Chiplet-Based Modular Computing Platform

COWOS/ EMIB / MCM Research Analysis

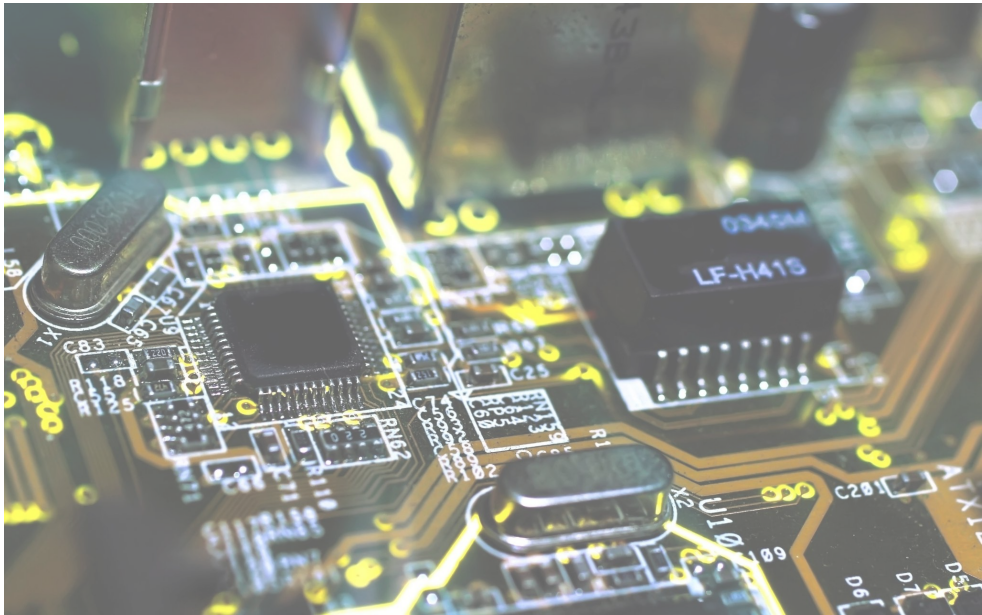
Siyuan Jing, Xindong Zhou, Pengtao Huang

Project Definition

- Our product enables seamless hardware customization through chiplet technology, providing cost-efficient and adaptable solution to meet diverse needs.
- For companies or individuals who are seeking customizable, high-performance computing solutions.

Three Directions

- Standardized Interfaces
- Encapsulation Technology
- Modular Development



Target Users and User Stories

Potential Companies and Businesses:

- Telecommunications Companies: Qualcomm, Ericsson

Integrate chiplet technology into networking hardware to enhance performance and scalability for 5G and IoT applications.

- Gaming Companies：Epic Games, Unity Technologies

Utilize the platform to create powerful gaming engines that can dynamically allocate resources based on user demands.

- Cloud Service Providers：Amazon Web Services

Use the modular platform to offer customized computing solutions to clients, enhancing their cloud services.

Develop Analysis Platform for Chiplet

An LLM (Large Language Model) based analysis platform allows customers to select the right chip technology and packaging solution to improve computing speed and productivity.

Input:

Product type: consumer-grade chip, industrial-grade chip or server-grade chip, etc

Design goals: high performance computing, package integration, etc

Production yield:

expected mass production scale, such as mass production or small customization

Cost constraints: customer's budget or cost target

Develop Analysis Platform for Chiplet

Analysis process:

Based on the parameters entered by the user, LLM is able to generate analysis reports from technical documents, market trends, manufacturers' production capacity, packaging technology characteristics and other information.

The output of LLM details the reasons for the recommendation, such as how the recommended technology solution meets the design goals.

Develop Analysis Platform for Chiplet

Outputs:

1. Selection recommendations:

Recommend the chip packaging technology, process node, and manufacturer that best suits the user's needs.

