

Heterogenous Integration – Enabling systems beyond Moore's Law

Dr.-Ing. Andreas Middendorf
Strategic Business Development
Fraunhofer IZM, Berlin

Fraunhofer IZM - Crossing Frontiers in Microelectronics

Three facts about our institute

1

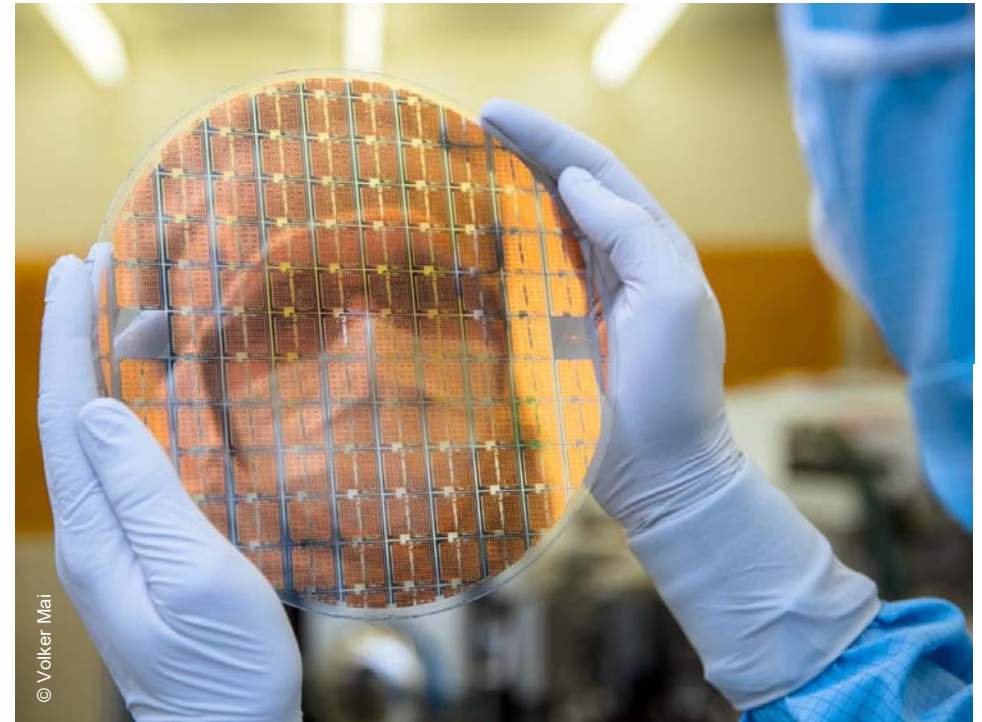
We are one of the world's leading institutes for applied research as well as the development and system integration of robust and reliable electronics.

2

We have over 30 years of experience with novel technological solutions developed in collaboration with partners from industry and academia.

3

We are the only fully integrated packaging institute covering everything from design, technology, reliability, and eco assessments.



Fraunhofer IZM at a Glance

30 years of experience



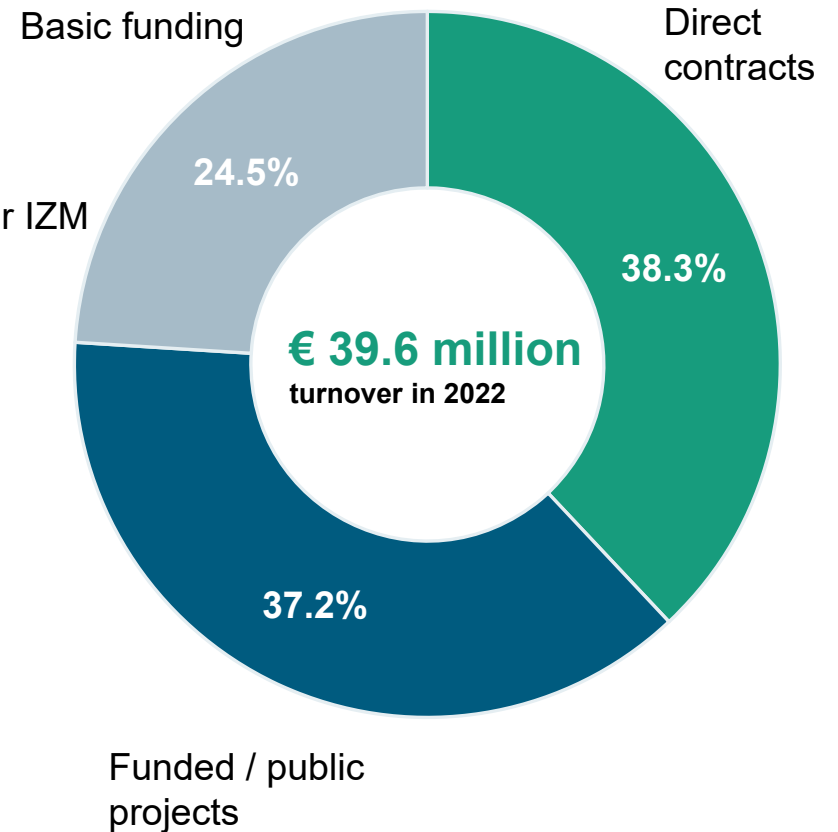
- 430 employees** (including students and trainees)
 - 121 interns, bachelor students, master students and student assistants have been supervised at Fraunhofer IZM
 - 8 trainees



