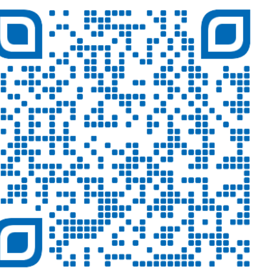
FUDAN UNIVERSITY

先进封装与集成芯片

Advanced Package and Integrated Chips



Lecture 2： Damascene, 2.5D Integration

Instructor: Chixiao Chen, Ph. D

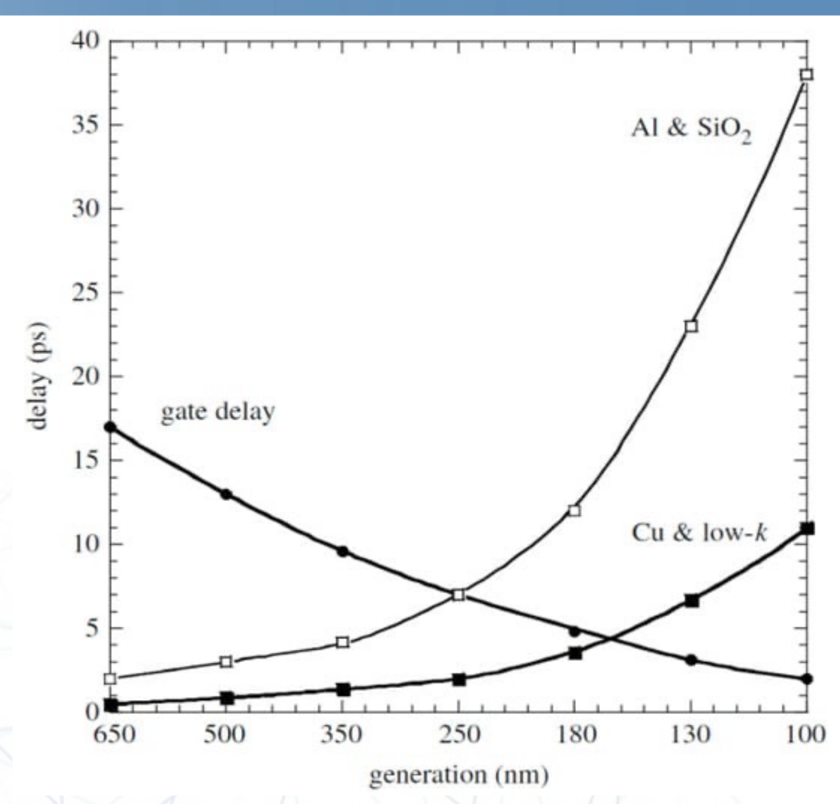
Overview

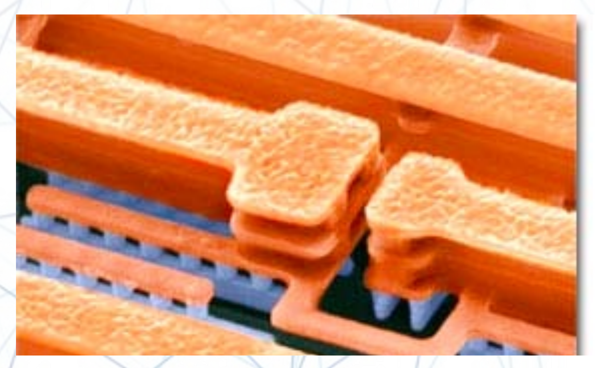
- Damascene process and its application in Advanced Package
- 2.5D (2.xD) Integration technologies
  - Silicon TSV interposer
  - Fan Out Package

Why copper (Cu) interconnect ?

- Advantage of Cu interconnect
  - better conductivity (37%)
  - much less susceptibility to electromigration (10x)
  - enhanced heat dissipation (60%)

- Why do not use Cu in CMOS technology initially?
  - Diffuse Silicon/SiO_2 (i.e. poisonous for devices)
  - Difficult (even impossible) to etch by plasma
  - Quickly oxidizes in air





Key idea of Damascene Process

- IBM introduces Damascene Process, forming Cu interconnect, in 1990
- Why called it Damascene?
  - To make intricately patterned, highly polished steel that was used for swords and knives, metal-smiths in Damascus developed a specific technology.
  - Now, Damascus is the capital of modern Syria.

- Key idea of Damascene Process
  - Replace the copper etching with Chemical Mechanical Planarization
  - Insert a special barrier, typciall Ta, TaN, TiN, TiW, to stop copper diffusion

Single Damascene Process

1. Dielectric (SiO_2/SiN) Deposition
2. Photolithography and Dielectric Etching, leaving trenches/vias
3. Barrier layer (TiN) deposition
4. Copper deposition by electroplating or chemical vapor deposition
   - Two step: thin seed layer + thick layer
5. Chemical-Mechanical Polishing (CMP), remove the excess

Note: TiN is conductor. Damascene is additive.

Dielectric
Etch and Deposit Barrier Layer
Ta/TaN liner
Copper must be encapsulated to prevent out-diffusion
Deposit Seed Layer
Copper Seed Layer
Ta/TaN liner
Wafer Surface
Electro plating
p-Well
n-Well
p-Well
n-Well
CMP
Copper connections between tungsten plugs
p-Well
n-Well
p-Well
n-Well

Fig. 2: Single Damascene Process Steps [1]