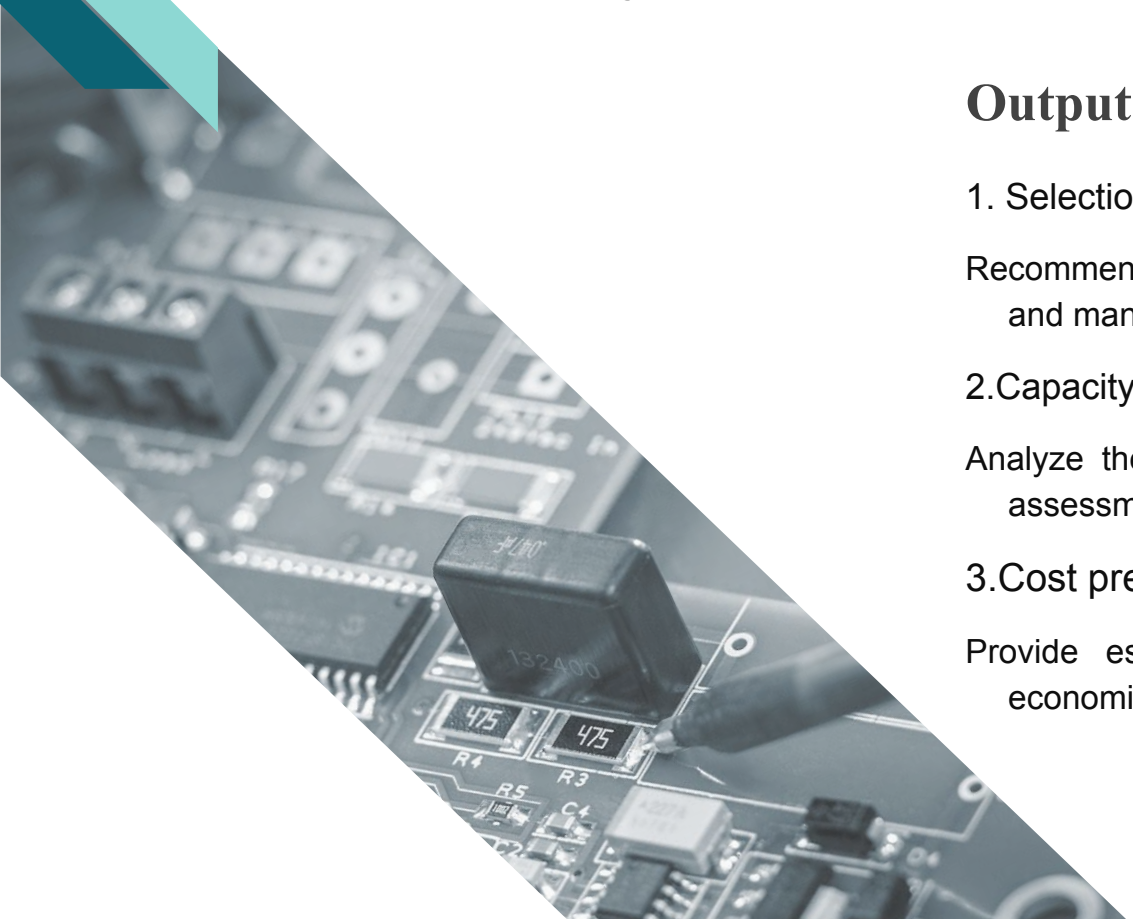
2.Capacity Recommendation:

Analyze the current vendor's capacity planning and give an assessment of the feasibility of mass production.

3.Cost prediction:

Provide estimated production costs to help users make economic decisions.

Develop Analysis Platform for Chiplet



Outputs:

1. Selection recommendations:

Recommend the chip packaging technology, process node, and manufacturer that best suits the user's needs.

