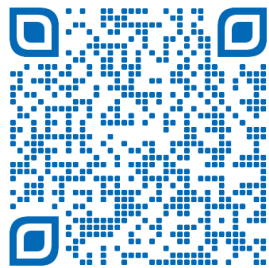




復旦大學
FUDAN UNIVERSITY

先进封装与集成芯片

Advanced Package and Integrated Chips



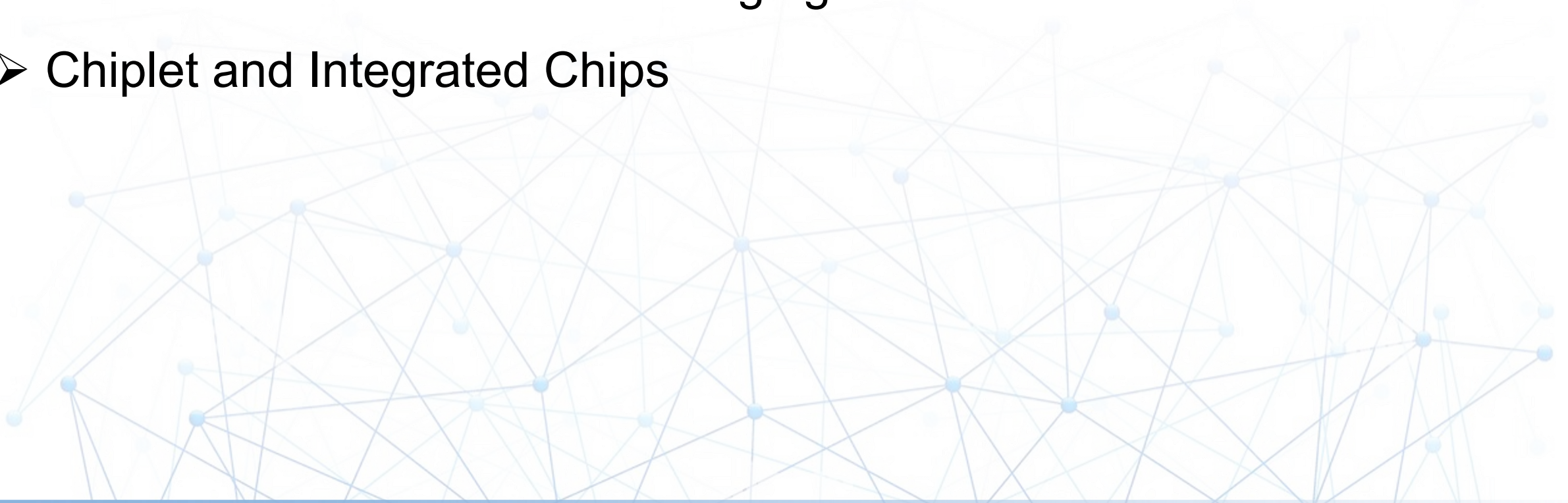
Lecture 1 : Introduction

Instructor: Chixiao Chen, Ph. D

Overview



- Course Overview
- Review of Conventional Packaging
- Introduction of Advanced Packaging
- Chiplet and Integrated Chips



Course Information

➤ Instructor I: **Chixiao Chen**

Email: cxchen@fudan.edu.cn

➤ Instructor II: **Wenning Jiang**

Email: wenningjiang@fudan.edu.cn



陈迟晓



江文宁

➤ Location: JA202, Jiangwan Campus

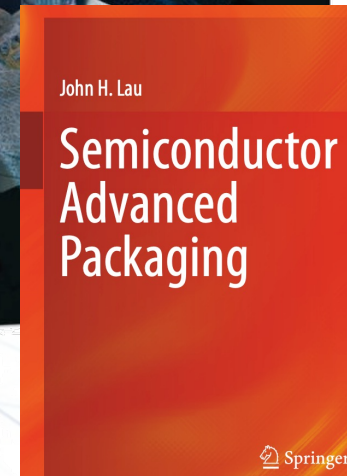
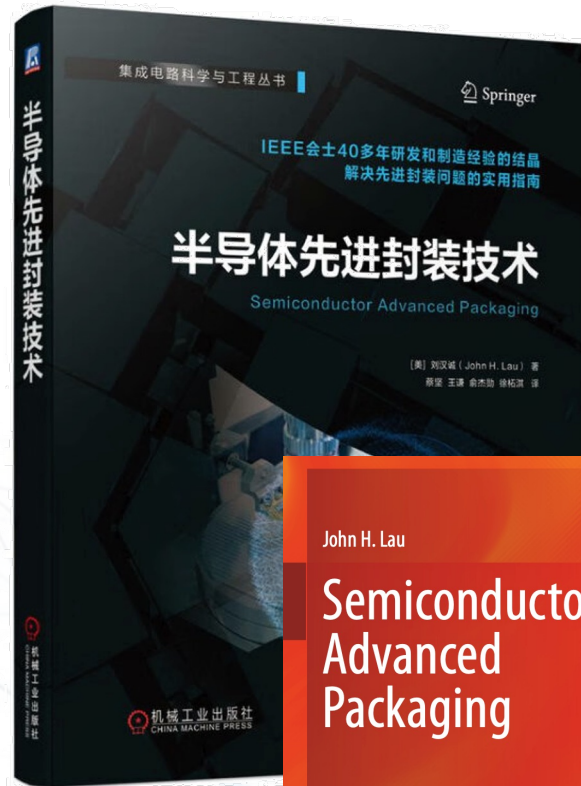
➤ Time: Monday Night 18:30 - 21:05

➤ Website: <https://cihlab.github.io/course/chiplet.html>

➤ WeChat Group Chat



- English version can be downloaded if you have Fudan account via Springer website.
- Chinese version can be purchased.
- High speed D2D Circuits and Systems:
<https://people.engr.tamu.edu/spalermo/ecen720.html>



Contents, Exam and Score



- Motivation: This course is not an packaging course, but a circuit and system design guide with advanced packaging knowledge.
- 3 parts:
 - 1) Advanced Packaging by Chixiao Chen
 - 2) Inter-chip communication circuit design by Wenning Jiang
 - 3) Integrated Chips Architecture by Chixiao Chen
- Exams:
 - 1) Option 1: Presentation (Related Papers form ECTC, ISSCC ...)
 - Option 2: Project Design (UCIe Package/PHY/Controller Design)
- Score: 20% Attendance/Quiz + 20 % Homework + 60% Presentation/Project

Schedule



- Lectures from 2.17 to 4.28
 - 4 weeks for Advanced Packaging
 - 4 weeks for die-to-die circuit design
 - 3 weeks for integrated chips system
- Project and Presentation: May 12/19/26
 - 5.5 is 劳动节-Holiday
 - 6.2 is 端午节-Holiday
- ISCAS Week (May 26-28) may affects pre.
- 清明节 校庆、运动会 do not affect us so far.

第二学期 2025年2月9日至2025年6月21日

周次	日	一	二	三	四	五	六	备注
0	2/9	10	11	12	13	14	15	9. 本科生线上申请补考, 2月12日至16日补考, 2月16日注册, 2月17日上课。
1	16	17	18	19	20	21	22	
2	23	24	25	26	27	28	3/1	10. 研究生线上申请补考, 2月12日至16日补考, 2月14日注册, 2月17日上课。
3	2	3	4	5	6	7	8	11. 妇女节、清明节、劳动节、青年节及端午节放假以学校办通知为准。
4	9	10	11	12	13	14	15	
5	16	17	18	19	20	21	22	12. 5月16日、17日第63届校园田径运动会暨第18届教工运动会, 5月16日停课一天。
6	23	24	25	26	27	28	29	
7	30	31	4/1	2	3	4	5	13. 5月27日建校120周年, 开展校庆学术活动等。
8	6	7	8	9	10	11	12	
9	13	14	15	16	17	18	19	14. 2025届本科生、研究生毕业典礼于第17周举行。
10	20	21	22	23	24	25	26	15. 通识教育课程考试安排在第16周, 第17、18周为停课考试周。
11	27	28	29	30	5/1	2	3	
12	4	5	6	7	8	9	10	16. 第二学期于2025年6月21日结束, 共计18教学周(包括考试)。
13	11	12	13	14	15	16	17	17. 2025年6月22日起开展各类本科生、研究生暑期教学活动。
14	18	19	20	21	22	23	24	
15	25	26	27	28	29	30	31	18. 研究生寒假、暑假时间由院系和导师根据培养计划妥善安排。
16	6/1	2	3	4	5	6	7	19. 教职工原则上每学期提前一周上班, 延后一周开始寒暑假轮休。具体时间安排由学校办另行通知。
17	8	9	10	11	12	13	14	
18	15	16	17	18	19	20	21	

Electronic Packaging

Electrical connectivity:

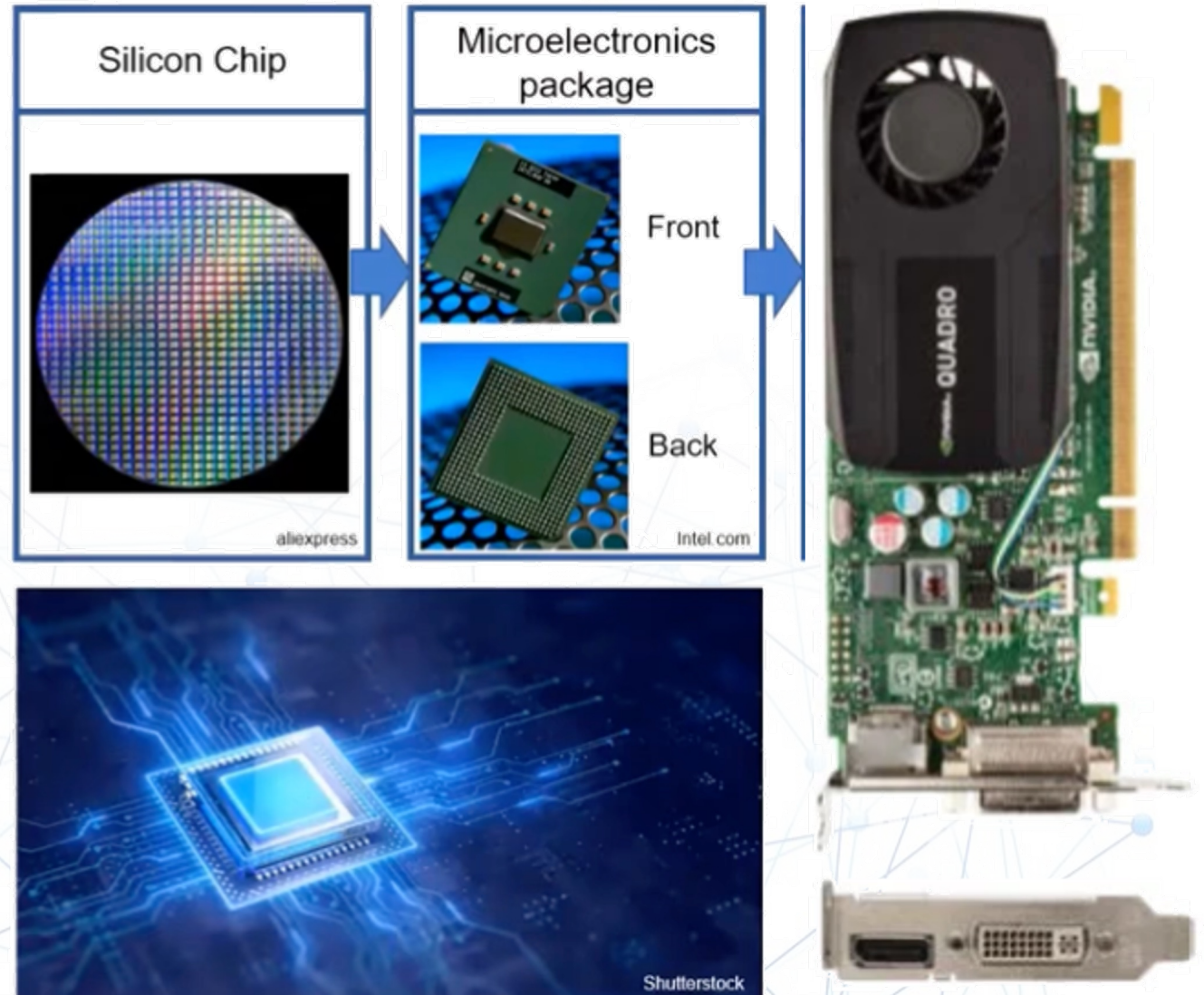
- Power distribution to all components and chip circuits.
- Connectivity between various chip circuits, devices and components.