Dual Damascene Process

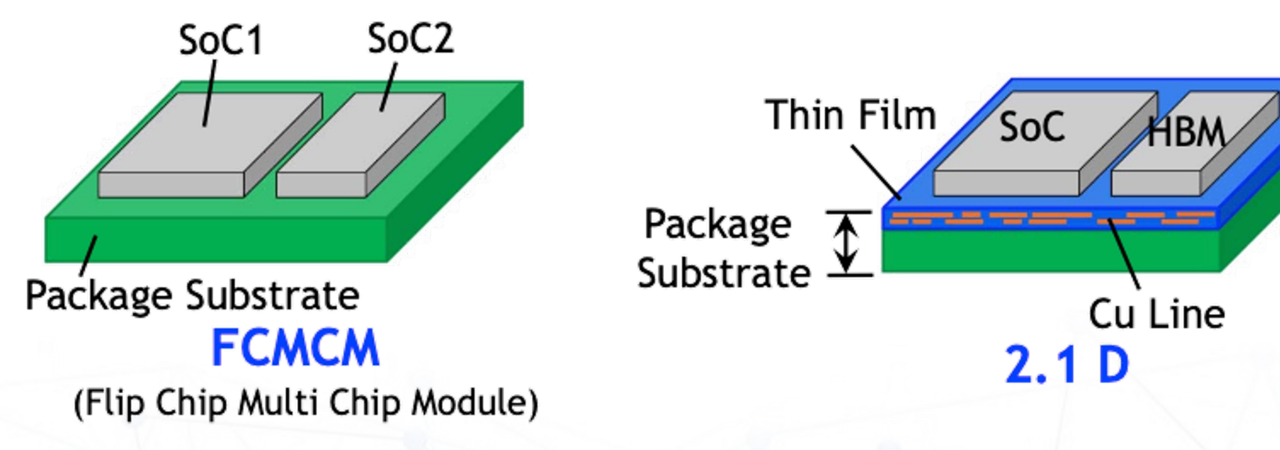
➢ Motivation: making the inter-layer vias without separate via process.

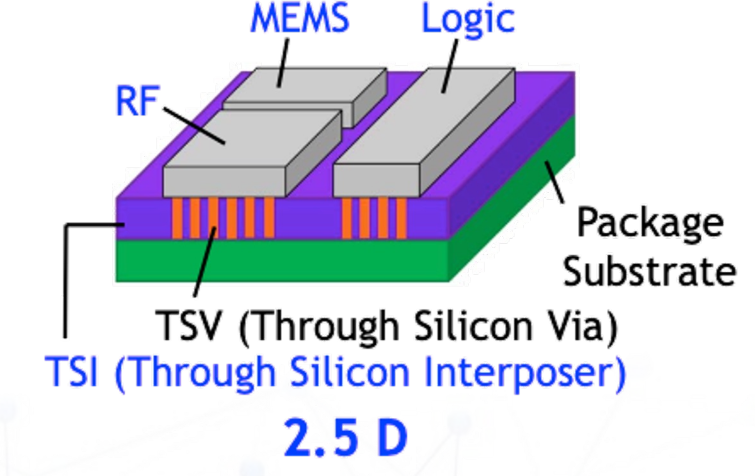
1. Two layer Dielectric Deposition
2. Trench etching (upper layer) using lithography and plasma etching
3. Via etching (lower oxide layer)
4. Barrier layer (TiN) deposition
5. Two step copper deposition
6. Chemical-Mechanical Polishing (CMP), remove the excess copper

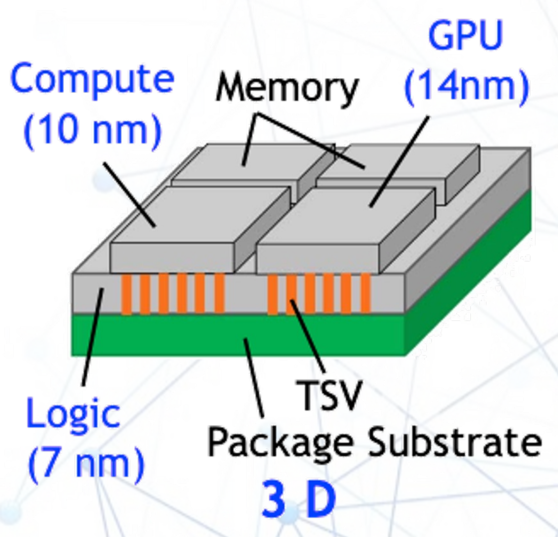
Resist
Dielectric
Etch Stop
Si Wafer
TiN
Copper Seed
Copper Fill

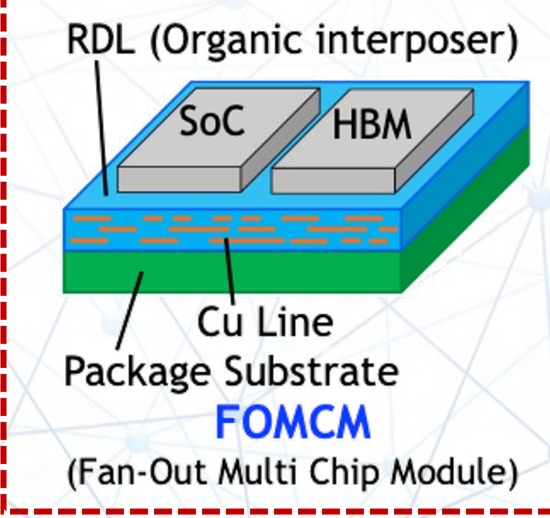
Fig. 4: Trench First then Trench [4]

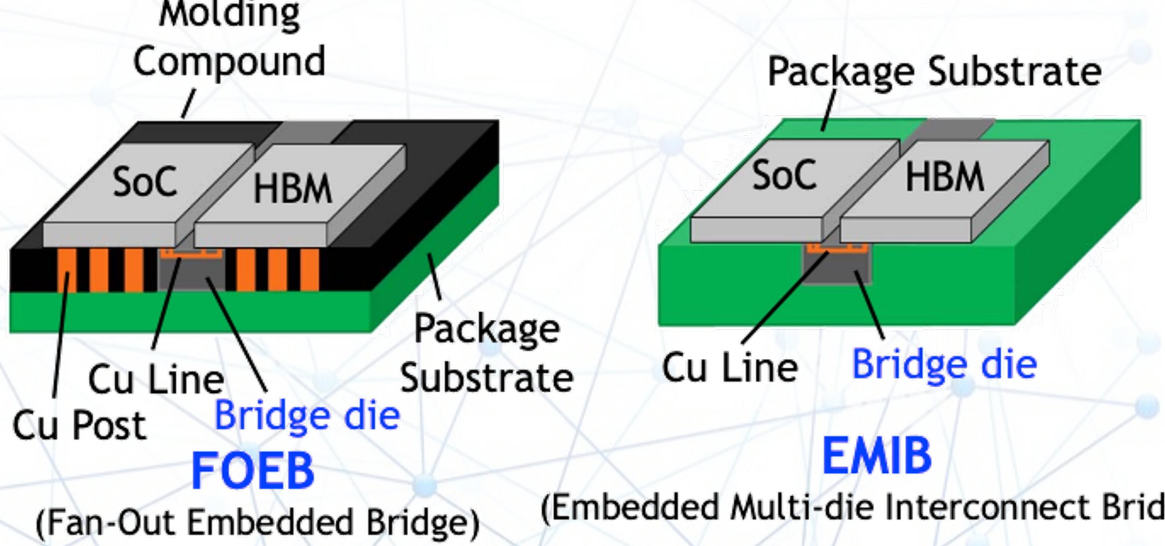
Categories of 2D-2.5D Chiplet Integration