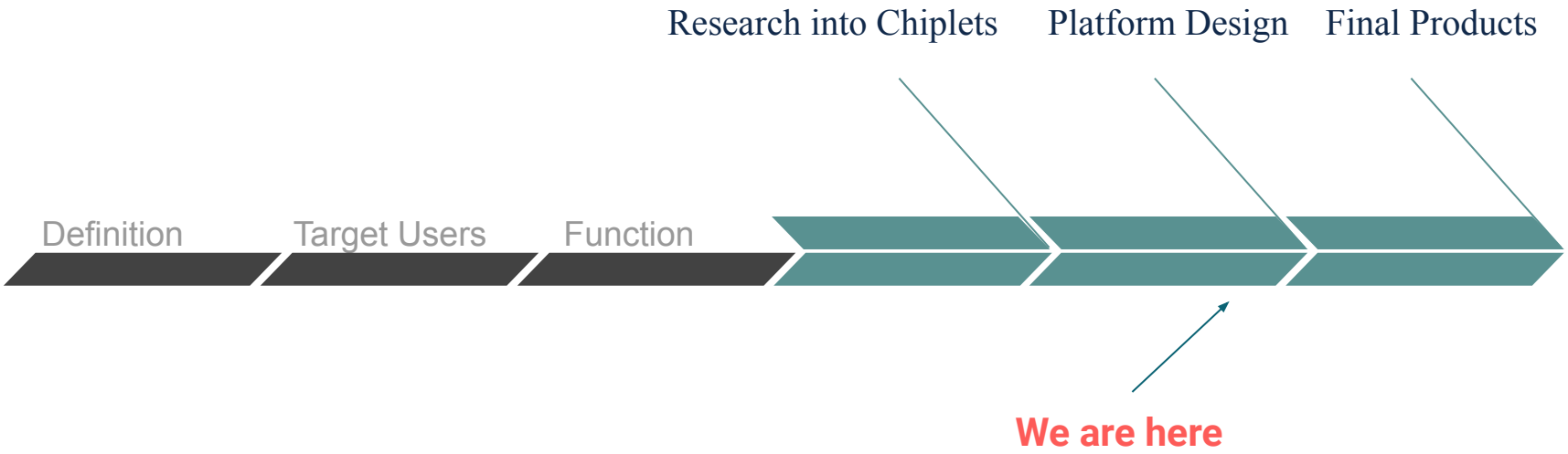
2.Capacity Recommendation:

Analyze the current vendor's capacity planning and give an assessment of the feasibility of mass production.

3.Cost prediction:

Provide estimated production costs to help users make economic decisions.

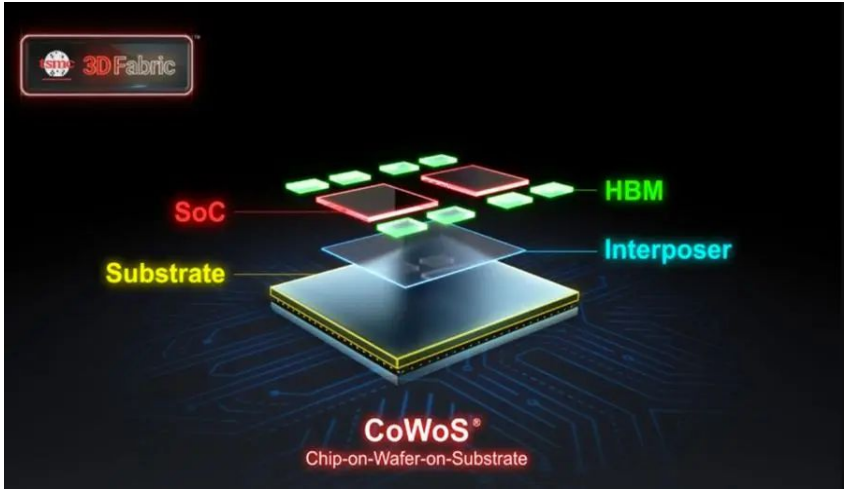
Project Timeline



Product Performance Characteristics

Cowos

(Chip-on-Wafer-on-Substrate)