Thermal sink:

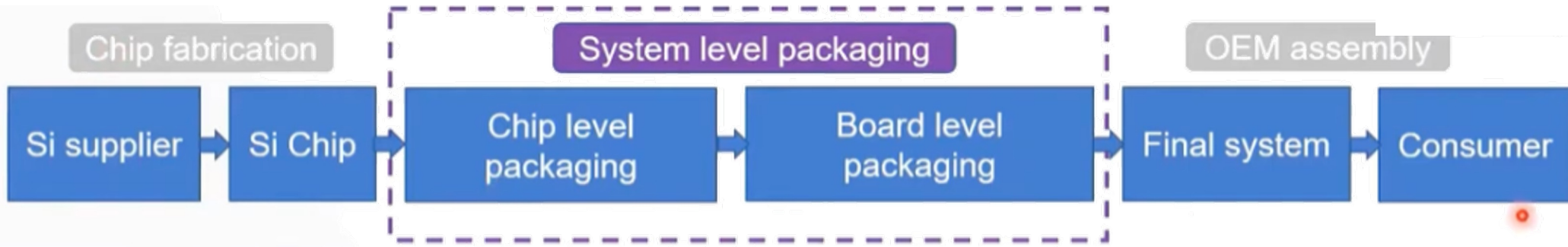
- Removal of heat generated by chip circuits during processing for continued performance.

Mechanical protection:

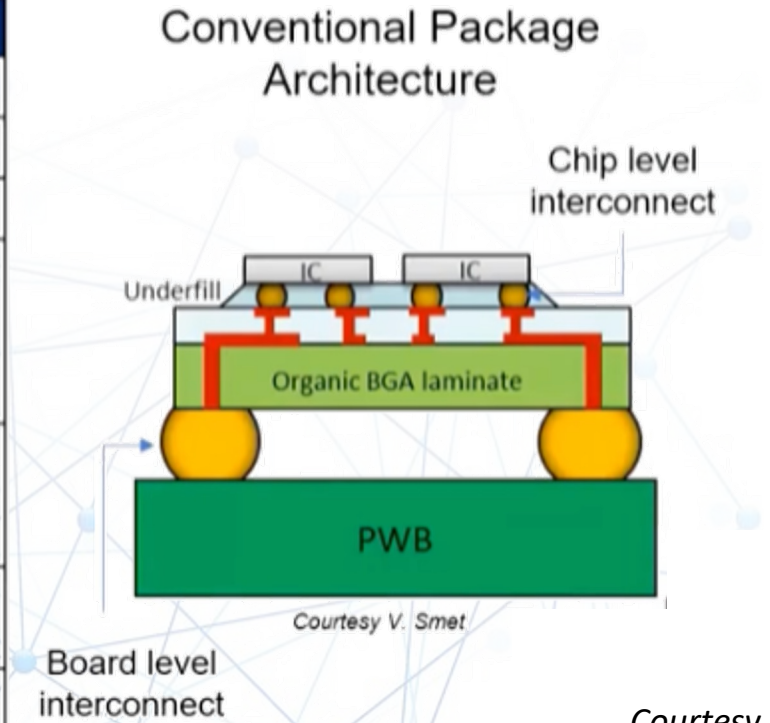
1. Mechanical structure and stability to allow for manufacturability.
2. Protection of devices and components from environmental exposure and damage.
3. Shielding chip circuits from external electromagnetic (EM) radiation or interference.



Interconnect via System Packaging



Parameter	Chip level interconnect	Board level interconnect
Level	Chip-to-package	Package-to-board
Pitch	10s of um to ~200um	Hundreds of um to ~1000 μm
Bump	Plating, jetting, dip transfer	Screen-printing, ball drop, jetting
Assembly	2D conventional, 2.5D w/ organic RDL 2.5D w/ Si interposer 3D Chip-wafer, wafer-wafer attach	Reflow, mount
Materials	Solder, flux Underfill, mold Thermal interface materials	Solder, flux Underfill (for limited applications)
Interconnection rework	Limited	Most applications
Cost	Low-moderate (depending on assembly architecture)	As low as possible



Courtesy
3DPRC, GATech

Conventional IC Package Types

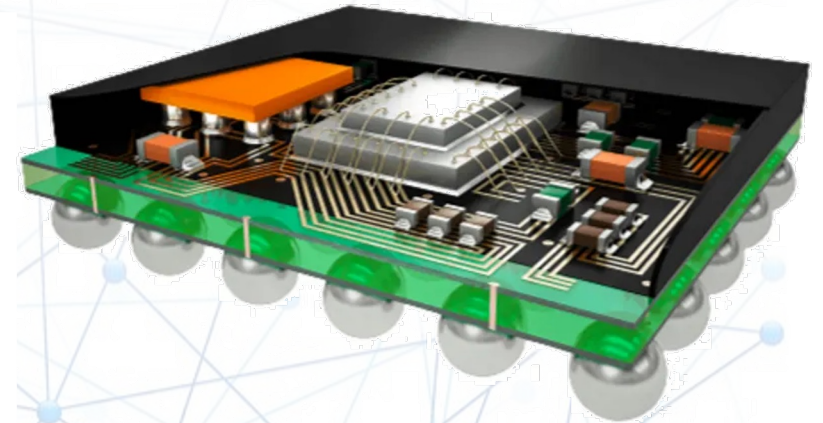
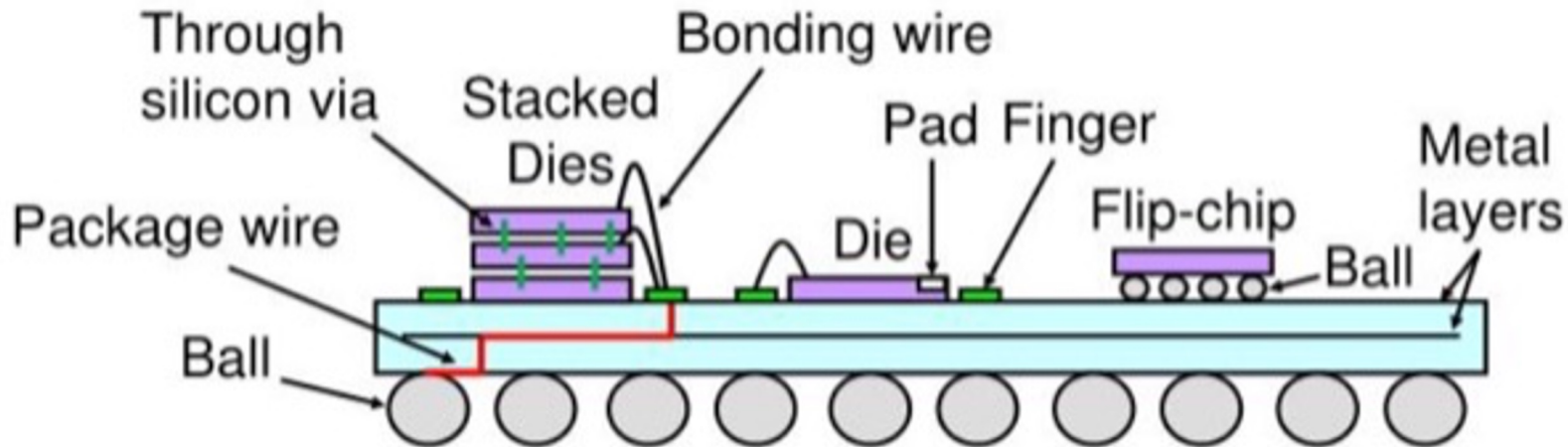
- There is a number of package types, among which BGA (ball grid array) has the most interconnect.