Source: C.K Chung VLSI 2020

A Deep Dive on 2.5D AI Chips

Back Side Metal

DRAM Memory

AI Logic Die

DRAM Memory

Memory Controller Die

Memory Controller Die

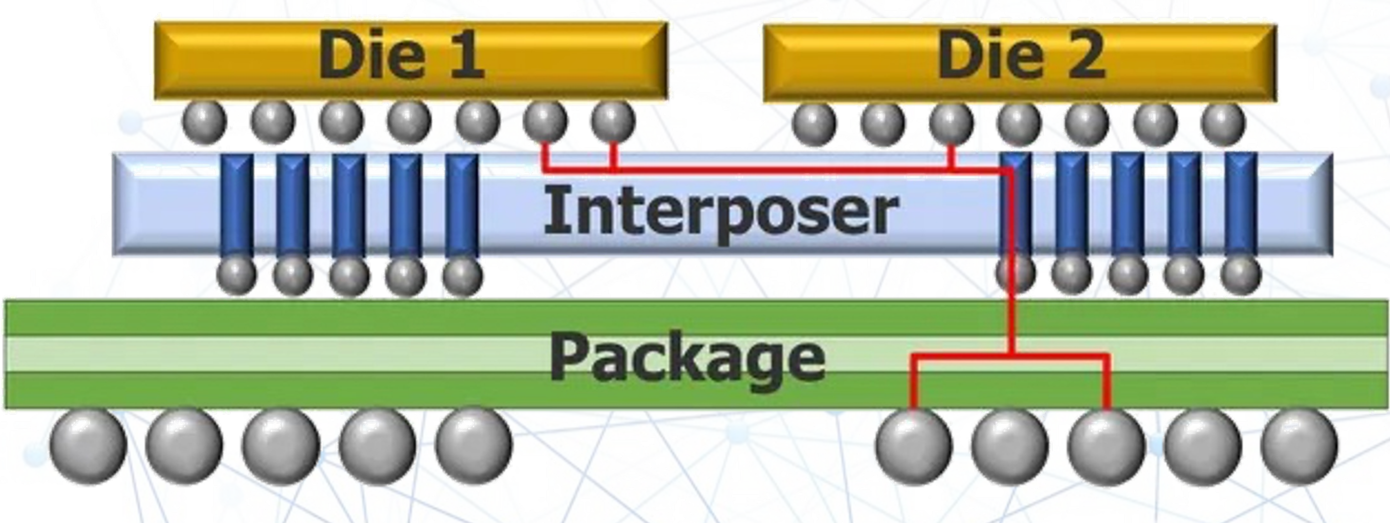
Silicon Interposer

Advanced Substrate

© 2021 Applied Materials, Inc.

Silicon Interposer

- Silicon interposer provides extreme dense interconnect between different dies. Line space and pitch can be set to 400um/RDL layer.
- Normally no active devices in interposer, thus no functions. (if exists, called active interposer)



Architecture	Parallel Interface	Serial Interface
Package	2.5D interposer	Organic substrate
Bump pitch	40 - 55 μm	130 – 150 μm
Interc. density	10^2 – 10^3 IO/mm ^{2}	10^1 IO/mm ^{2}
Line space	>0.4 μm	> 10 μm
Interc. length	<5 mm	<50 mm
Data rate/lane	2 – 8 Gbps	2.5 – 112 Gbps
BW density	2-3 Tbps/mm	1.6-2 Tbps/mm
Power	<0.5 pJ/bit	1.0-1.5 pJ/bit
Latency TX+RX	~4.5 ns	~5.5 ns
Bit error rate	<<1E-15	<1e-15 for NRZ
Standards	HBI, OpenHBI, AIB2.0	OIF, CEI 112G, USR/XSR

The Damascene Process in TSV fabrication

- A similar process is applied for TSV application, which has approximately 10 μm opening in diameter and about 105 μm depth, which give an aspect ratio of 10.5.

Si substrate

SiO_2

SiO_2 by thermal oxidation or PECVD

Photoresist

Photoresist

Mask, Litho/ Patterning

SiO_2 etch

TSV

DRIE

Strip resist

SiO_2

Ti/Cu or Ta/Cu

PECVD or SACVD SiO_2, PVD barrier layer (Ti or Ta) and seed layer (Cu)