Manufacturers and Reasons

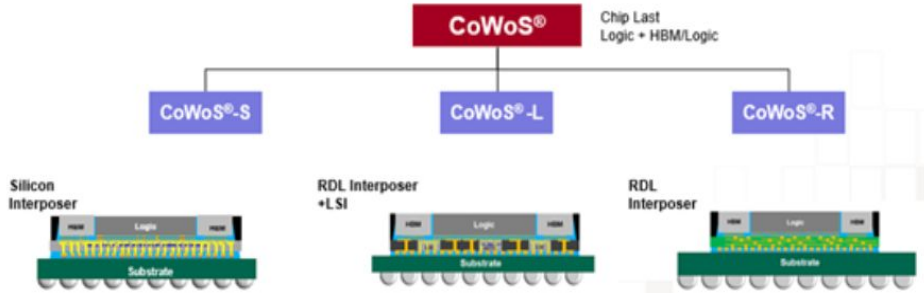
Due to the advanced packaging requirements of AI chips, TSMC's 2.5D advanced packaging CoWoS technology is currently the primary technology used for AI chips.

Clients

This includes NVIDIA's H100 GPU, which utilizes TSMC's 4-nanometer advanced process, as well as the A100 GPU, which uses TSMC's 7-nanometer process, both of which are packaged using CoWoS technology.

Technological Route

The CoWoS® family covers three technologies – CoWoS®-S, CoWoS®-L, CoWoS®-R



Resource:

Y. -C. Hu *et al.*, "CoWoS Architecture Evolution for Next Generation HPC on 2.5D System in Package," *2023 IEEE 73rd Electronic Components and Technology Conference (ECTC)*, Orlando, FL, USA, 2023, pp. 1022-1026, doi: 10.1109/ECTC51909.2023.00174.

Cowos

(Chip-on-Wafer-on-Substrate)

Comparison Between CoWoS-L and CoWoS-S

CoWoS-L can provide higher capacitance than CoWoS-S. Due to yield concern, the maximum area of eDTC has an upper limit around 300 mm2 on a single silicon chip.

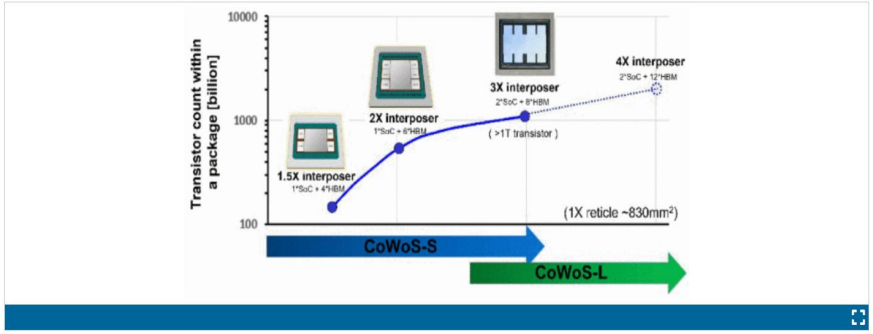
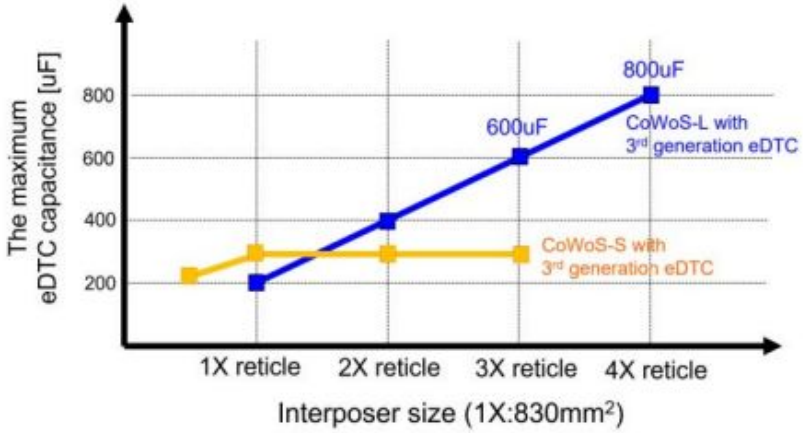