- 5,374 m²** laboratory space
- 69** labs and measurement spaces



- Long-term contracts with
 - Technical University of Berlin
 - Technical University of Dresden
 - Brandenburg University of Technology



Fraunhofer IZM

A selection of our clients and partners

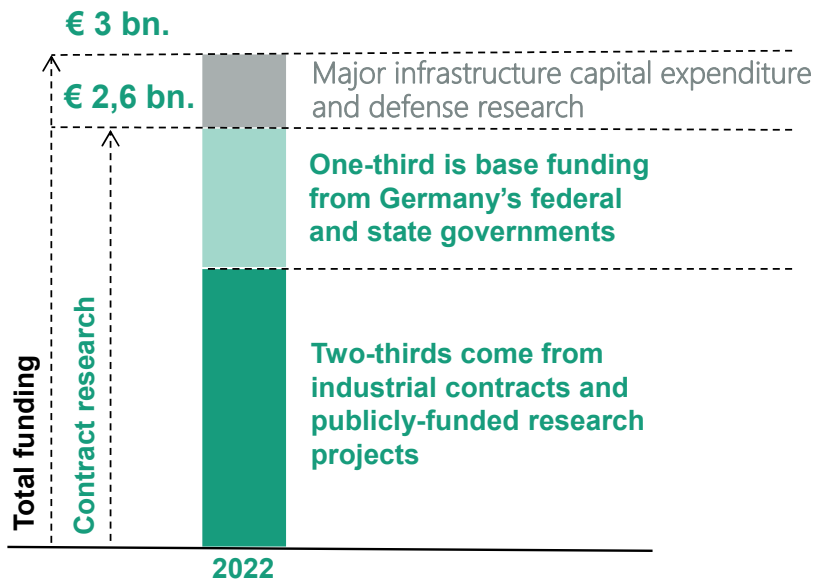


The Fraunhofer-Gesellschaft

Applied research focusing on key future-relevant technologies and the commercialization of findings in business and industry. A trailblazer and trendsetter in innovative developments.

 **>30,350 employees**

 **76 institutes and research institutes**



Research Fab Microelectronics Germany (FMD)



11 Institutes of the Fraunhofer-Gesellschaft

2 Institutes from Leibniz-Gemeinschaft



Approx. **4,500** employees with **2,800** scientists



€ 560 m. Budget



€ 200 m. industrial revenue
€ 235 m. public funding
€ 125 m. basic funding



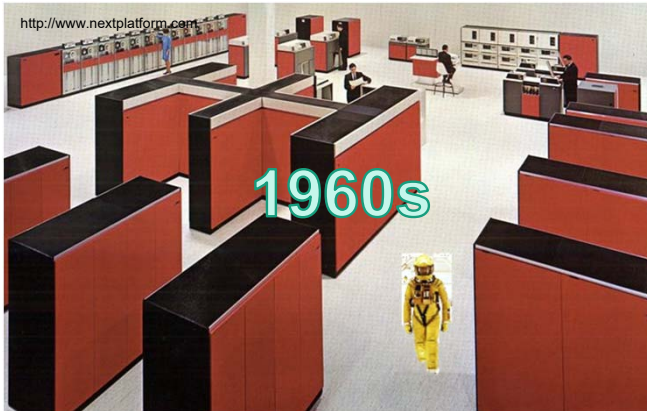
€ 2.2 bn. assets and investment



Design (down to 10/12 nm)
Wafer Processing:
 GaAs/InP (4"), SiC (6")
 Si, SiGe, GaN (8"), Si (12")
Advanced Packaging up to 12"; 600 mm panel
system integration, testing & characterization

High Performance Computing Cluster

Some Historical Facts...



IBM System/360
(1960s) 10 MIPS (ca. 1 MFLOPS)

1960s



70s

Heterogeneous
Integration for

HPC

1970s

IBM System/370 (1970s)
100 MIPS (ca. 10 MFLOPS)



Node Card

<http://ummr.altervista.org>

15 W

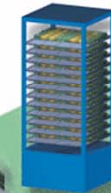


Cluster

Cabinet

2007

IBM Blue Gene/P (2007)
131072 CPUs: 360 TFLOPS



Cabinet

Node Card

Route cards

IO cards

Nodes

CPUs

4x2

10 GF/s

B* DDR

2 midplanes

1024 nodes

(2,048 CPUs)

(8x8x16)

2.9/5.7 TF/s

512 GiB* DDR

15-20 kW

System

64 cabinets

65,536 nodes

(131,072 CPUs)

(32x32x64)

180/360 TF/s

32 TiB*

1.2 MW

2,500 sq.ft.