Cu

Cu Plating. Then anneal at 400°C for 30m

SiO_2

Cu TSV

Ti/Cu or Ta/Cu

CMP the overburden Cu, Cu seed layer, and barrier layer

100μm

TSV

0.40μm

0.44μm

10.1μm

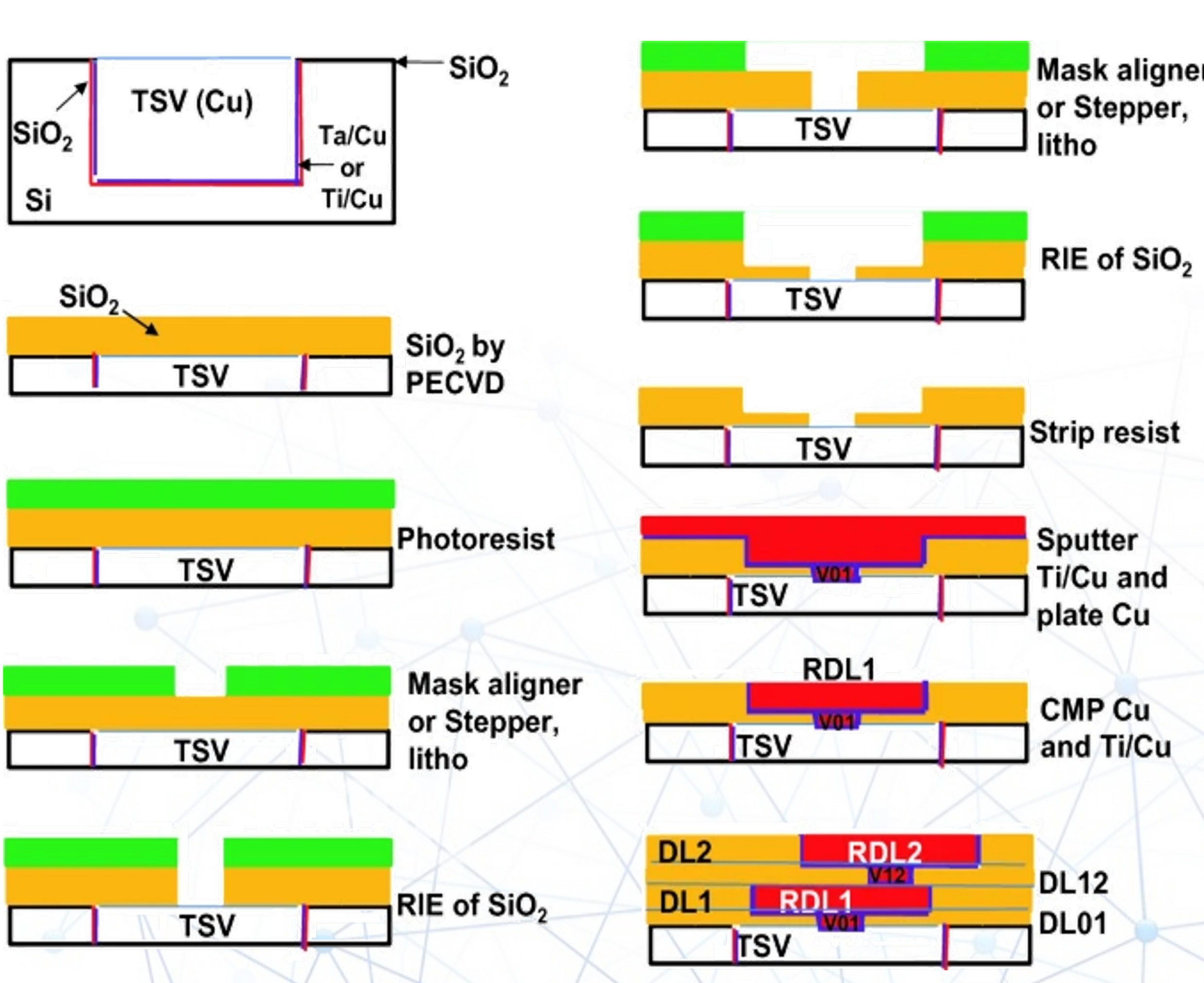
0.83μm

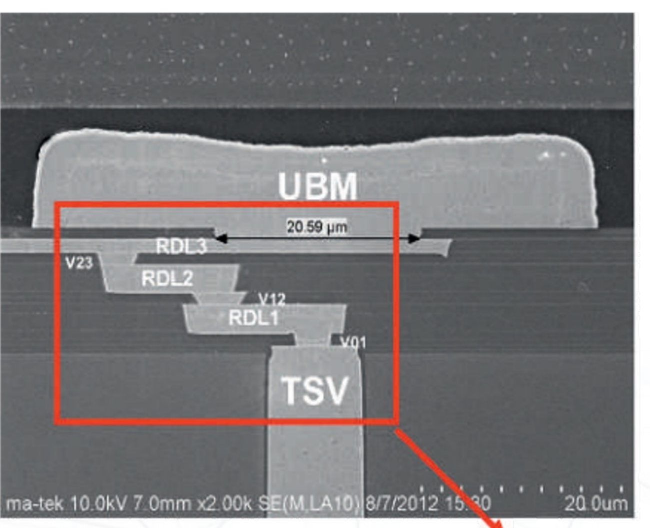
8.67μm

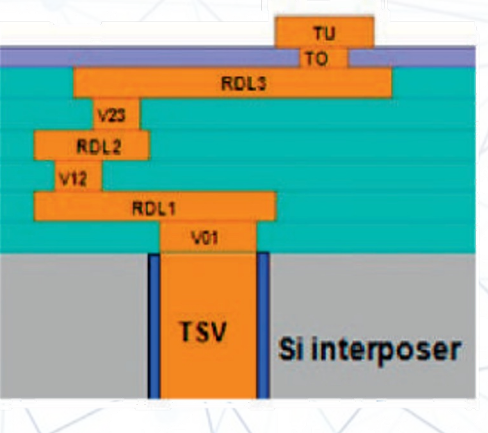
Deep Reactive Ion Etching

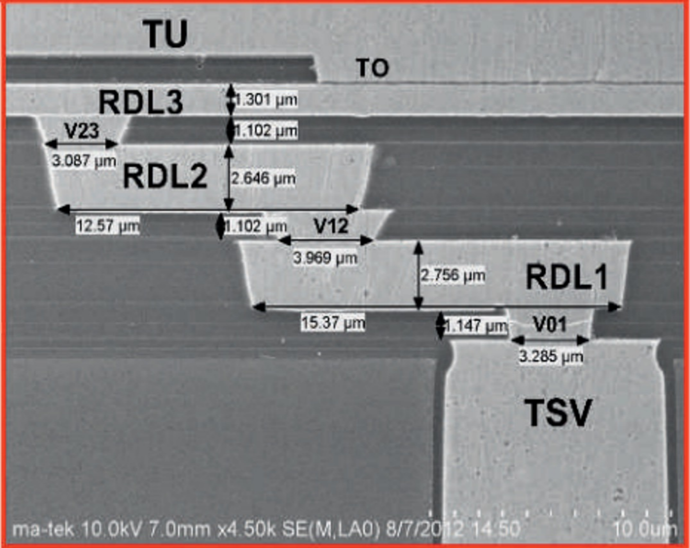
- Traditional wet etching methods tend to produce tapered vias, while DRIE is capable of etching these deep, narrow holes while maintaining a high aspect ratio.
- Bosch process, which alternates between etching and passivation (on the sidewalls of the via to protect them during the etching process) cycles,