Fig. 1.
Cowos development progress



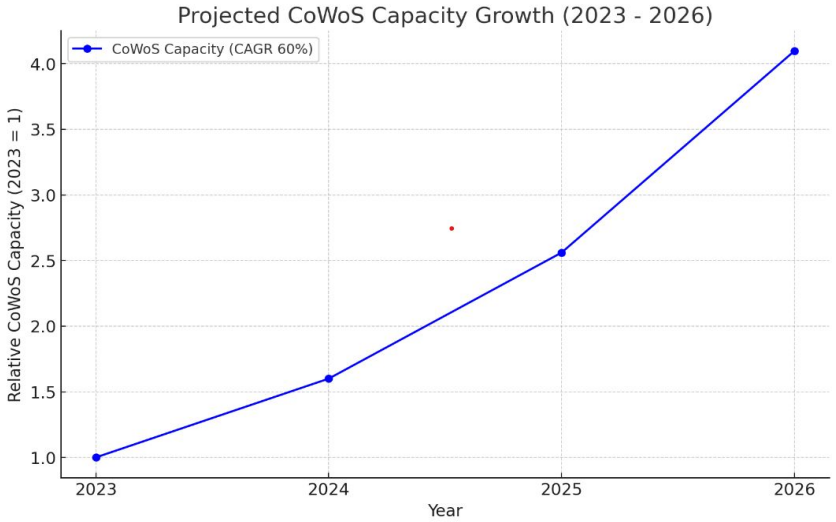
Maximum eDTC capacitance between CoWoS-S and CoWoS-L.

CoWoS-L with multiple LSI chips can dramatically increase the total eDTC capacitance on RI by connecting the capacitance of all LSI chips.

Cowos

(Chip-on-Wafer-on-Substrate)

TSMC More Ambitious on CoWoS Capacity Expansion, Targeting 60% CAGR by 2026



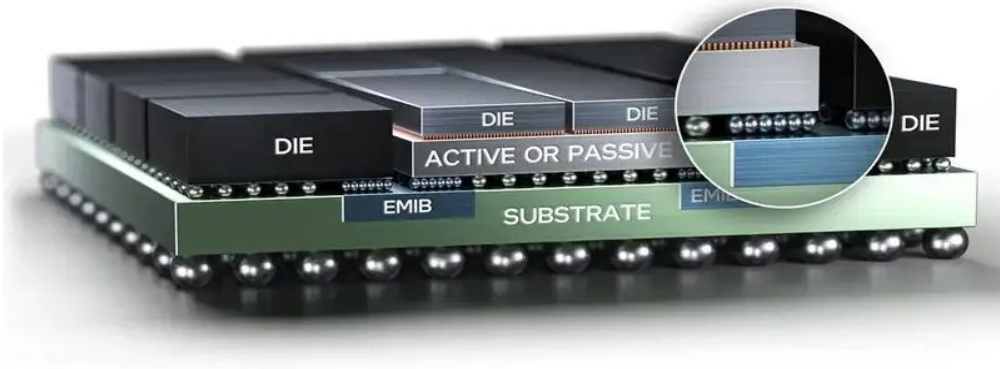
Capacity Expectations:

TSMC also plans to expand its system-on-integrated chips (SoIC) capacity at a CAGR of 100% through 2026, indicating that its SoIC capacity will increase eight-fold from 2023 levels by the end of 2026, according to AnandTech.