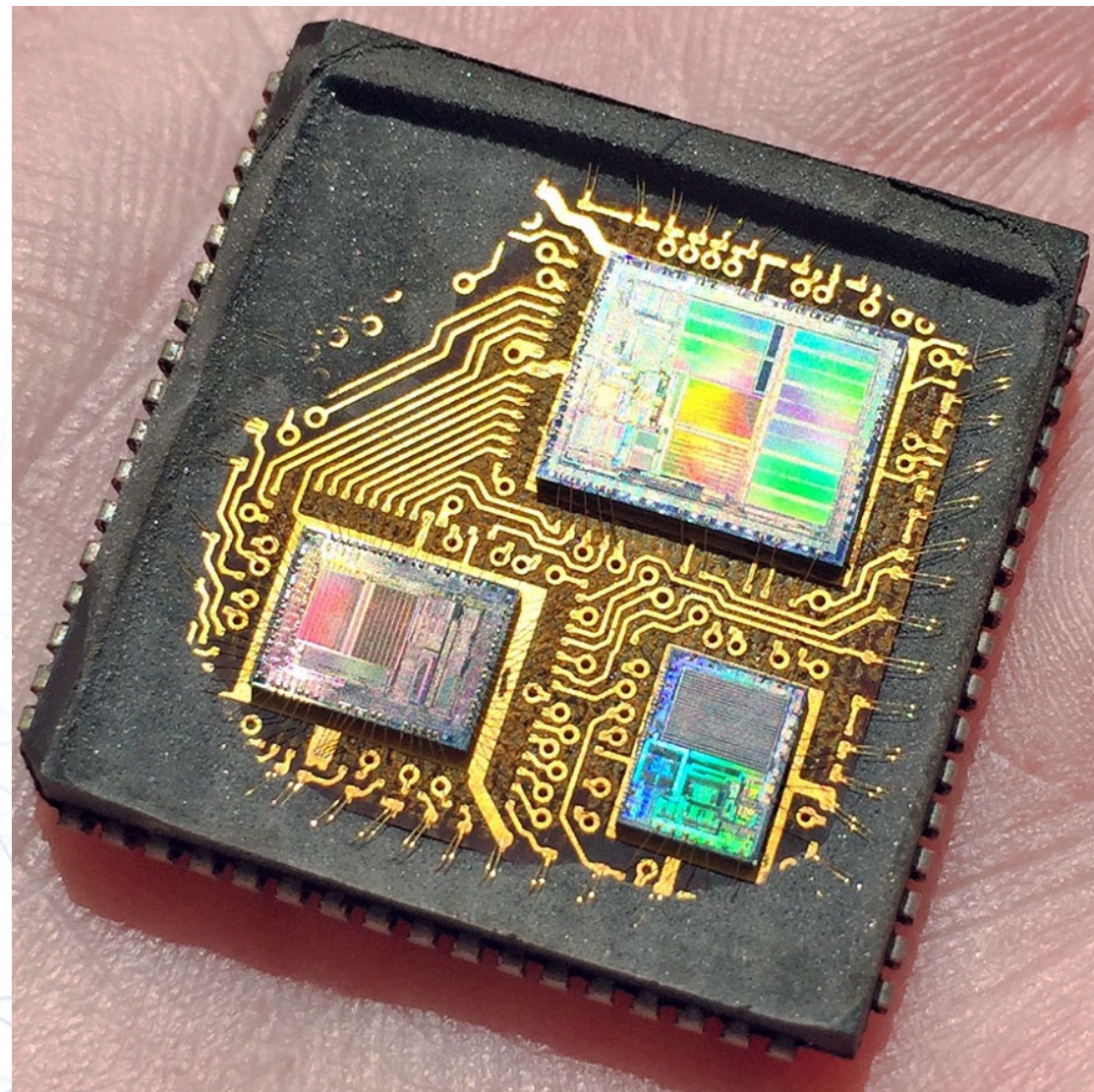
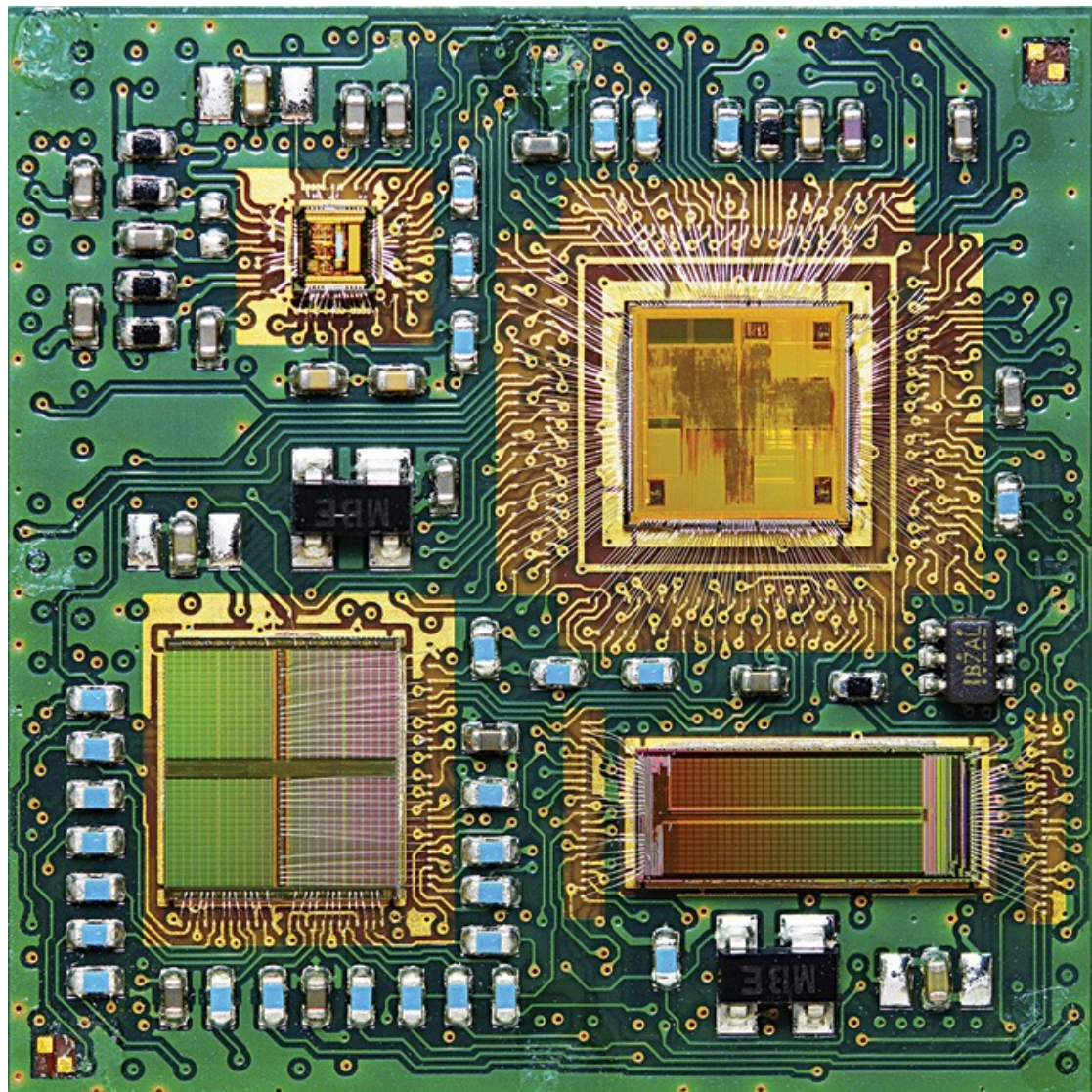
System-in-Package



- For smaller form-factor, and higher performance, multiple dies and flip-chips are placed within one substrate
 - Die-stacking was available
 - Locate fingers are placed around each die with minimal wire bonding
 - multiple-layer of substrate using new special films, Ajinomoto Build-up Film

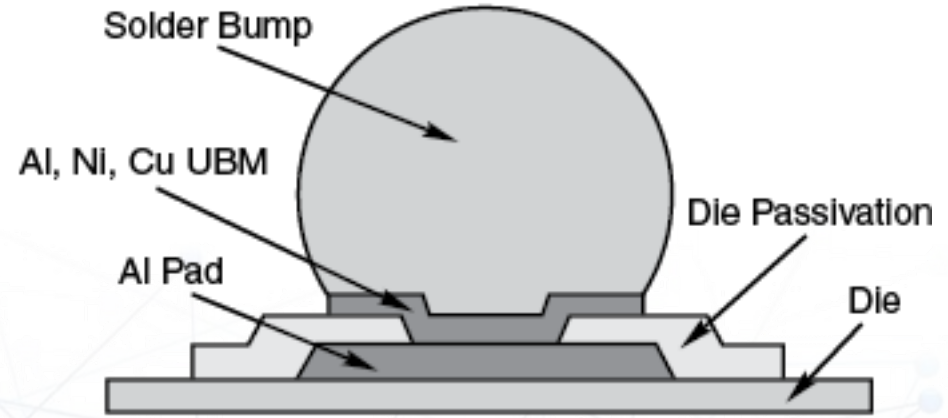


SiP Gallery (From Internet)



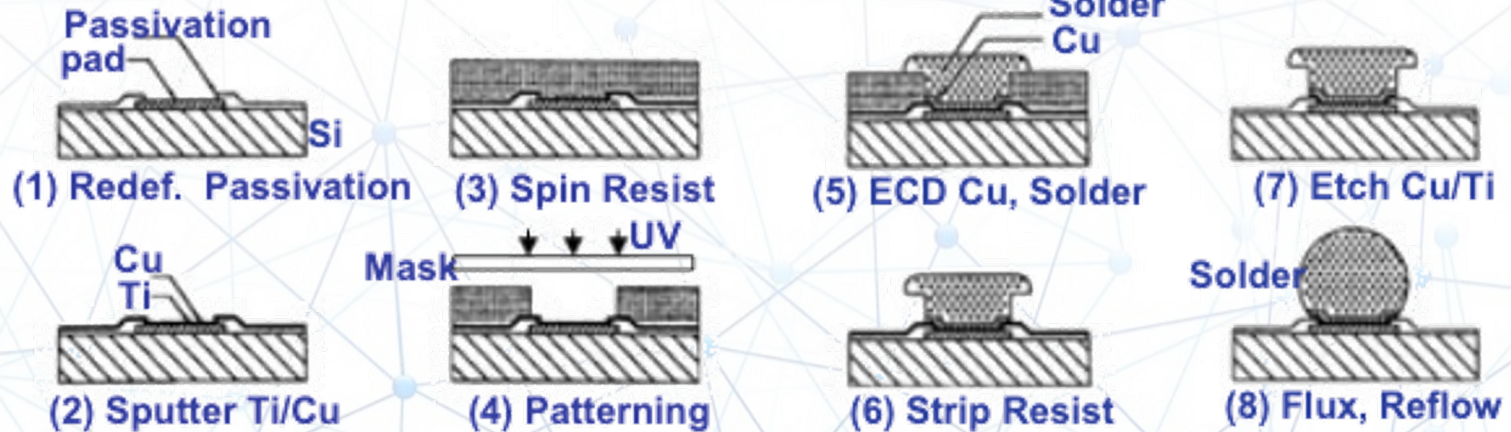
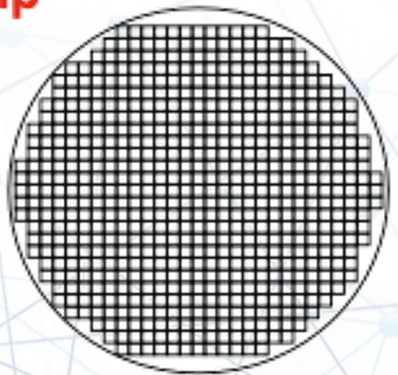
Flip Chip and C4 Bump

- Today, the **controlled collapse chip connection (C4)** bumps are the most used interconnection structure.
- C4 bumps uses SnAg and has a pitch of 150 μm , and 90 μm pad opening