Connecting Metal Routing to TSVs









Under Bump Metallization

- The UBM serves as an interface layer between the chip's bond pad and the solder or copper pillar, ensuring a reliable mechanical and electrical connection.
- Flow: PI patterning, UBM sputtering, PR patterning, bump plating, PR stripping, UBM etching and flow.
- Nickel/Ti to serve as barrier to prevent the diffusion of copper or aluminum into each other.

Photoresist
Underbump metallization
Solder
Ni barrier
Cu pillar
a)
b)
c)
d)
e)

Packaging Flow for Silicon Interposer

- Die-last Chip-on-Wafer (CoW) flow
- Interposer can be tested before CoW process
- Thin wafer to CoW
- Need 2x glass carrier
- Question: Which company complete the Front/Middle/Back end process?
  - OSAT: Outsource Semiconductor Assembly and Test
  - Semiconductor fab

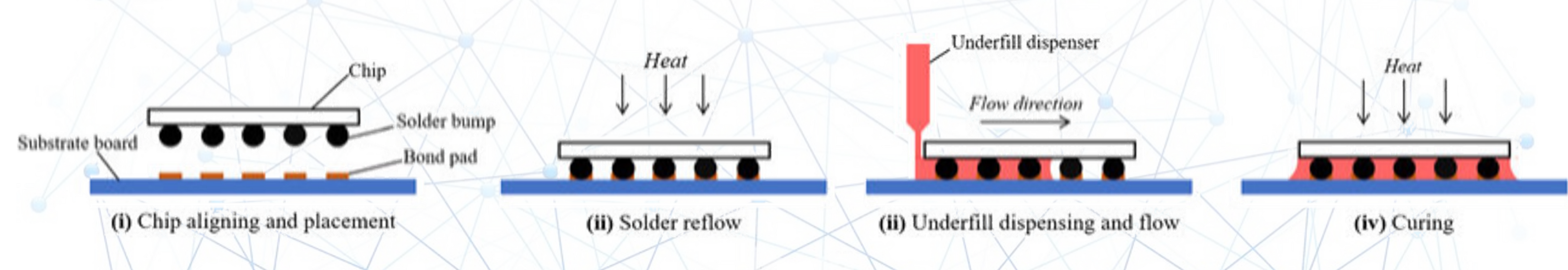
Passivation
Si Substrate
TSV
UBM
RDL/Pass.
Carrier #1
Interposer
Temporary Bonding
Carrier #1
Backgrinding
Carrier #1
Si Etching, Pass., and Cu Revealing
Carrier #1
UBM
Carrier #1
C4 Wafer Bumping
Carrier #1
Carrier #2
Temporary Bonding Carrier #2
Carrier #2
De-bonding Carrier #1
Carrier #2
Chip-on-Wafer. Underfill
De-bonding Carrier #2 and dicing
Chip
TSV
Interposer
Package Substrate
Assemble the TSV module on package substrate, then test. Underfill

Not-to-scale

Underfill

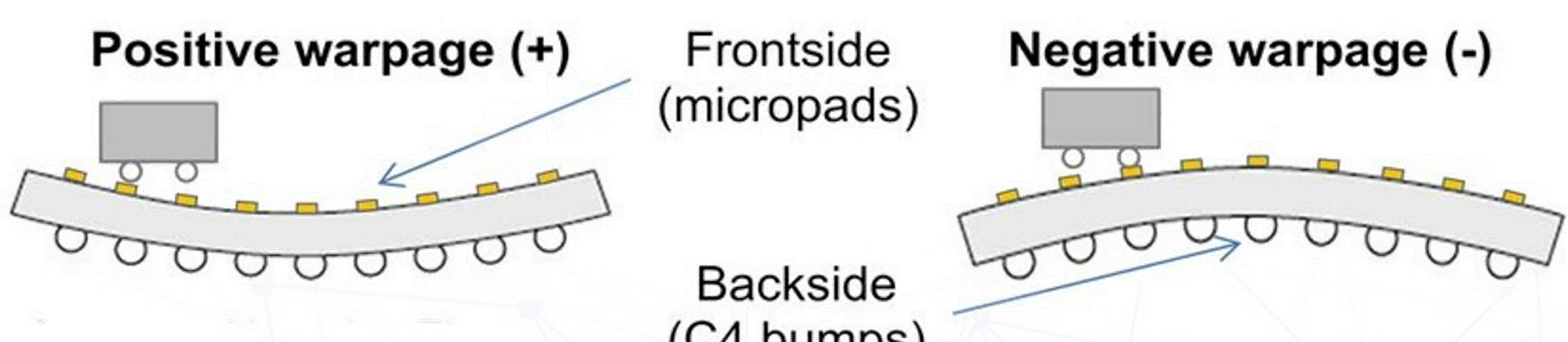
➢ Underfill refers to an epoxy-based (organic) material applied between a die and interposer, providing additional mechanical support and distributing thermal stresses evenly and enhancing electrical reliability due to solder joint fatigue.