MTBF 6.16 Days

* <http://physics.nist.gov/cuu/Units/binary.html>

High Performance Computing Cluster

Contemporary HPC Cluster with Nvidia A100



Ohio Supercomputer Center: Ascend HPC Cluster (2023)
48 AMD EPYC CPUs and 96 Nvidia A100 80GB Tensor Core GPUs: 2 PFLOPS

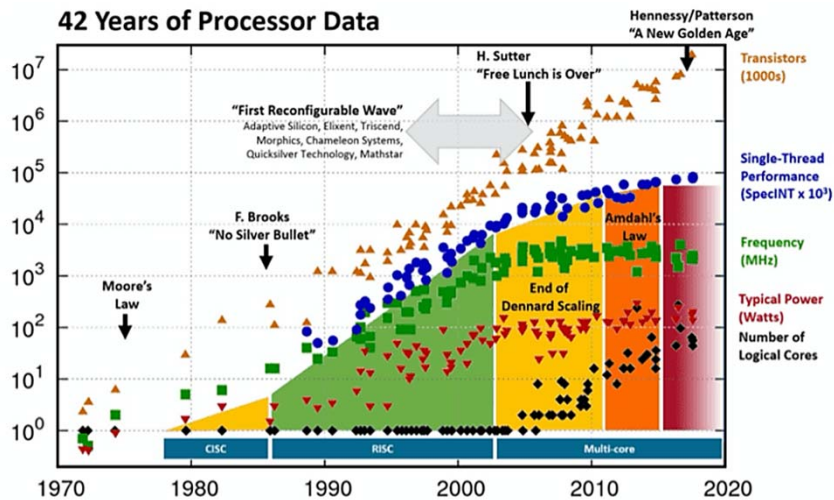


High Performance Computing

Drivers and Challenges

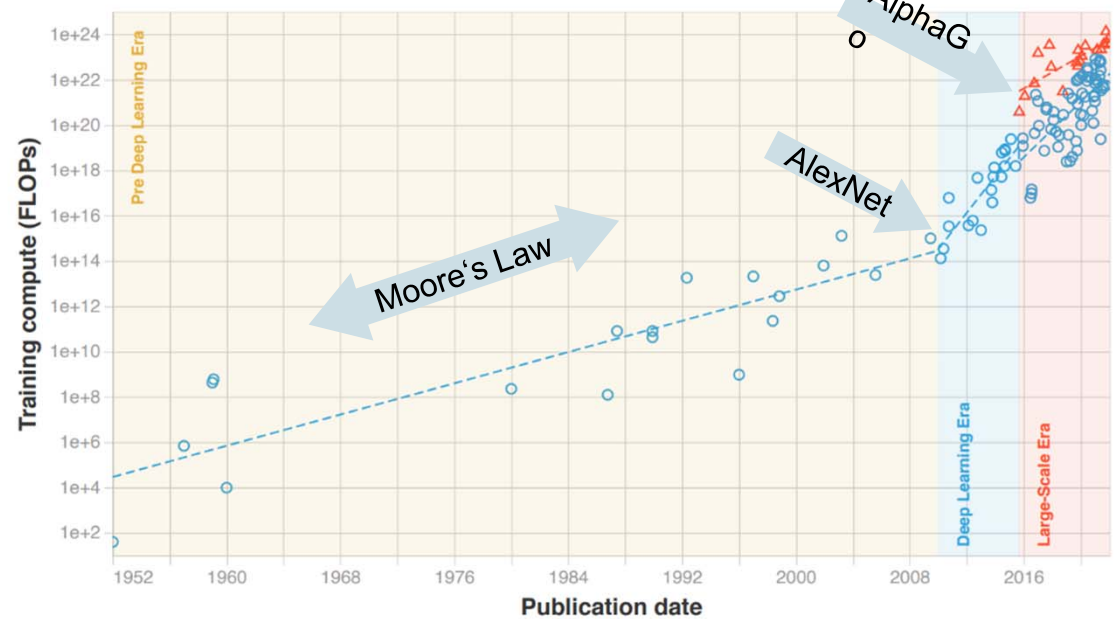
MEMORY WALL: Microprocessor speed increased faster than memory speed

POWER WALL: End of Dennard Scaling (= as transistors get smaller, their power density stays constant, < 65nm)



Hennessy and Patterson, Turing Lecture 2018, overlaid over "42 Years of Processors Data"
<https://www.karlsruhp.net/2018/02/42-years-of-microprocessor-trend-data/>; "First Wave" added by Les Wilson, Frank Schirrmeyer
 Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
 New plot and data collected for 2010-2017 by K. Rupp

Training compute (FLOPs) of milestone Machine Learning systems over time
 n = 121



