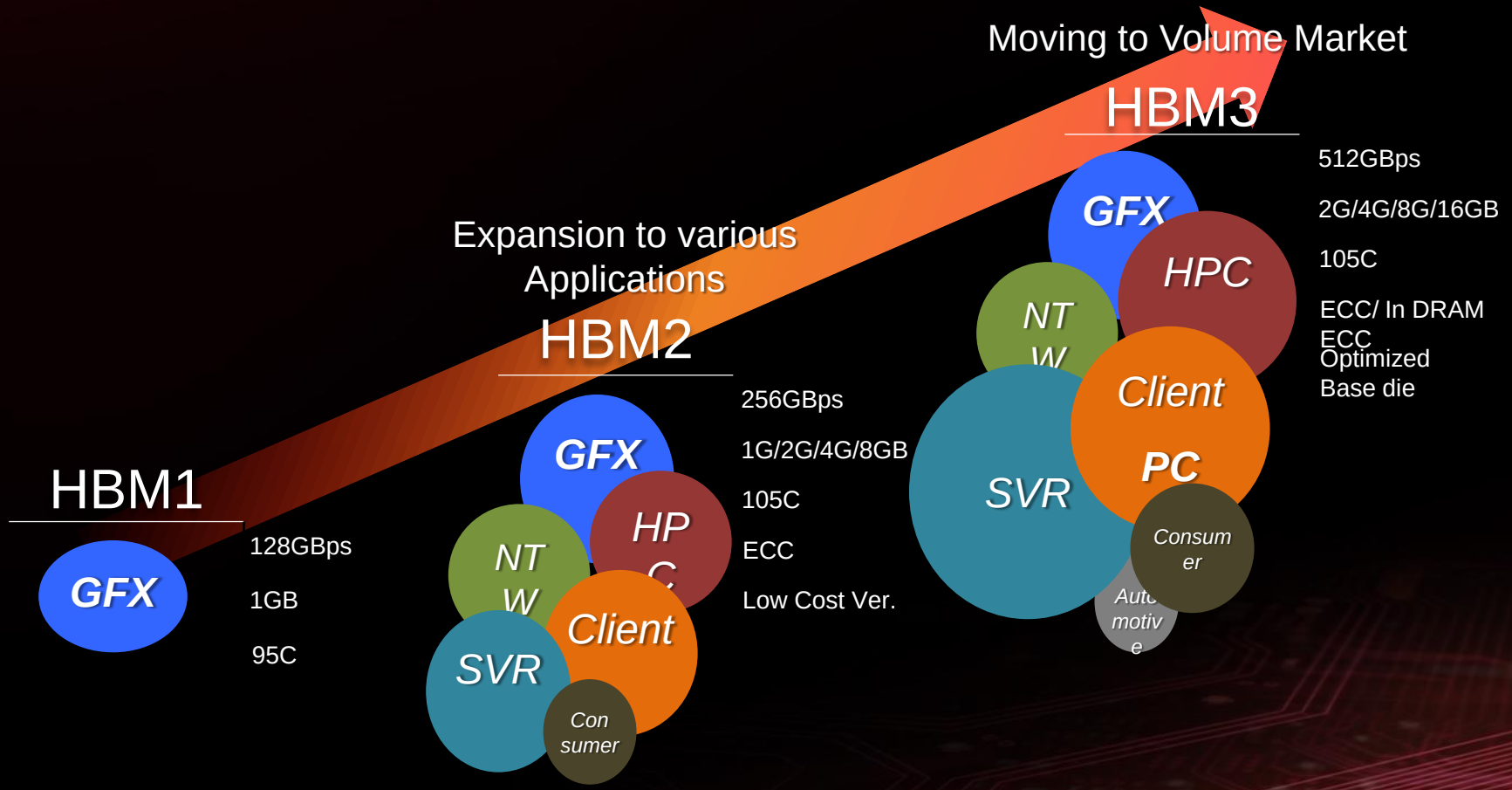


Collaterals Available from HBM vendors

Item	Remarks
Functionality	Datasheet (Jedec/Vendor)
	Verilog (mission mode and DFT)
	IBIS
	Hspice
Mechanical/Interposer design	GDS
	Bump pad netlist
	Bump Ballout
Thermal Simulation	Flotherm
	Icepak

Future of HBM Solution

HBM would penetrate various market segments in the short future.



Q&A


SK hynix
Thank you