The CoWoS platform allows integration with High Bandwidth Memory (HBM) and advanced GPU cores, which are essential for AI and HPC applications. This high-speed interface capability has positioned CoWoS as a preferred solution for AI chip companies that need rapid processing and

EMIB

(Embedded Multi-die Interconnect Bridge)

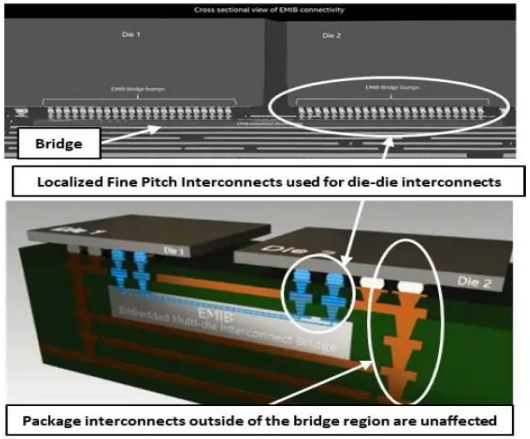


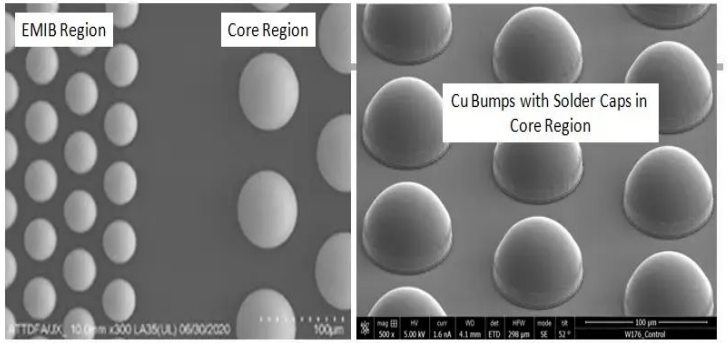
Capacity Expectations:

Intel is currently the sole manufacturer of EMIB technology, with its Fab 42 facility in Arizona dedicated to EMIB and other production, and has invested $3.5 billion in Rio Rancho, New Mexico, to increase capacity for advanced packaging.

Despite the lack of specific wafer capacity data, we expect EMIB capacity growth to slow or even shrink as intel's current cash flow is severely scarce and the foundry business is not expanding as expected.

Compared with cowos, the current EMIB is only used for its own products and lacks popular products to drive demand growth, but its technical characteristics are not backward and it has the potential for rapid production expansion.

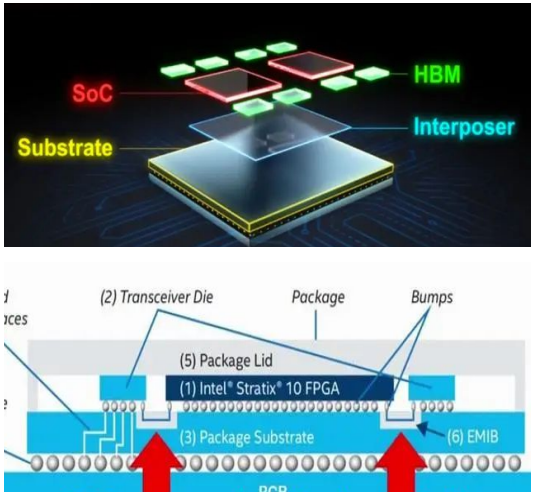




Technology Upgrade:

The next upgrade of EMIB is to reduce the solder bump spacing from 55 microns to 45 microns to improve interconnect density and performance, and will be combined with Foveros technology in the future.