AlexNet: Convolutional neural network (CNN) used for image recognition

AlphaGo: 1st computer program to defeat a human Go player

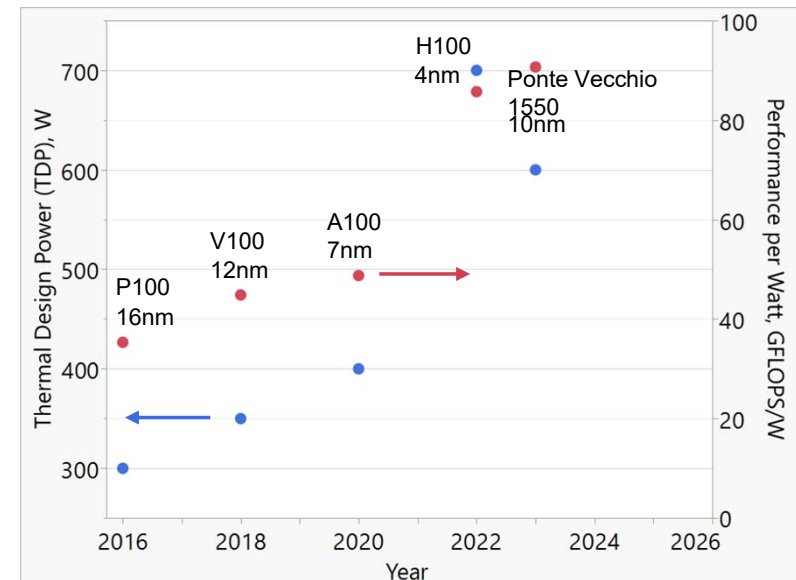
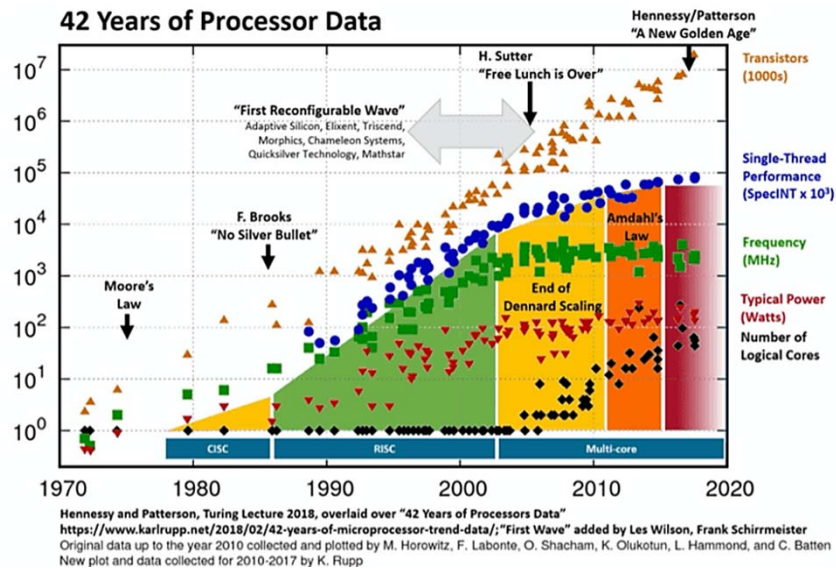
High Performance Computing

Drivers and Challenges

MEMORY WALL: Microprocessor speed increased faster than memory speed

POWER WALL: End of Dennard Scaling (= as transistors get smaller, their power density stays constant, < 65nm)

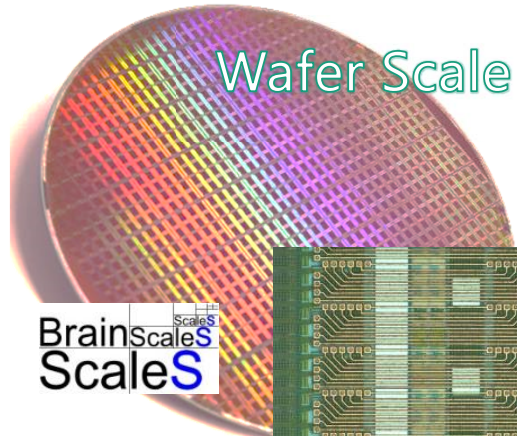
CPU	GPU	TPU
Several cores	Thousands of cores	Matrix based workload
Low latency	High data throughput	High latency
Serial processing	Massive parallel computing	High data throughput
Limited simultaneous operations	Limited multitasking	Suited for large batch sizes
Large memory capacity	Low memory	Complex neural network models



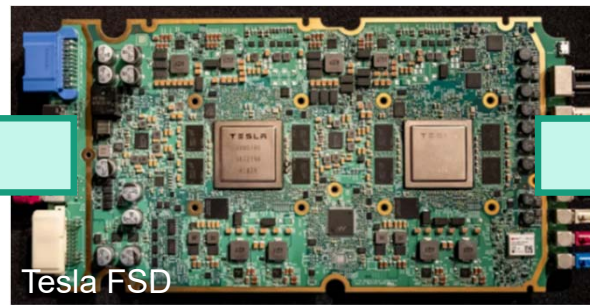
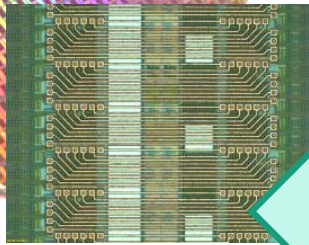
Demanding High-End Performance Packaging