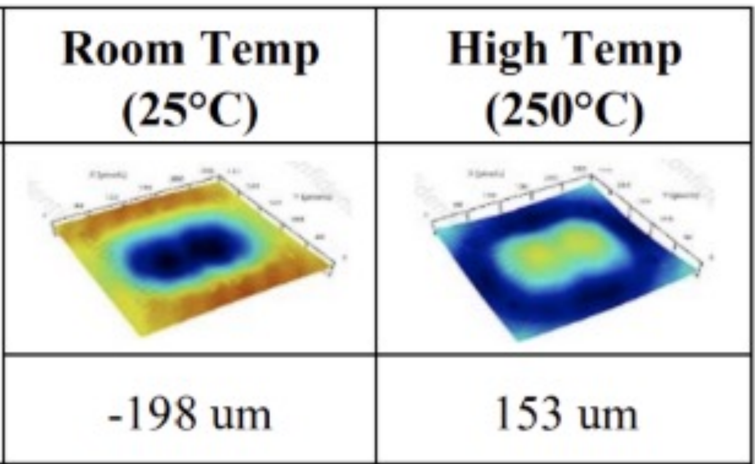
➢ Capillary Underfill: underfill material is applied at one edge of the chip, and through capillary action, the material flows into the gaps between the chip and the substrate. Underfill flow: CoW alignment, solder reflow, flux cleaning, dispensing, curing.



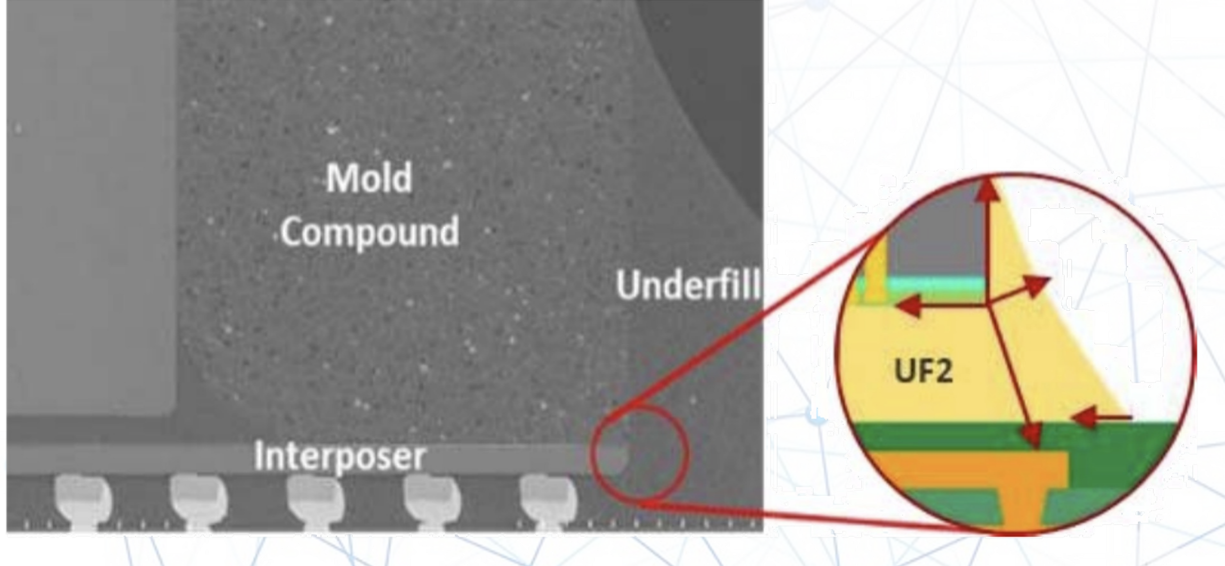
Stress Analysis for Interposer Warpage

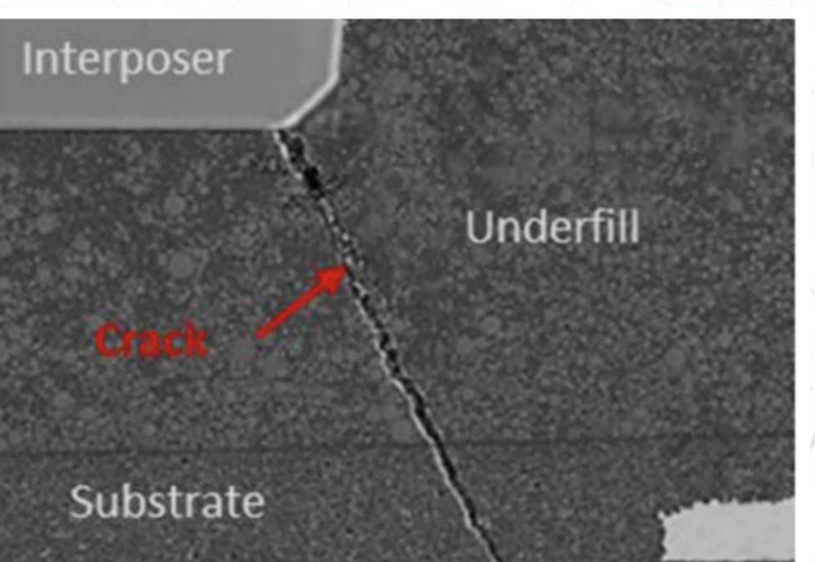
➢ The interposers are thinning to 100um resulting warpage.