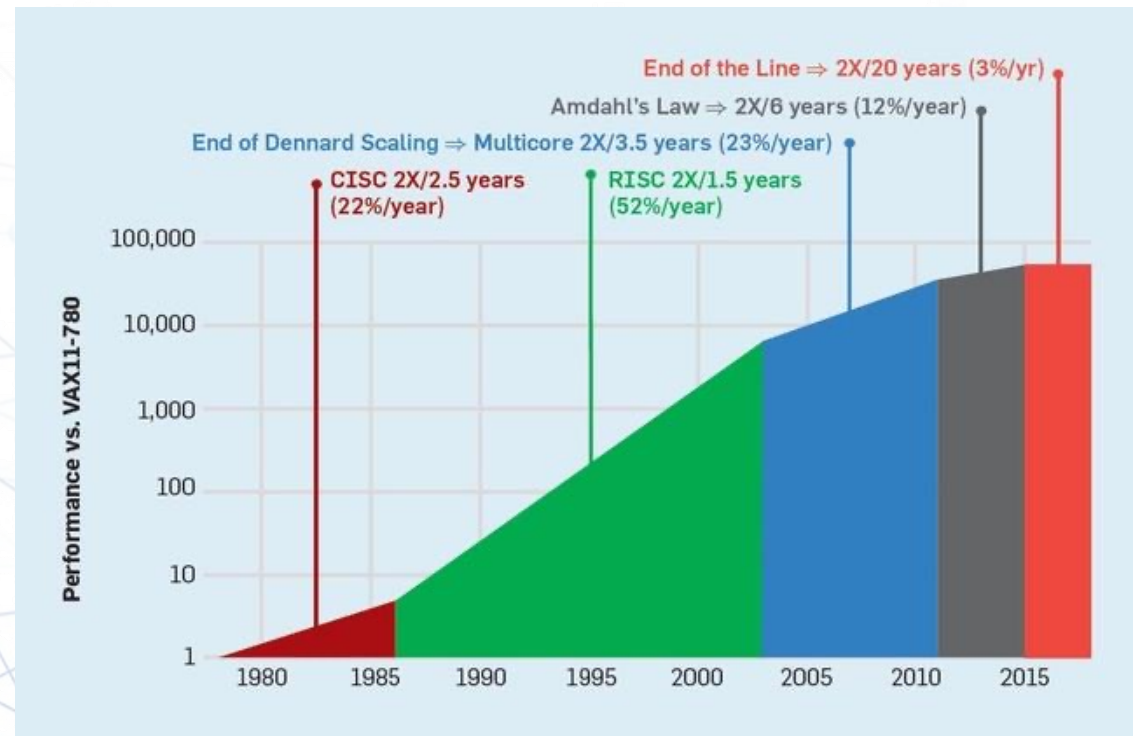
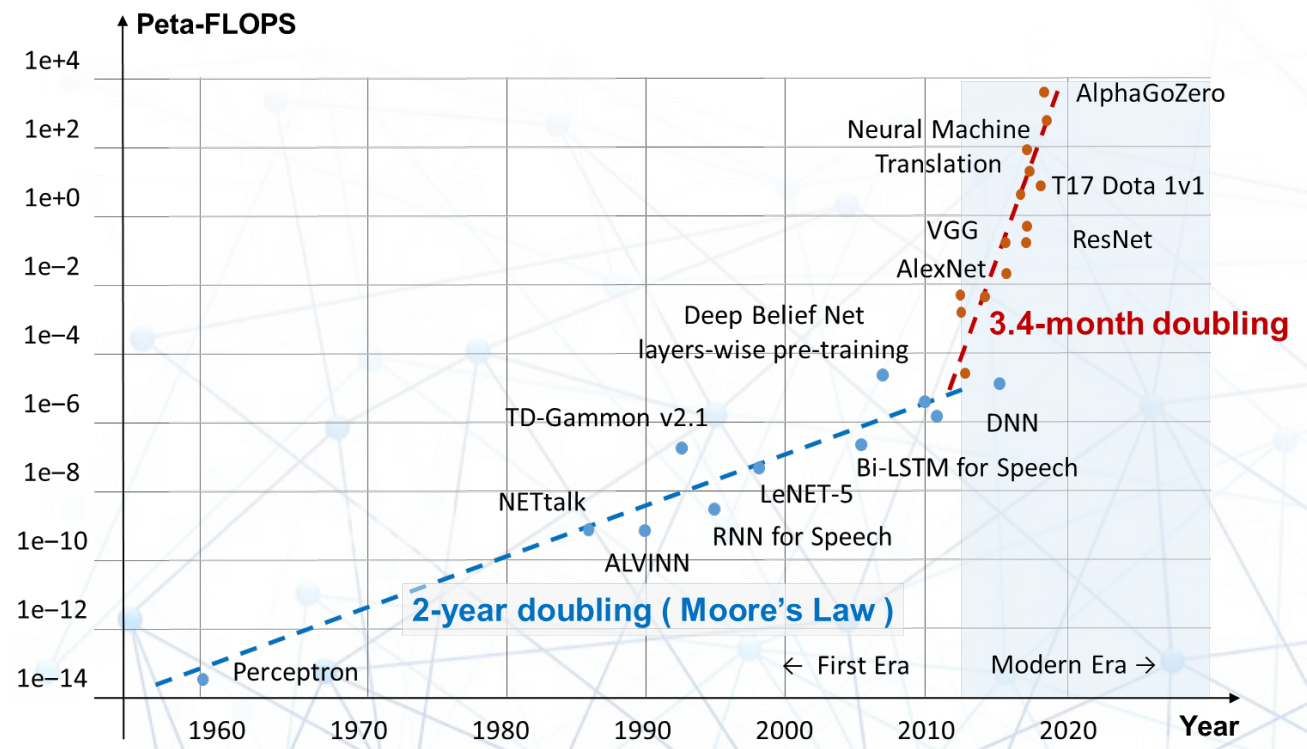
(a) C4 bump

Device Wafer



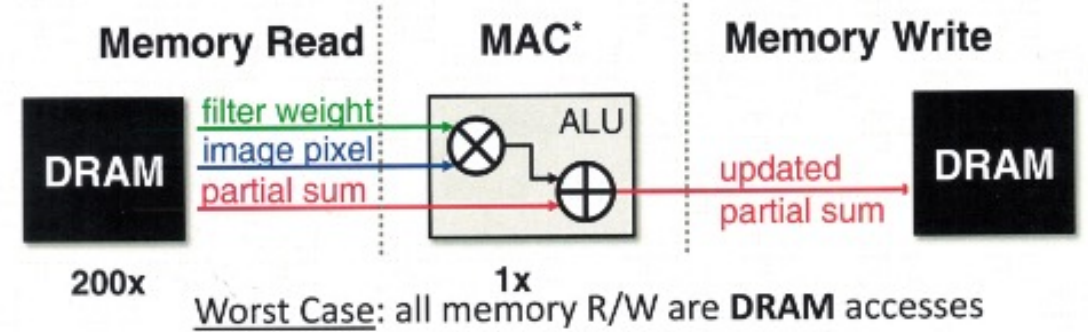
Why We Need More (Than Moore)?

- Deep Learning Algorithm doubles per 3.4 months (10+x every year)
- Moore's Law doubles per 1.5~2 year
- Von-Neumann Architecture (CPU) doubles per 20 years

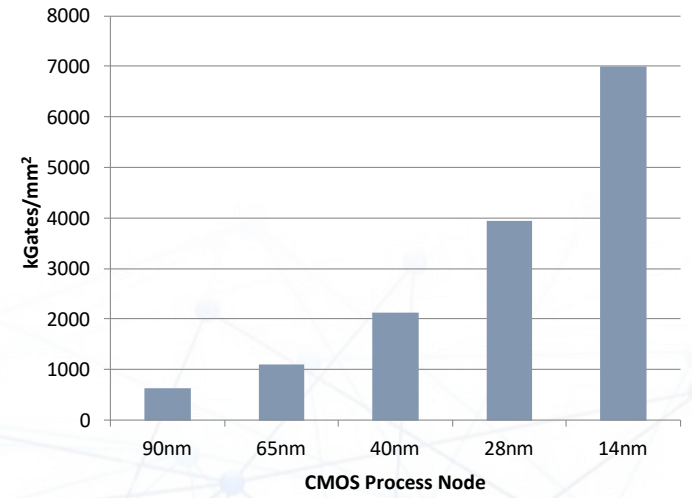


The Interconnect Problem / Memory wall

➤ High performance computing requires both more logic gates and higher rate processor-memory communication.



- Example: AlexNet [NIPS 2012] has 724M MACs
→ 2896M DRAM accesses required



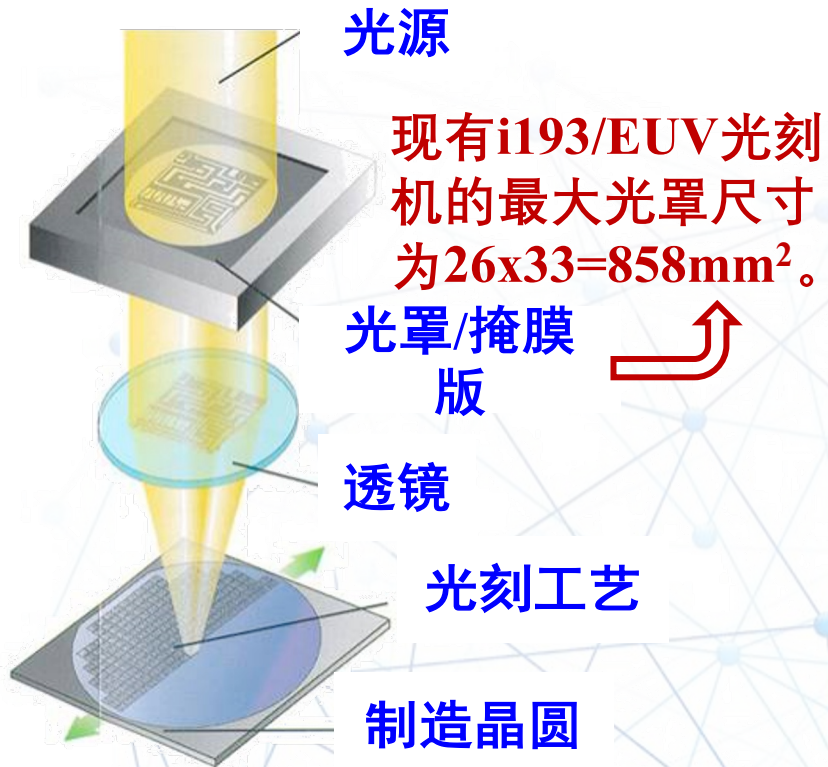
➤ But memory interface grows more slower than Moore's Law, 16x in 20 years.

➤ Logic density increase by 1000x in 20 years.