SoC vs. Chiplet vs. Wafer Scale Engine

Wafer Scale Engine

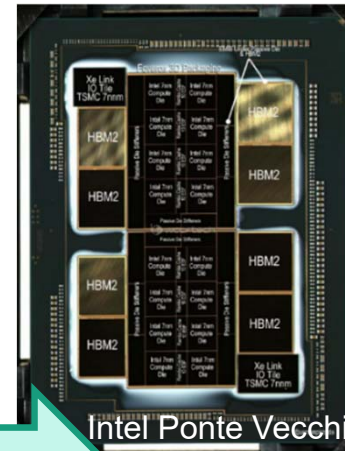


Brainscales
Scales
Scales



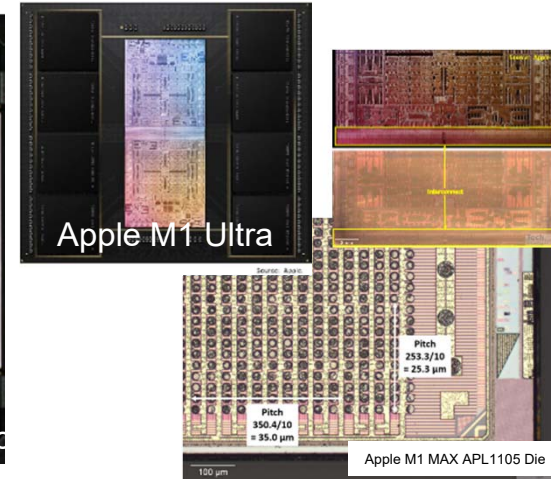
Tesla FSD

SoC



Intel Ponte Vecchio

Chiplets



Apple M1 Ultra

Apple M1 MAX APL1105 Die



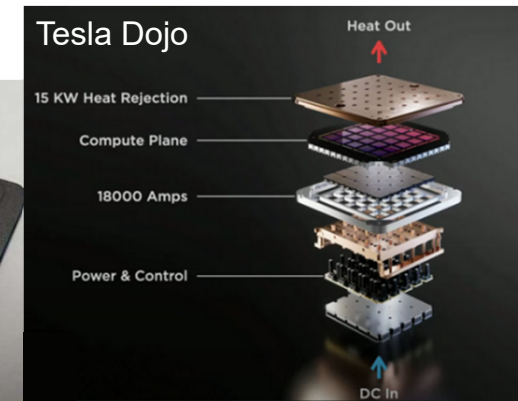
cerebras

Cerebras WSE-2
46,225mm² Silicon
2.6 Trillion transistors

Largest GPU
826mm² Silicon
54.2 Billion transistors



Ayar Labs