Room Temp (25°C)	High Temp (250°C)
-198 um	153 um

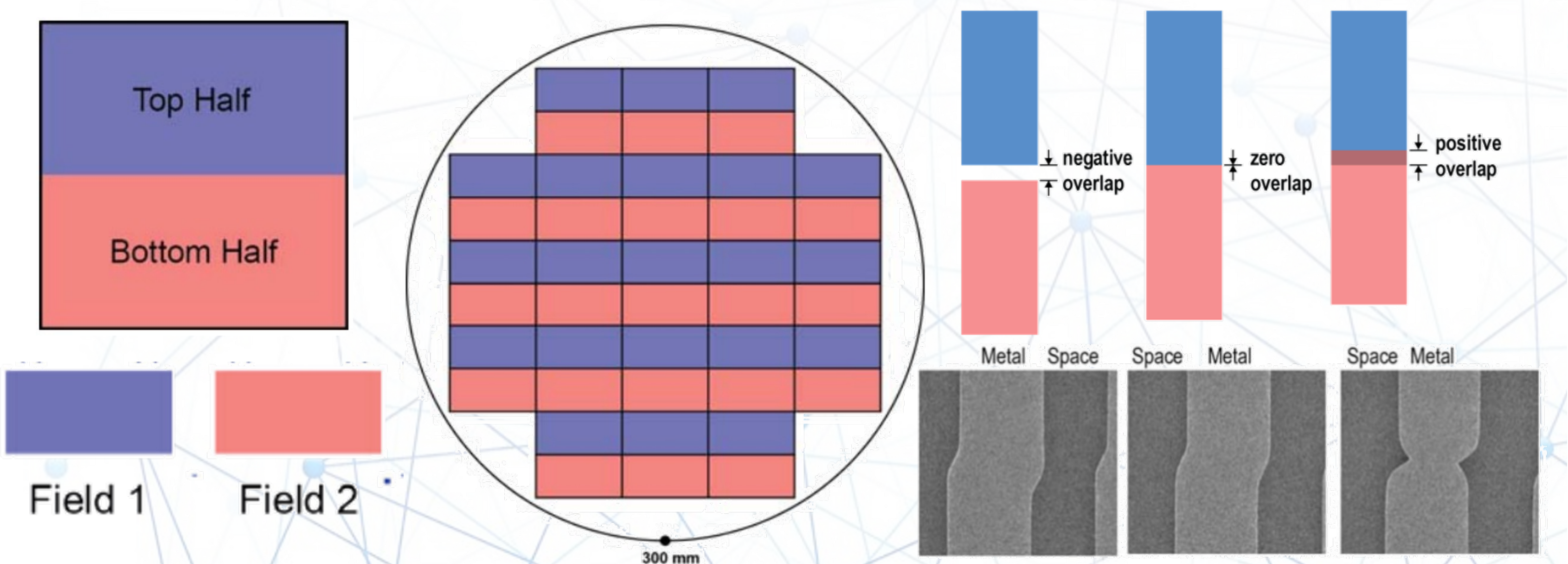


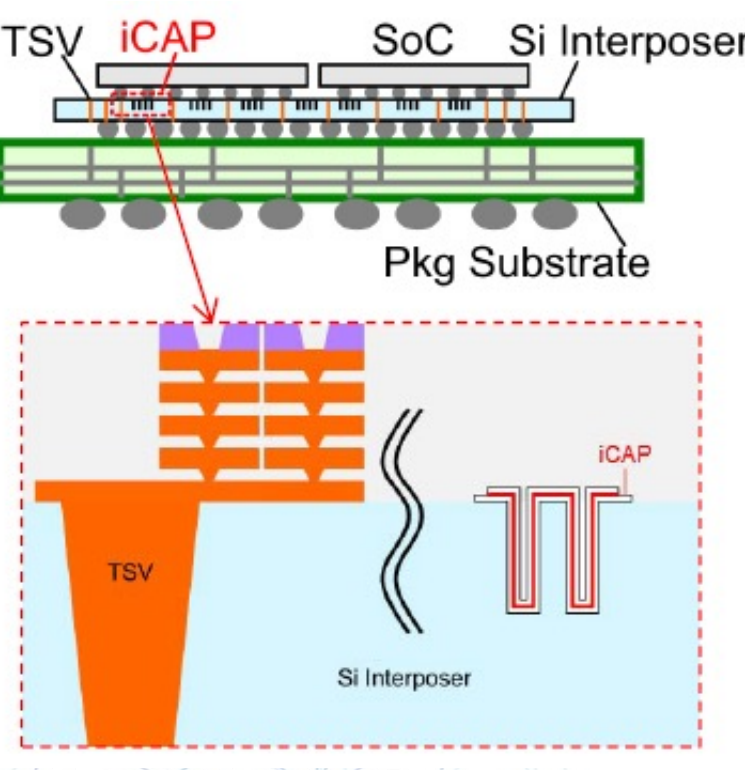


https://ieeexplore.ieee.org/document/10195766

Advanced Silicon Interposer Technology

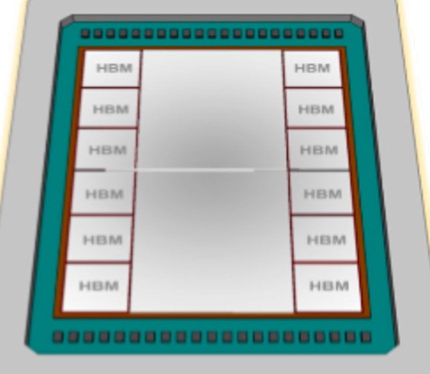
- To fabricate an interposer whose area is larger than reticle size, splitting and stitching is needed. Lithography stepper has less resolution at stitching boundaries.
- To enhance the power integrity, deep trench capacity (DTC) is required, like DRAM. DTC is embedded in silicon interposer with high-k dielectronic.

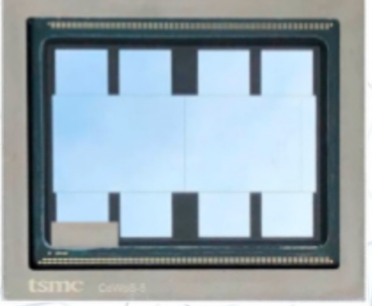




Roadmap of CoWoS-S Technology

- TSMC has developed 5+ generation of Silicon interposer (CoWoS-S) technology
- The main progress is the area of the interposer
- Challenges include warpage and cost.





Epoxy Molding Compound

- Conventional packages use an epoxy mold compound to fully embed the dies, rather than placing them upon a substrate or interposer.
- Epoxy resin reacts with a curing material under 180-220 C temperature, forming solid layer excellent mechanical strength and thermal stability.

RDL on Epoxy Molding Compound

- After a chip is encapsulated with epoxy molding compound (EMC), a dielectric layer, often a photosensitive polyimide (PI), is applied over the exposed I/O pads. The metal layer, typically copper (Cu), is then deposited and patterned to form the desired interconnects.