DDR SDRAM Standard	Release Year	Prefetch Buffer Size	Vdd	Maximum Transfer Rate (MT/s)	Chip Density
DDR1	2000	2n	2.5	200-400	256Mb-1Gb
DDR2	2003	4n	1.8	400-1066	512Mb - 4Gb
DDR3	2007	8n	1.5	1066-2400	1Gb-8Gb
DDR4	2014	8n	1.2	2133-4800	4Gb-32Gb
DDR5	2021 (expected)	16n	1.1	4266-6400	16Gb-32Gb

Monolithic SoC Performance Limit

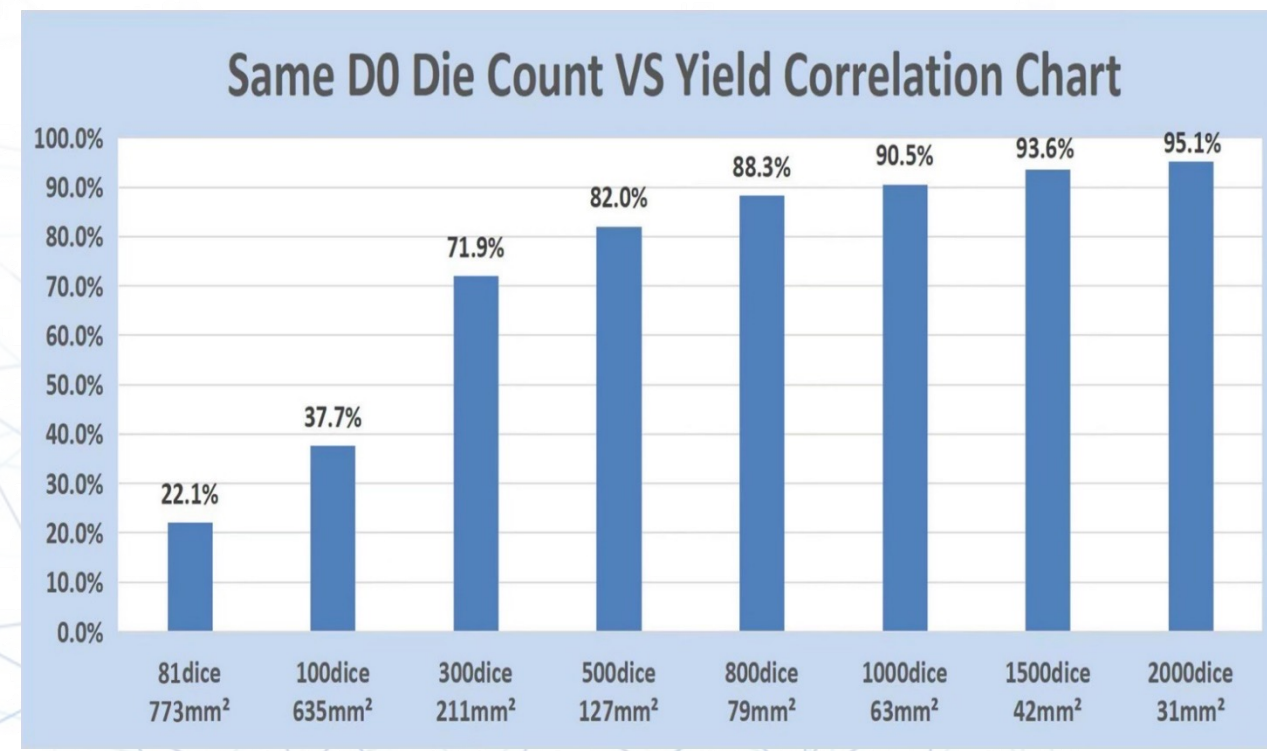
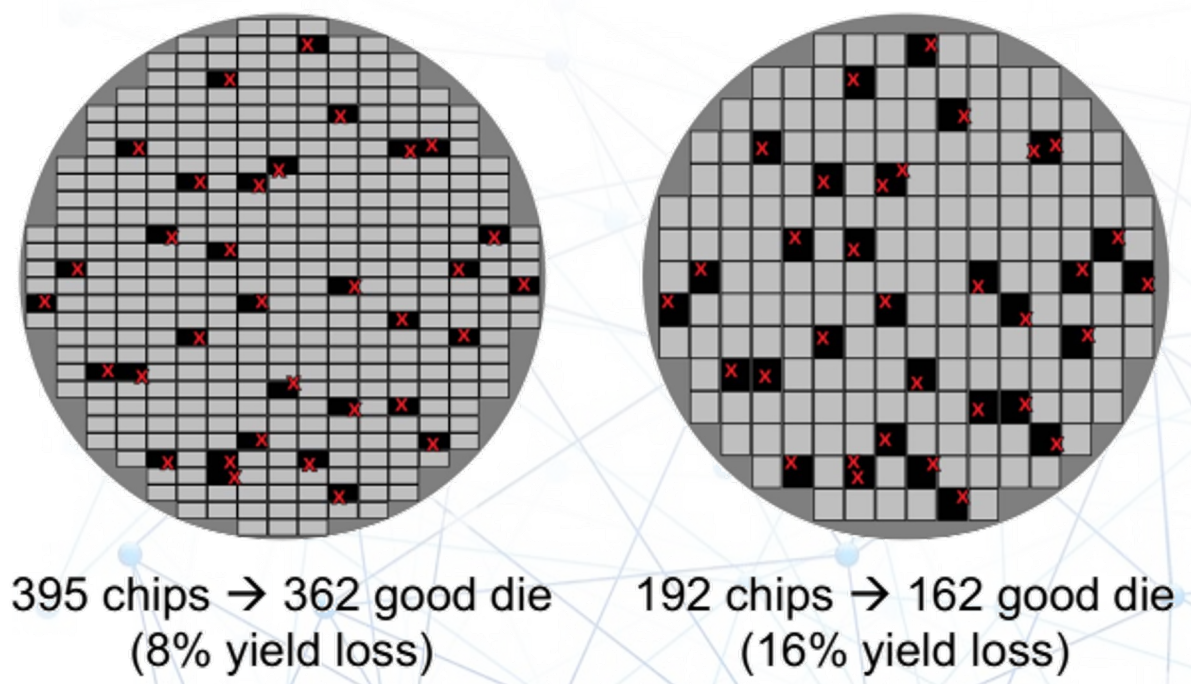
- Monolithic (Single Chip) SoC is limited by the maximum fabrication area due to lithography, known as reticle size (since DUV, it has been 858 mm²)
- State-of-the-arts GPU maintains its area near 820 mm².



GPU	NVIDIA H100	NVIDIA A100	NVIDIA V100	NVIDIA Tesla P100
晶体管数	80B	54.2B	21.1B	15.3B
芯片尺寸	814mm ²	828mm ²	815mm ²	610mm ²
架构	Hopper	Ampere	Volta	Pascal
工艺节点	TSMC 4N	TSMC N7	12nm FFN	16nm FinFET+
GPU 集群数	132	108	80	56
CUDA 核心	16,896	6912	5120	3584
年份	2022	2020	2017	2016

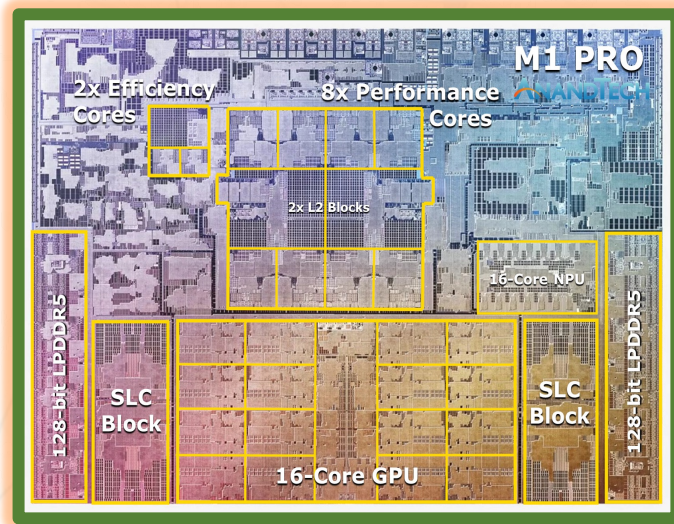
Yield Limit of the Large Chips

- Advanced technology involve high cost due to yield loss: If chip area is greater than 500 mm², yield is normally less than 20%.
- 200 mm² is selected as a sweet yield point for most advanced technology.

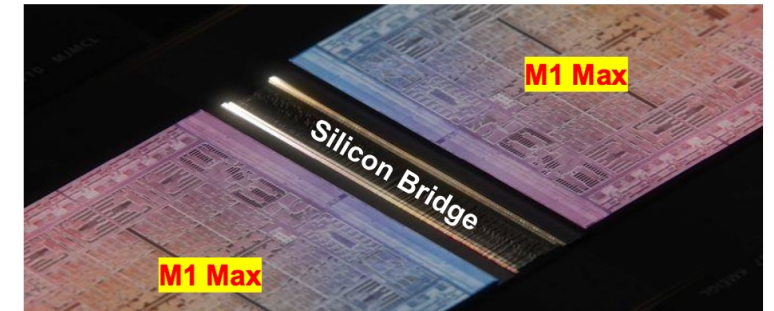


From SoC to Integrated Chips

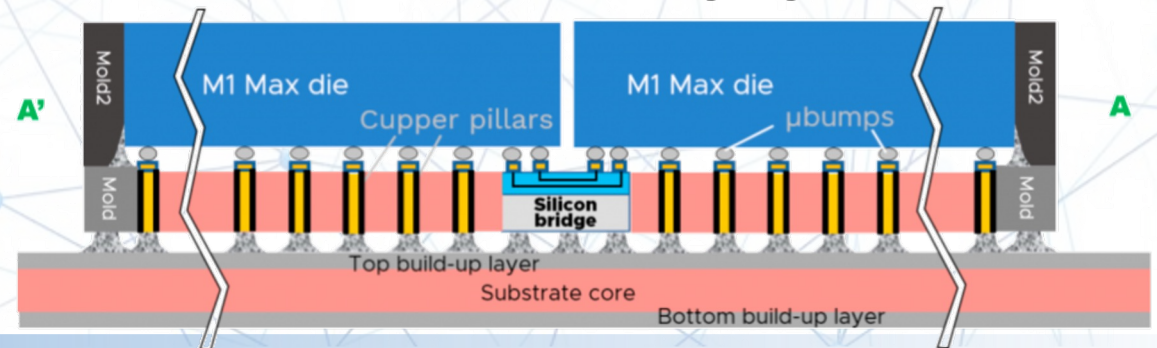
- System-on-Chip (SoC): Integrating multiple IPs and billions of transistors inside a monolithic die.
- Integrated Chips: Design and fabrication **Chipllets** with certain functions, and integrating them into a larger scale system by advanced fabrication technology.