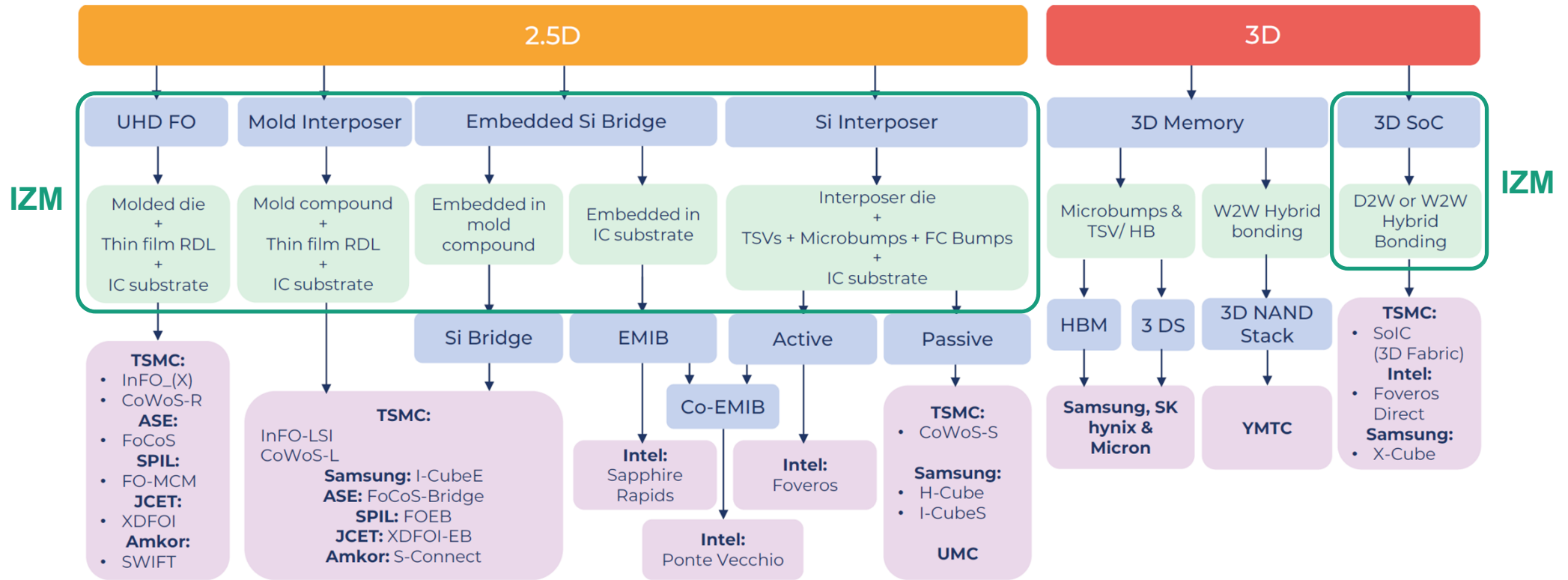


Tesla Dojo

- 15 KW Heat Rejection
- Compute Plane
- 18000 Amps
- Power & Control

Technology Overview and Manufacturers

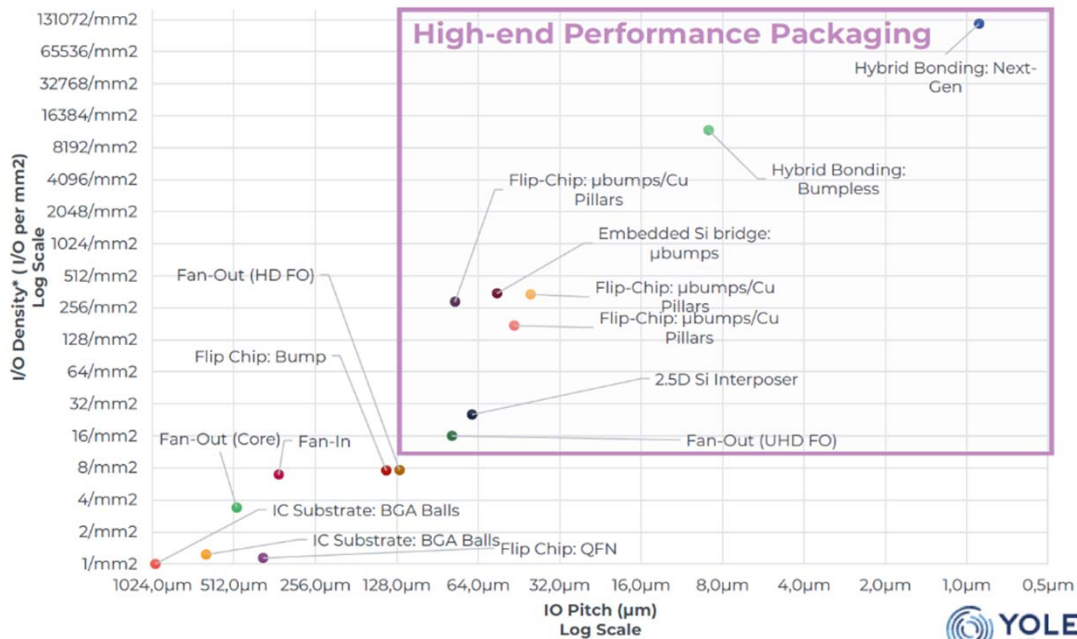
High-End Performance Packaging



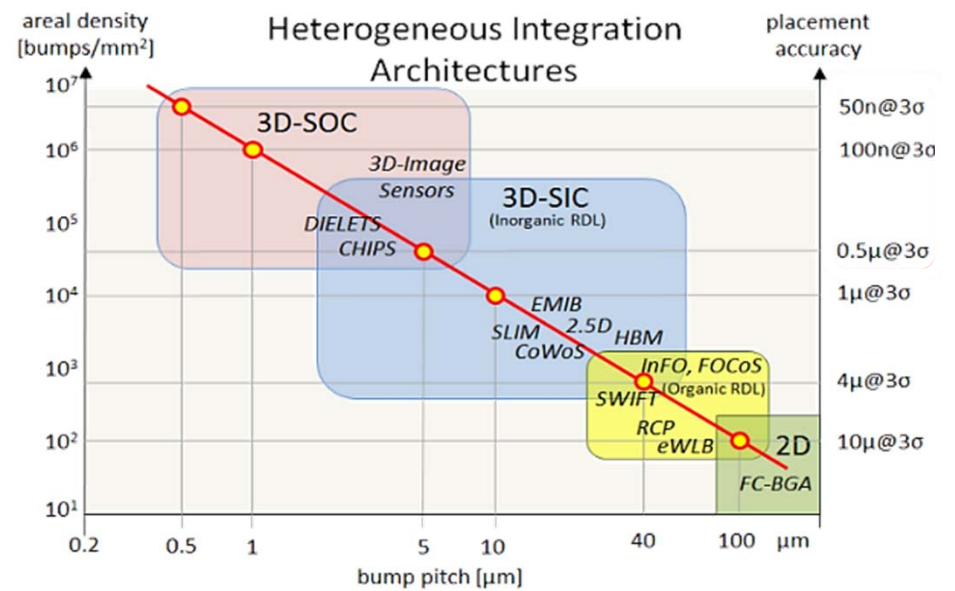
Advanced Packaging for High-Performance Computing

Major Technology Challenges for Chiplets Systems

IO Density vs. IO Pitch for Advanced Packaging (Log Scale)