Intel plans to implement more complex hybrid packages in 2025, such as the EMIB 3.5D, which is capable of supporting up to 47 active chips and more than 100 billion transistors



Cowos vs EMIB

Cost Quotation:

The cost is lower than the full-size silicon interlayer (like Cowos), the silicon bridge can be manufactured using the mature process to reduce cost, and also compatible with different process modules

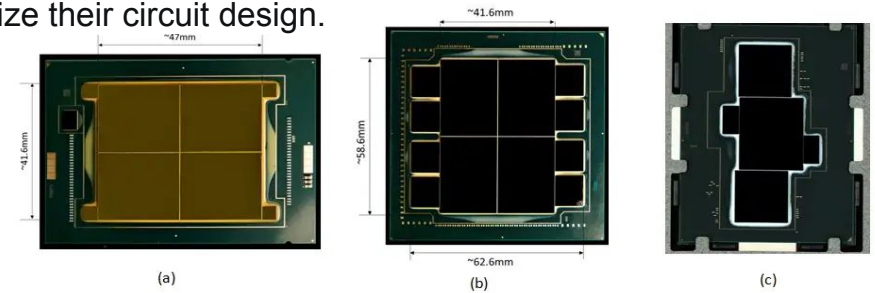
EMIB

(Embedded Multi-die Interconnect Bridge)

Expansibility:

EMIB occupies a small area, with high expansibility, not only can be compatible with a variety of process modules, the future can also be combined with other packaging technology, such as intel is ready to combine it with its own 3D packaging technology foveros, equivalent to a chip level wire, and does not need to occupy the area of the logic layer.

In the future, it can also be applied to BOM technology to help solve the heat dissipation problem in 3D packaging of logic chips. In addition, EMIB also supports considerable flexibility in bridge placement, which gives designers more options to optimize their circuit design.

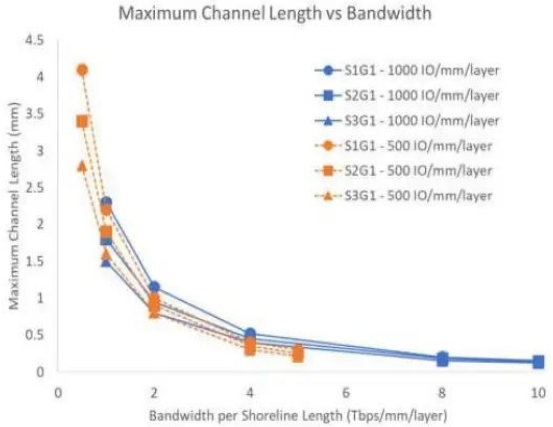


Various EMIB Package Configurations: In (a) and (b) the total Die Area is Substantially Greater than Reticle Size (c) shows an irregular and asymmetric Layout

BOM:

The metal interconnect layer is placed below the logic layer so that the logic layer with the largest heat generation is in direct contact with the heat sink.

EMIB
(Embedded Multi-die Interconnect Bridge)



Channel length across the EMIB bridge versus target datarate

Different signal/ground shielding patterns on the metal layers are evaluated.

Routing Density (IO/mm/layer)	Trace Width/ Trace Spacing (μm)		
	S1G1	*S2G1*	*S3G1*
1000	0.25/0.25	0.25/0.41	0.25/0.50
500	0.50/0.50	0.50/0.82	0.50/1.00

Very aggressive routing density configurations are used, with sub-um L/S.

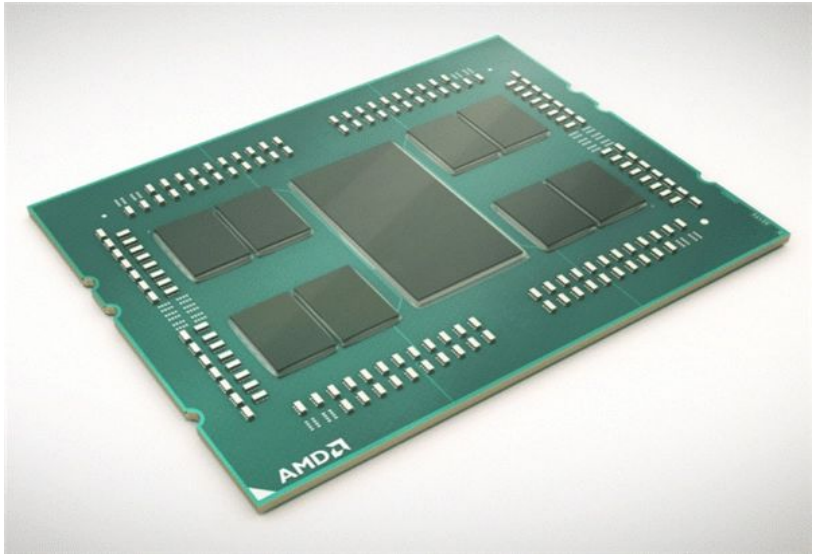
Interface Rate:

The silicon bridge of EMIB has a small footprint and does not affect the IO signal and power balance, enabling higher frequency communication.

The interface rate is closely related to the channel length and pin density of the silicon bridge, with the maximum allowed channel length decreasing rapidly as the bandwidth increases (see the above figure).

In high bandwidth scenarios, the interconnect distance of EMIB must be as short as possible to ensure the communication quality, which should be taken into account by the module designers.

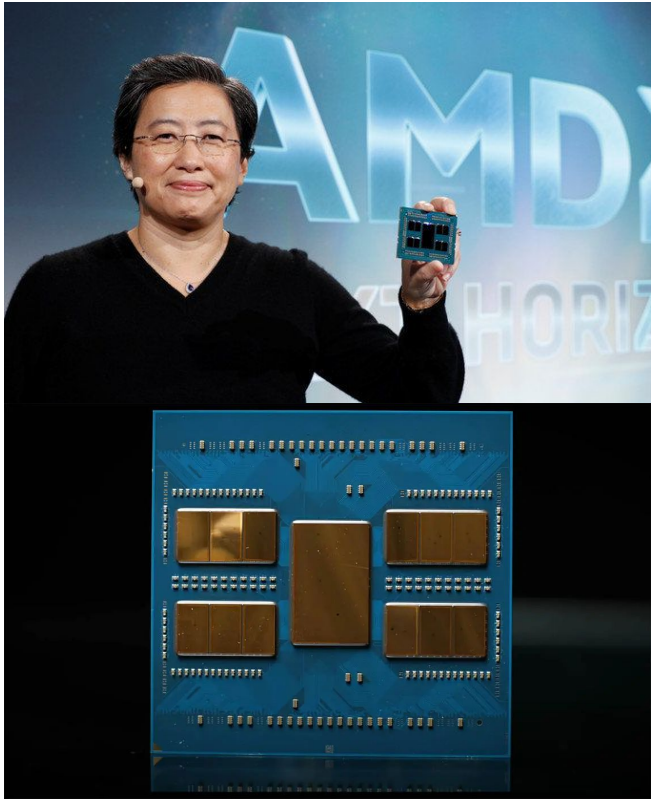
MCM Organic Substrate Base
（Multi-Chip Module ）



Capacity Expectations:

The Global Multi-Chip Modules (MCM) market is anticipated to rise at a considerable rate during the forecast period, between 2023 and 2031. In 2022, the market is growing at a steady rate and with the rising adoption of strategies by key players, the market is expected to rise over the projected horizon.

MCM Organic Substrate Base



(EPYC based on **MCM Organic Substrate Base**, the middle one is IO die, four pieces around is the computing cores)

Interface Speed and Scalability:

The MCM package connected by organic substrate is a 2D package, which is the lowest rate among the three types of technologies mentioned in this paper.

It is not suitable for high IO density module integration, such as integration of video memory and memory. At present, AMD epyc platform and intel Zhiqiang series are used.

In terms of expansibility, because the organic substrate belongs to the outermost connection mode, it is also the most compatible and expandable, and can be used with the previous 2.5D or 3D packaging technology.

Thank you !

The project is in progress!