Apple M1 Processor Monolithic SoCs

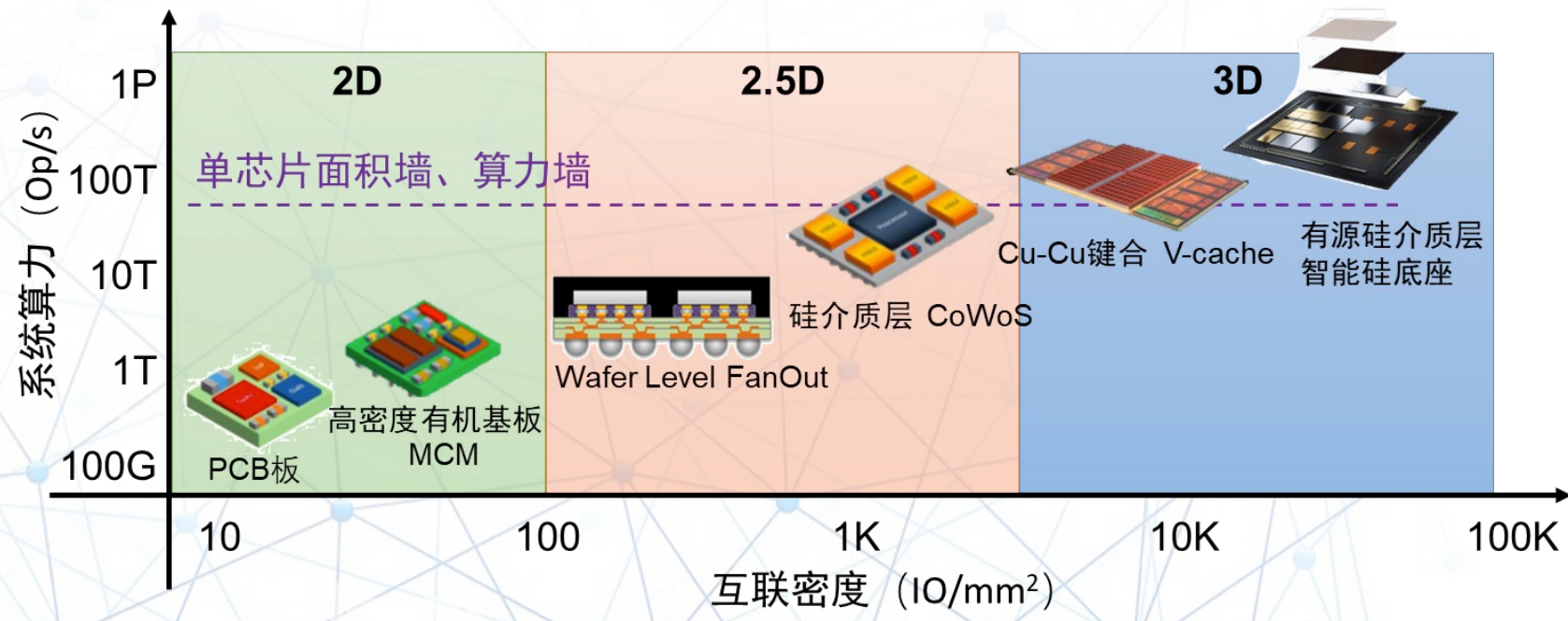
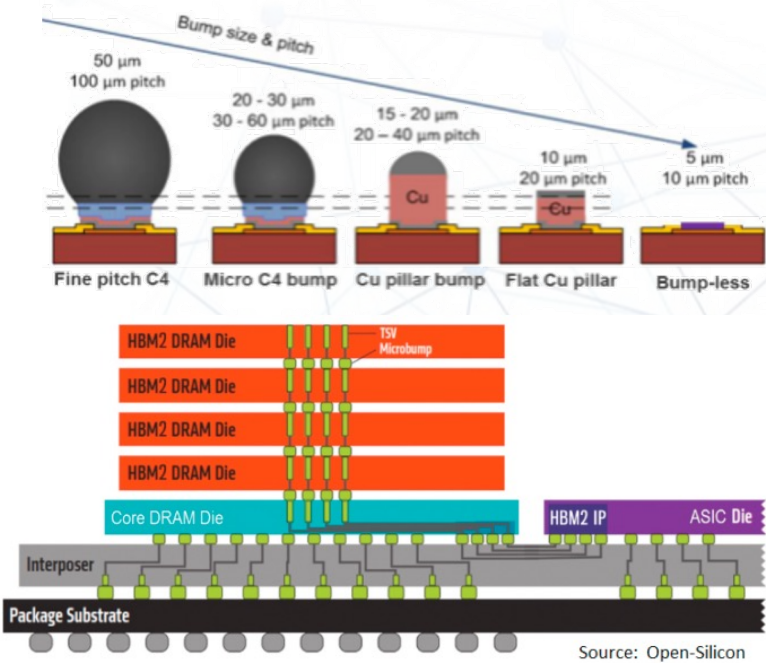


Apple's M1/M2-Ultra chip bridging two processors



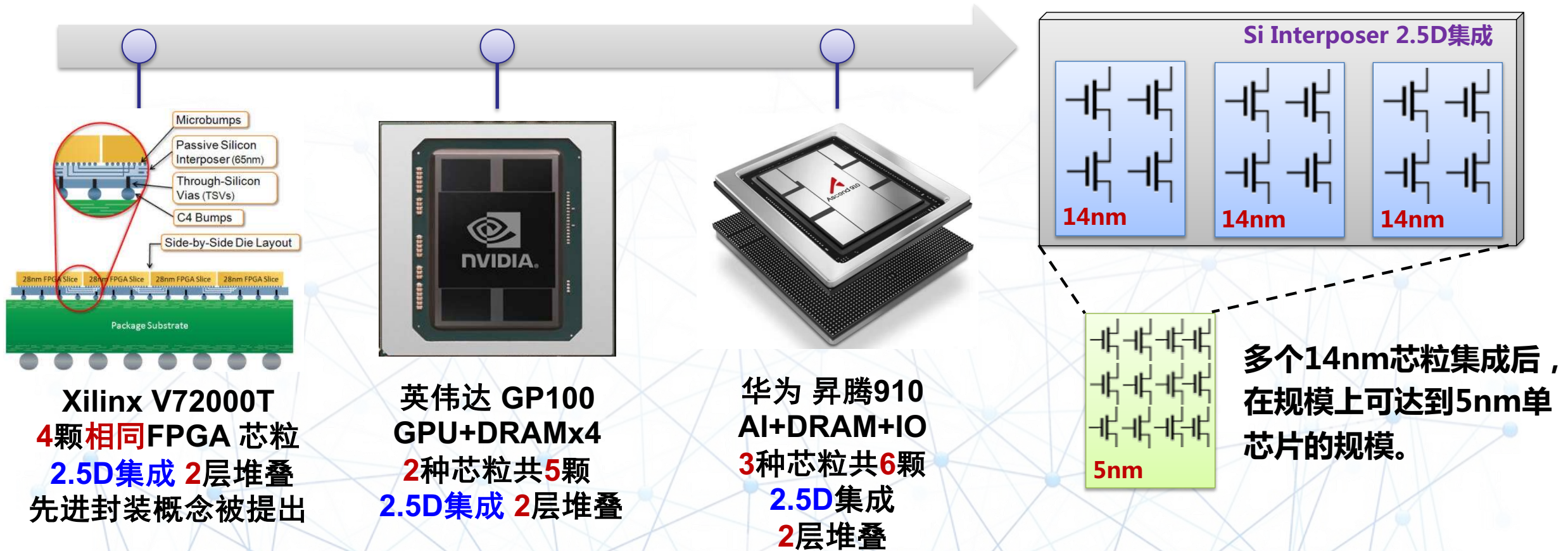
Roadmap of Heterogenous Integration

- Compared with conventional PCB/SiP substrate, advanced packaging introduces semiconductor fabrication technology to implement finer interconnect
- The bump size is reduced to 10um pitch, the line width/space is reduced to sub 1um , therefore the number of interconnect can improve by 10-1000x.

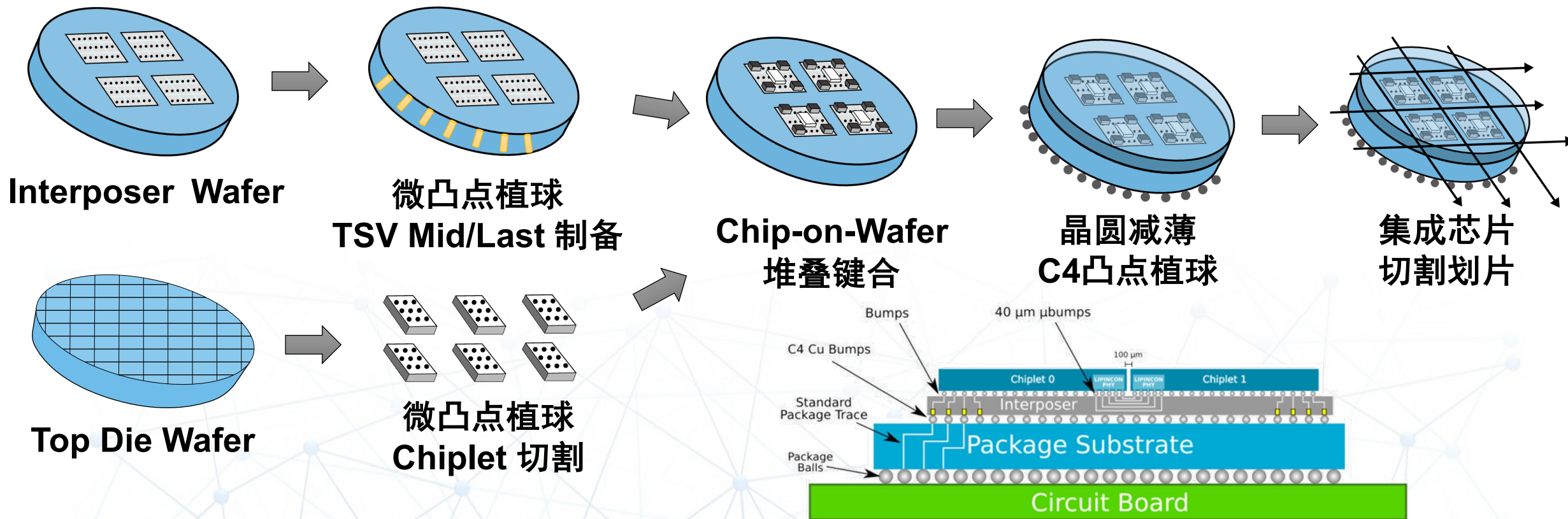


2.5D Integrated Chips improves Performance

➤ 2.5D集成技术基于半导体工艺的无源硅Interposer/RDL Interposer/硅桥，大幅缩小互连凸点的尺寸/间距，**增加互连线密度和通信带宽**，并降低延时。