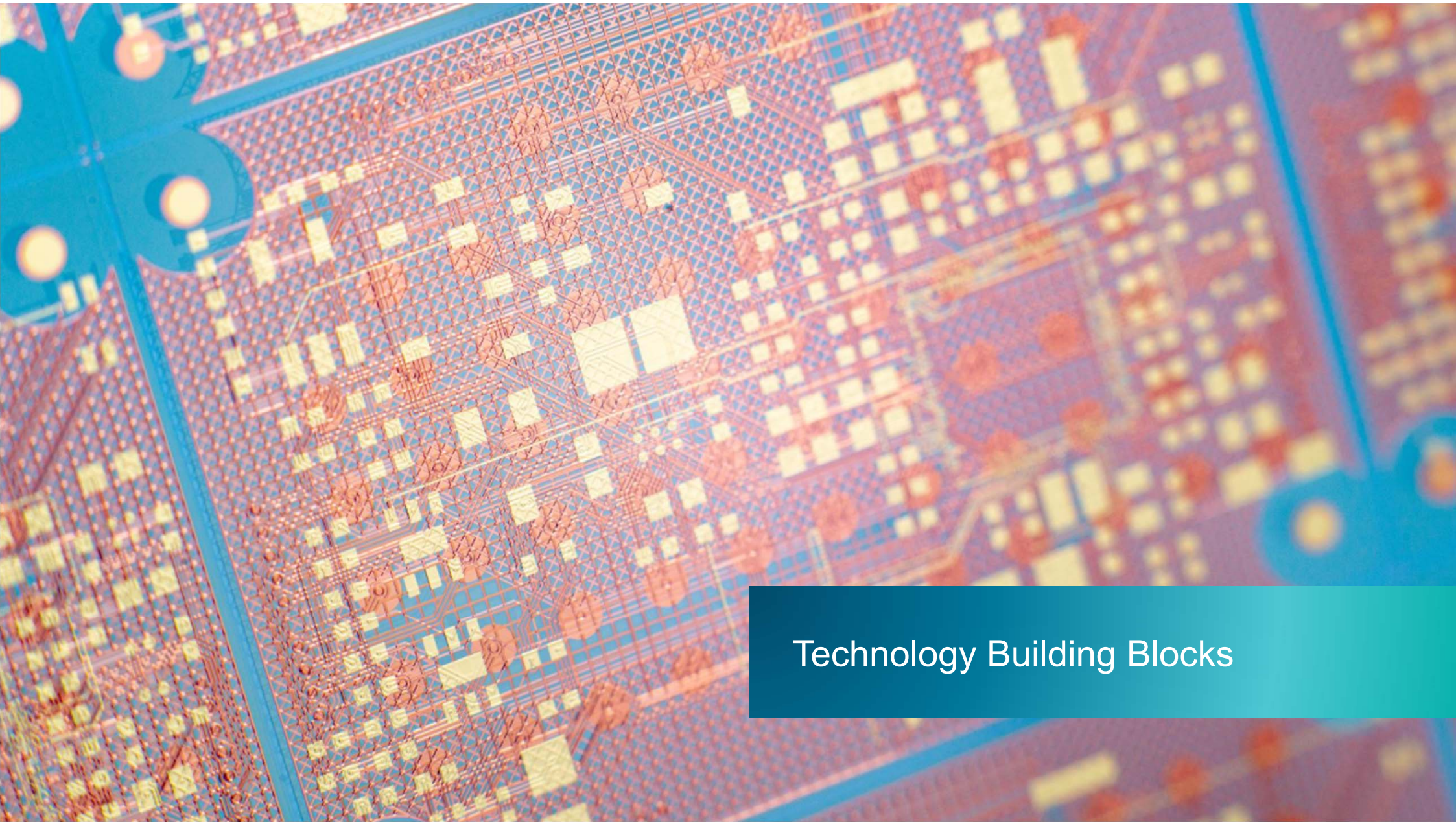
Smaller and smaller dimensions on a **single wafer**



<https://www.ectc.net/files/68/Behler%20Besl.pdf>

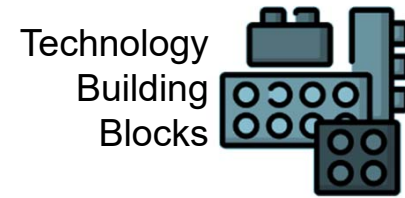


Integration of components **combined to a system**

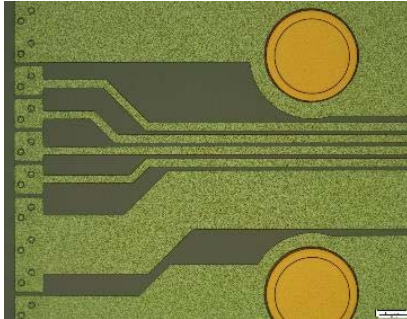


Technology Building Blocks

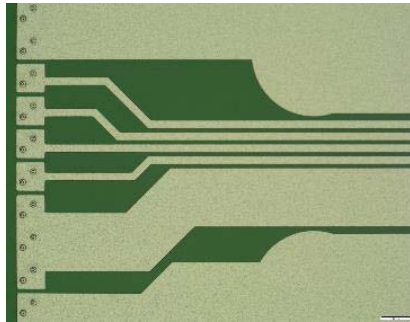
TSV Formation (TSV Frontside – Via Last)



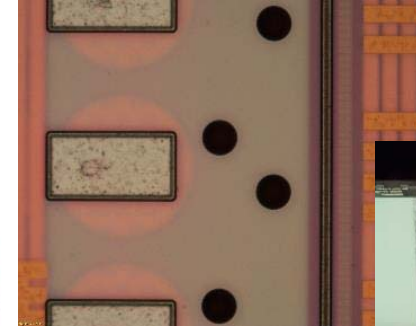
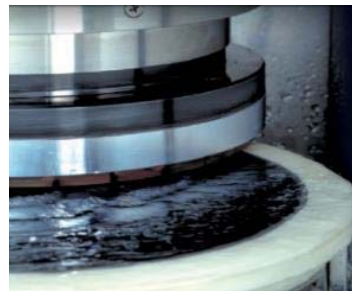
- 10. Backside passivation
- 11. Backside BGA Bump Pad Formation



9. Backside RDL



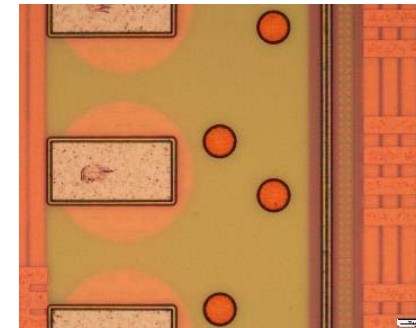
- 7. Backside TSV Reveal
- 8. Backside Passivation