Al Pad
Passivation
KGD
EMC
Photosensitive polyimide (PI)
Spin Polymer
Mask aligner or Stepper and Etch Polymer
TiCu
Sputter TiCu
Photoresist
Mask aligner or Stepper (Litho)
Cu Plating
RDL1
Polymer
Strip Resist & Etch TiCu
TiCu
RDL2
RDL1
KGD
EMC
UBM-less Cu-pad
Solder ball
Solder ball
UBM
Contact pad
RDL2
Dielectric2
RDL1
Dielectric1
KGD
EMC
Al Pad
Passivation
TiCu

Fan-Out Wafer Level Packaging Technology

- Fan-In vs. Fan-Out, an cost effective way.
- EMC has a higher coefficient of thermal expansion (CTE) (10-20 ppm/°C) compared to silicon (2.6 ppm/°C). This mismatch causes die shift and warpage necessitating larger line widths and spacing.

Fan-In WLCSP
Fan-in wafer-level chip scale package

Fan-Out WLCSP
Fan-out wafer-level chip scale package

Chip
Chip
Epoxy mold compound (EMC)
Chip
Package size

cess time-temperature-profile

Temperature (°C)
Time (min)
1. Assembly
2. Molding
3. Post Mold Cure
4. Debonding
5. Cooling down

heating up to molding temperature → EMC cure → debond and cooling down
carrier CTE
EMC chemical shrinkage
EMC CTE
EMC
die
carrier

Fan-Out Wafer Level Packaging Technology

- Chip-first with die face down: ① place on a carrier ② Epoxy Molding ③ RDL
  Chip-first with die face up: ① Cu stud and Die attach ② Molding ③ grindling ④ RDL

Test for KGD and Dice
Device Wafer
Passivation
Al or Cu Pad
CHIP
Chip-first (Die face-down)
2-side (thermal release) tape
Temporary wafer/panel carrier
(a)
KGD
KGD
KGD
KGD
Reconstituted carrier
(b)
Compression molding of EMC on reconstituted carrier
EMC
(c)
Remove carrier and tape
(d)
Build RDLs and mount solder balls
(e)
RDLs
Solder balls

Fan-Out Wafer Level Packaging Technology

- Chip-first with die face down: ① place on a carrier ② Epoxy Molding ③ RDL

  Chip-first with die face up: ① Cu stud and Die attach ② Molding ③ grindling ④ RDL

Test for KGD
Device Wafer
UBM
Cu contact-pad
(a)
CHIP
Passivation
Al or Cu Pad
Sputter UBM and electroplate contact pad
UBM
Polymer
Contact pad
(b)
CHIP
DAF
Polymer on top, die-attach film on bottom of wafer, and dice the wafer
Coated with a LTHC sacrificial layer
LTHC
Temporary glass wafer carrier
(c)
Die face-up
KGD
KGD
KGD
Reconstituted carrier
(d)
Compression molding the reconstituted carrier and post mold cure
EMC
(e)
Backgrind the EMC to expose the Cu contact-pad
(f)
Solder balls
RDLs
Build RDLs on contact pads and mount solder balls
(g)

Organic Interposer Using Chip Last

- Issues of Chip first technology: die drift due to molding process. Limited routing metal L/S (10μm for Face down, 5μm for face up)
- Chip-last / RDL-first is developed to reduce the minimum L/S to 2μm.

(a) Sacrificial layer, Glass wafer — Coating a release (sacrificial) layer on a glass wafer

(b) Cu-pad, DL1, Opening — Cu-contact pads and UBM fabrication and first dielectric layer (DL1) opening

(c) RDL1 — Fabrication of the first RDL1

(d) DL2, Opening — Second dielectric layer (DL2) opening

(e) RDL2 — Fabrication of the second RDL2

(f) DL3, μbump pads — Third dielectric layer (DL3) opening and μbump pads formation

(a) Chip, Cu-pillar with solder cap, Chip, Pad, Glass wafer, Sacrificial layer, RDLs — Chip-to-wafer with RDLs bonding

(b) Underfill — Underfill dispensing and curing

(c) EMC (epoxy molding compound) — Compression molding of EMC

(d) Debonding the glass wafer by a laser

Packaging Flow Comparison for Fan Out

Method	Chip First		Chip last
	Face up	Face down	Face down
Process flow	(a) Chip DB (b) Reconstitution (c) RDL (d) Backend (C4 + sawing)	(a) Reconstitution (b) RDL (c) Backend (C4 + sawing)	(a1) Chip (a2) RDL (b) Flip Chip assembly (c) Backend (C4 + sawing)
Advantages	• Low die shift • Good RDL routability	• Low cost • Good RDL routability	• RDL No Die lost • Better RDL routability • Lower thermal budget • Higher reticle sizes • Faster cycle time • Better reliability
Dis-advantages	• Die lost concern • Higher thermal budget • higher cost (2 glass carriers)	• Die lost concern • Higher thermal budget • Die shift • Reticle size Limitation	• Higher cost (2 glass carriers) • Additional Underfill processes

Silicon Interposer Vs. RDL Interposer.

- ➢ Silicon interposer adopts BEOL CMOS technology, achieving 400nm line width/spacing, 200nm via diameter.
- ➢ RDL interposers's width and pitch is 5-10x higher, but cost effective and large area.

Type	Silicon	RDL
Dielect Layer	SiO_2	Organic Epoxy/PI
Line width / Space	400nm	2-5μm
Via Diameter	200nm	2μm
Cost	**High**	**Low**

Overall Area

Soufce: TSMC

Silicon Interposer

RDL/FanOut interposer

1.5x reticle

2x reticle

3.3x reticle

1.4x reticle

3.3x reticle

5.5x reticle

8x reticle

1 reticle ~ 830mm²

2016 2020 2023 2024 2025 2026 2027

Fan-Out for Ultra High Performance Computing

- Full wafer integration (System-on-wafer) is an emerging technology for ultra-high performance computing and high bandwidth die-to-die communication.
- Leveraging Fan-Out technology is a SoW solution with Known-Good-Dies. It also allows heterogeneous integration of compute/IO/… chiplets and integrated power.

Tesla Dojo supercomputer is the first industry full-wafer heterogeneous integration technolgy with good process control and high quality RDL.

Mimic power module
Connector
InFO wafer
Mimic thermal module (Cold Plate)

Power Module
Connector
Chip 1
Chip 2
Chip 3
Thermal Module