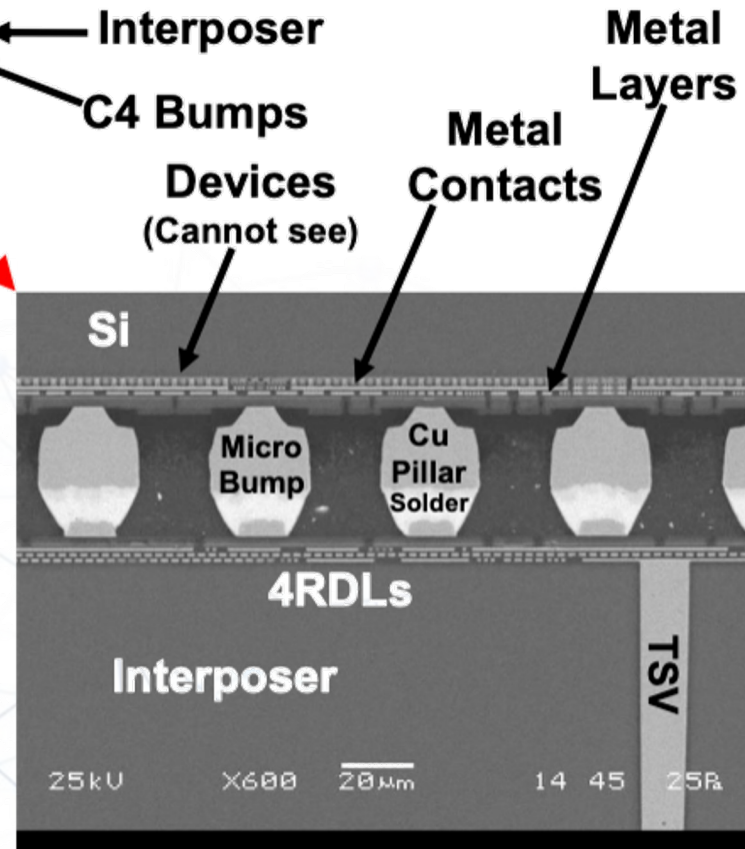
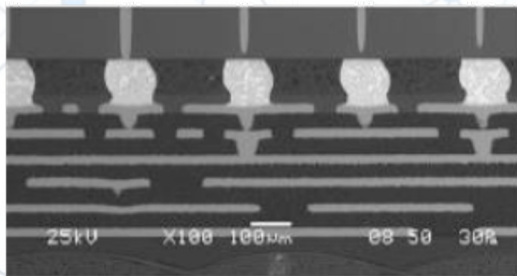
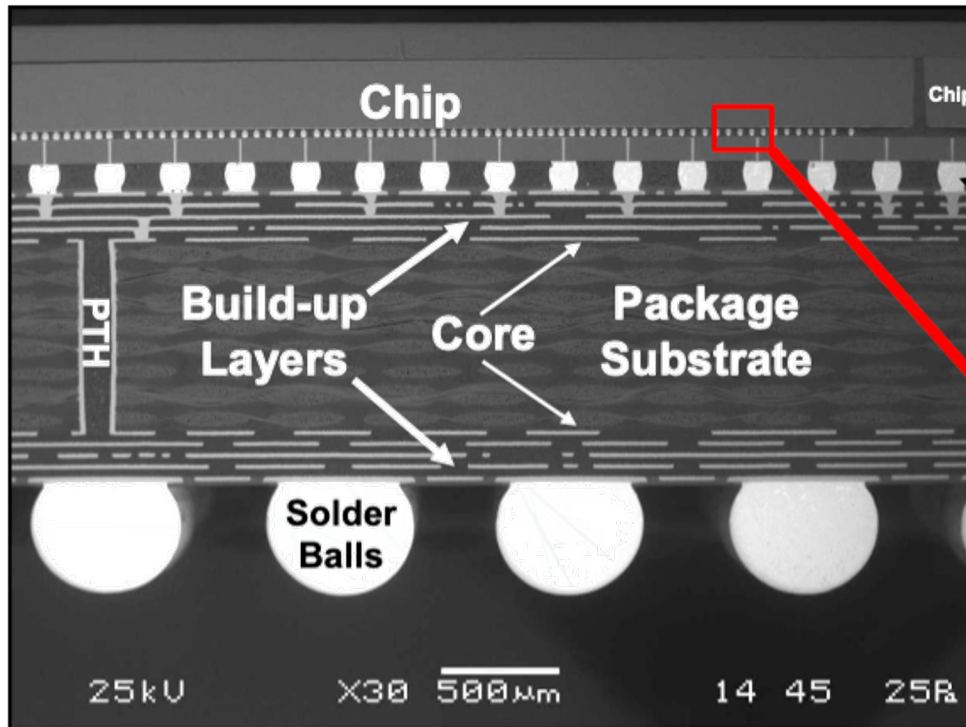


Technology of Advanced Packaging



- Silicon interposer (semiconductor wafer) replaces the conventional substrate in advanced packaging.
- Die to wafer bonding / through silicon via (TSV) become new technologies.

Cross section photo of 2.5D Integration



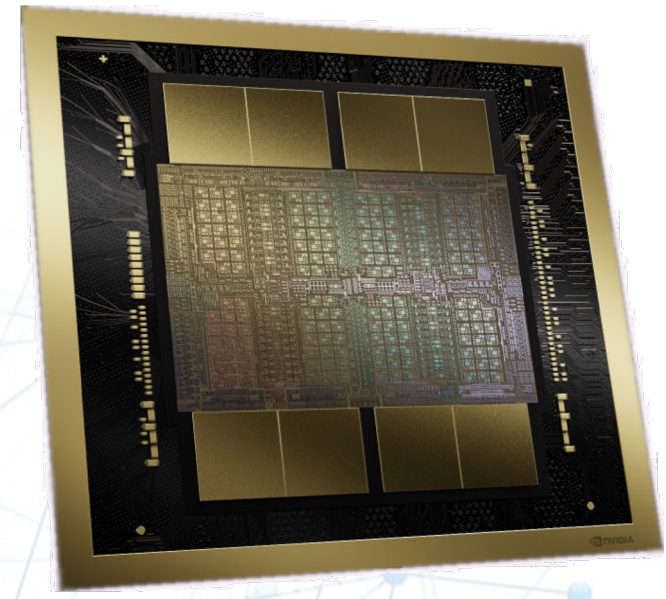
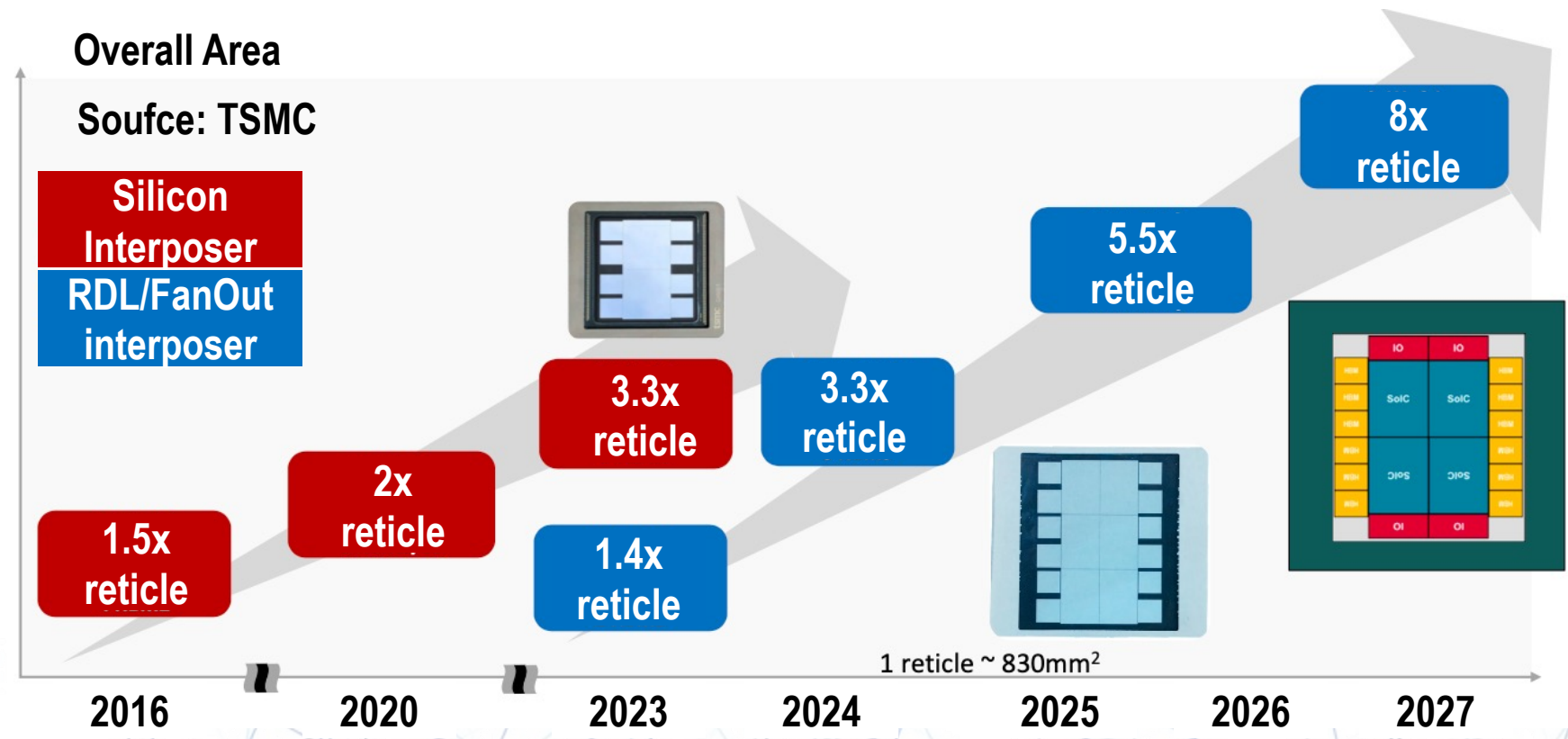
- Note the Pitch size
- Micro bump – 40µm
 - C4 bump – 130 µm
 - Solder ball – 0.5mm

The package substrate is at least (5-2-5)

- RDLs: 0.4µm-pitch line width and spacing
- Each FPGA has >50,000 µbumps on 45µm pitch
- Interposer is supporting >200,000 µbumps

2.5D Integration Exceeds the Area Limit

- The computing performance inside one package can be continuously scaling through 2.5D integration, another path beyond Moore's Law.

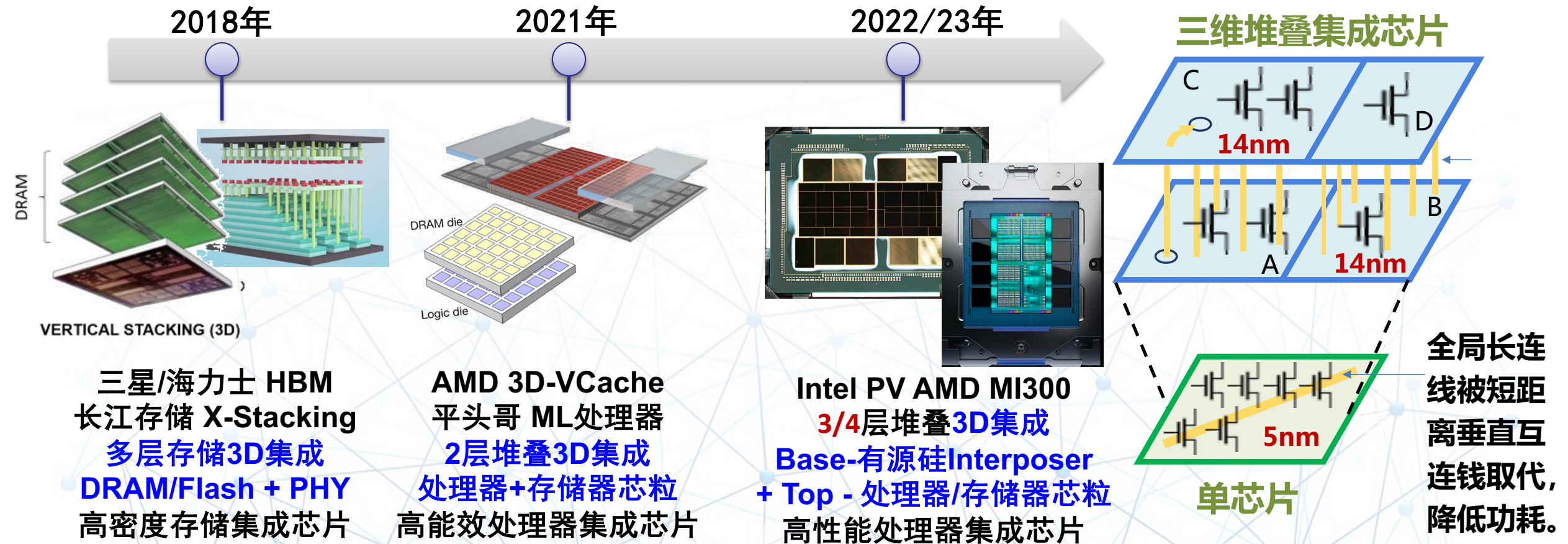


Nvidia B200 GPU with 2 GPU dies and 8 HBM dies

3D Integrated Chips improves Performance



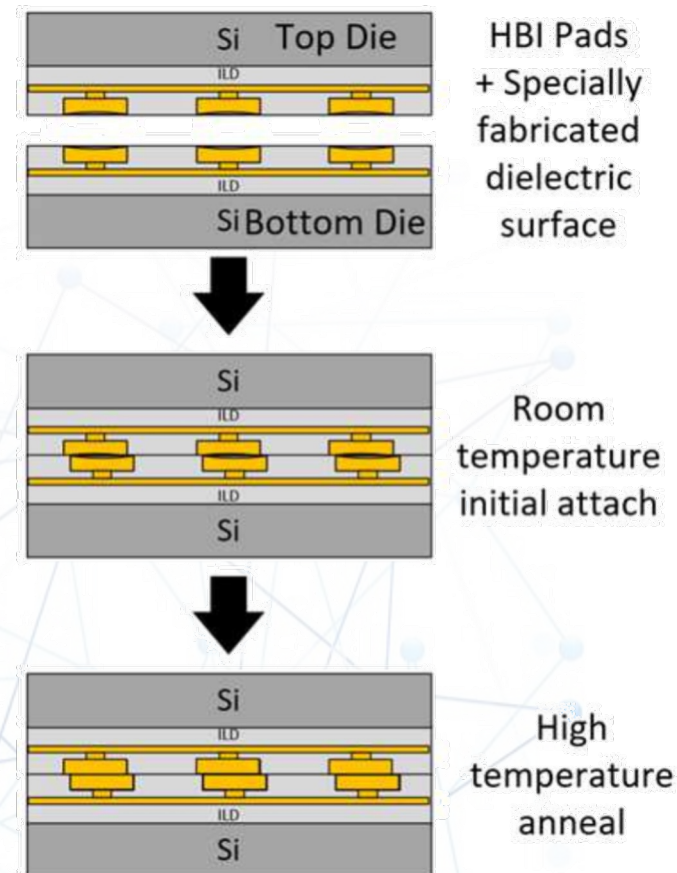
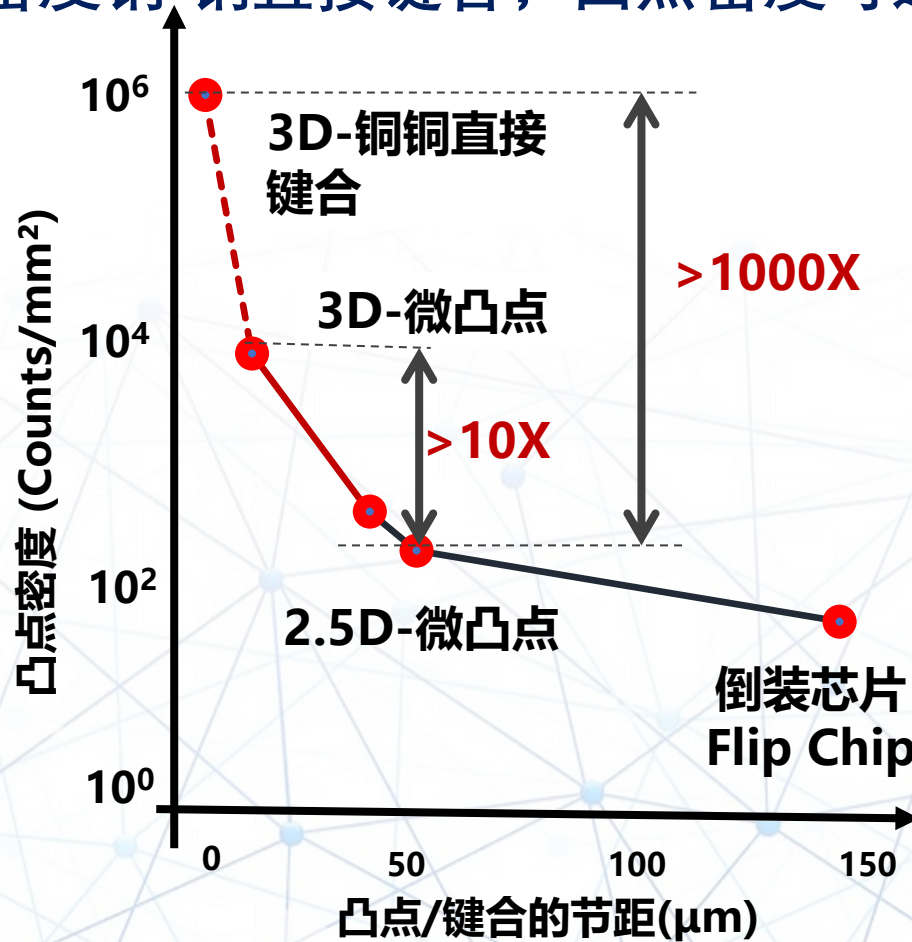
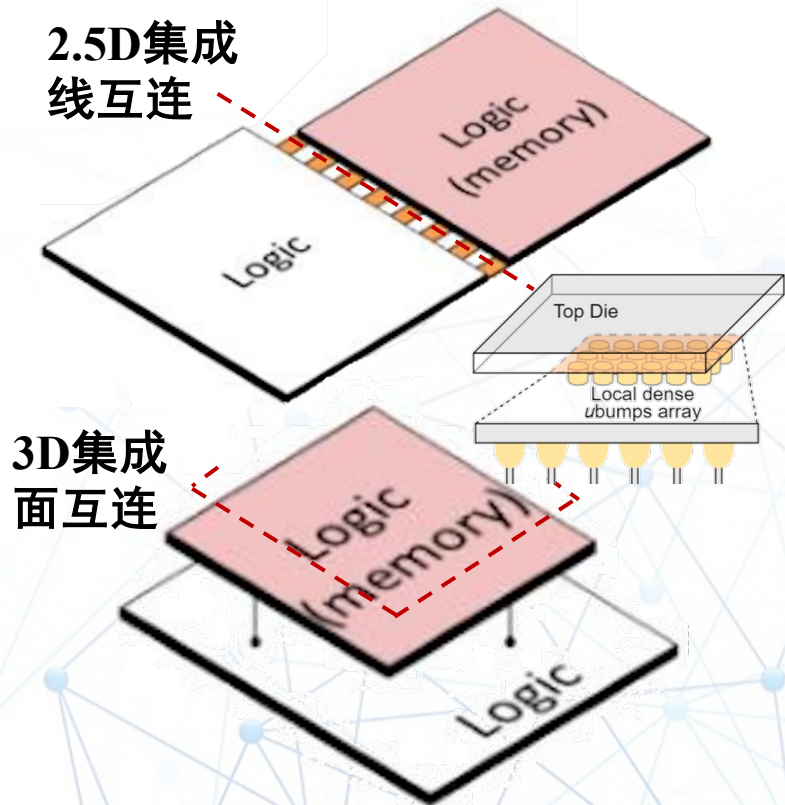
- 3D集成技术通过2个以上芯片的堆叠，实现升维，提升投影面积上的晶体管密度。
- 基于上述方法，集成芯片的能效提升接近于“摩尔定律”尺寸微缩。



3D Integrated Chips improves Performance

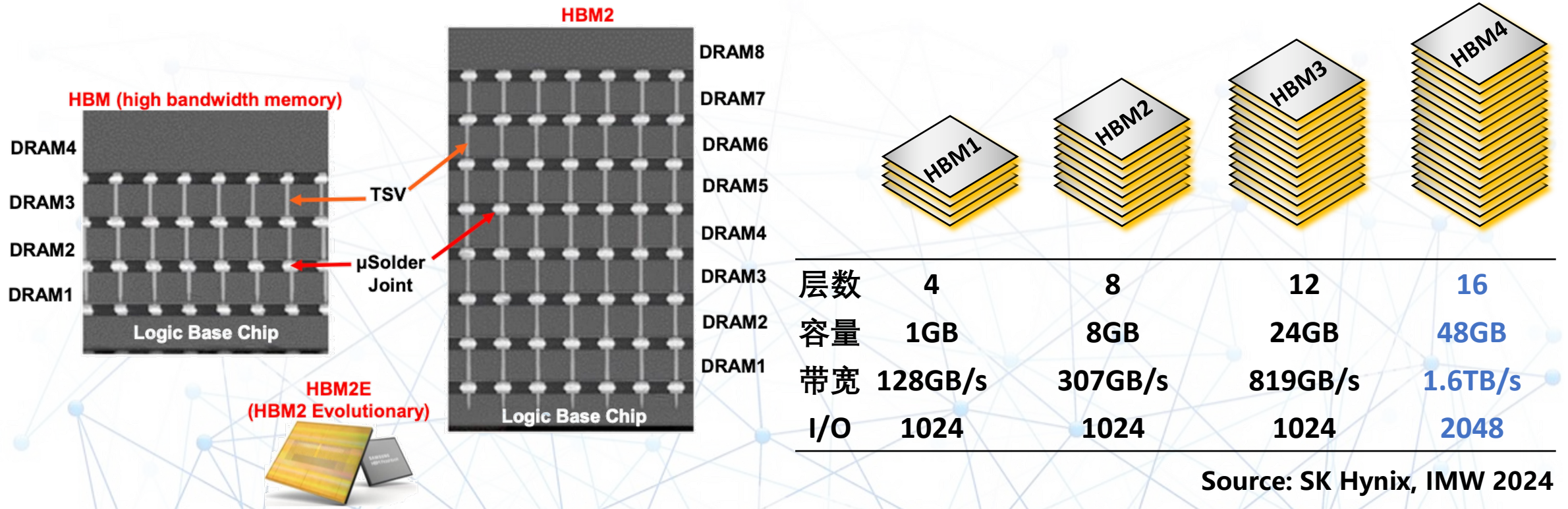


- N层三维堆叠集成芯片的**逻辑与存储算力密度**可提升N倍，（N=2时等效2-3代微缩）
- 三维堆叠的互连界面支持高密度铜-铜直接键合，凸点密度可达 10^6 个/ mm^2 。



3D Stacking Memory Technology

- High bandwidth memory (HBM) is a type of multi-die stacking DRAM for 2.5D integrated systems.
- It is evolved from HBM1 to HBM 4, which has 16 die stacking.



Integrated Chips is a major trend in HPC chips



➤ Using advanced packaging for high performance computing chip design is a very **HOT** topic.

排名	2023 新进超算一	2023 新进超算二	2022 Top 500 第一	2022 Top 500 第二
超算中心	 <p>美国 橡树岭 Frontier</p>	 <p>美国 劳伦斯 EL CAPT</p>	 <p>美国 橡树岭 Frontier</p>	 <p>日本 富岳 Fugaku</p>
总算力	2 EFLOPS	2 EFOPS	1.102 EFLOP	0.442 EFLOP
芯片组	 <p>Ponte Vecchio</p>	 <p>MI300/300X</p>	 <p>AMD Zen3+MI250X</p>	 <p>Fujitsu AF64x</p>
集成芯片 Chiplet数	GPU+SRAM+HBM +Act Int. (47)	CPU+GPU+HBM +Active Int. (21)	CPU: 9 智能计算: 2+HBMx8	1+HBMx4