1. TSV blind via etch and isolation



- 2. TSV filling
- 3. Front side RDL
- 4. Front Side Passivation / IO Formation

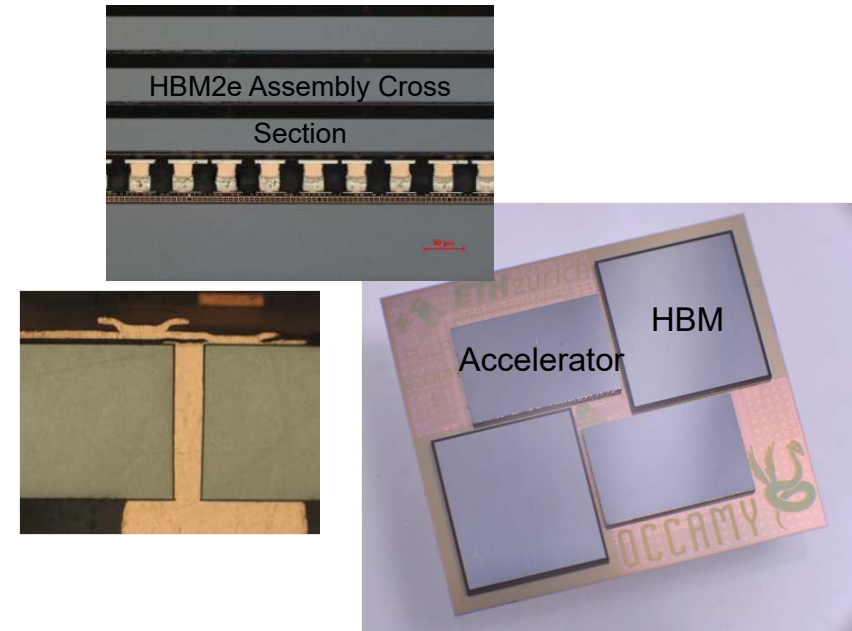


- 5. Temporary bonding
- 6. Si wafer thinning

Interposers

Silicon, Glass and Silicon Carbide

- **Si-interposers** w/ polymer based RDL and TSV → established technology, approved by industry
- **BEOL Si-interposers** back side finishing → established technology, approved by industry
- **Glass interposers** → technology path explored, various TGV and cavity options, interest from industry
- **SiC interposers** → technology setup is done but why SiC?
High thermal conductivity (>350 W/mK, on par with Cu!), high Young's modulus (>370 GPa), dielectric constant (6.5-10)

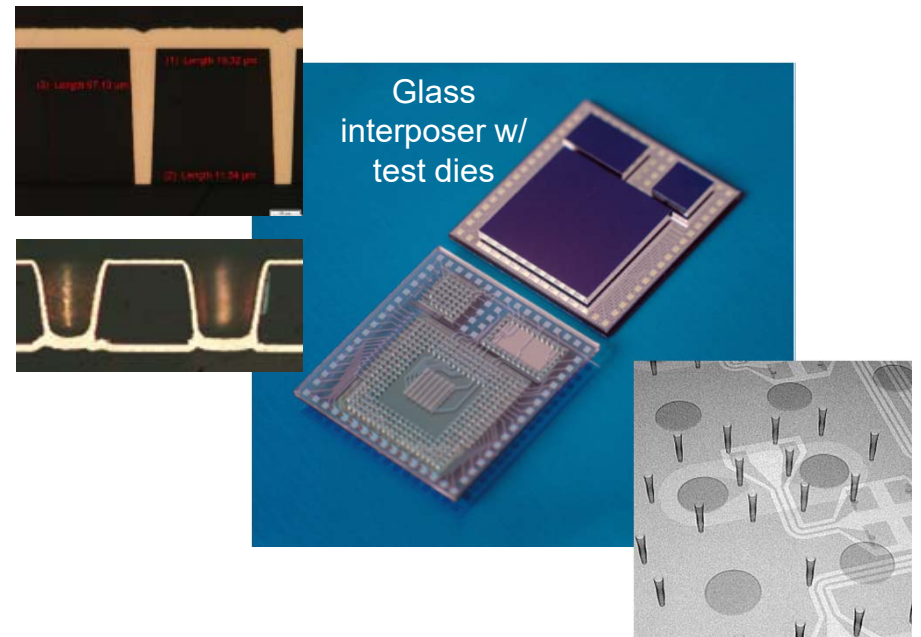


Chip / System Design: ETH Zürich
Interposer Finish / Assembly: Fraunhofer IZM

Interposers

Silicon, Glass and Silicon Carbide

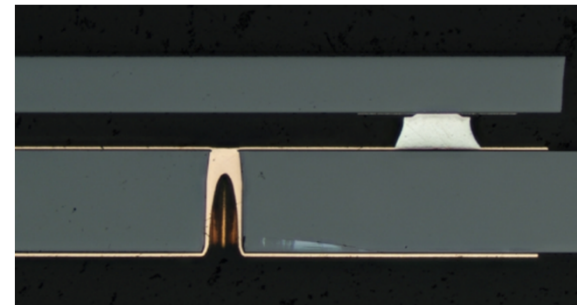
- **Si-interposers** w/ polymer based RDL and TSV → established technology, approved by industry
- **BEOL Si-interposers** back side finishing → established technology, approved by industry
- **Glass interposers** → technology path explored, various TGV and cavity options, interest from industry
- **SiC interposers** → technology setup is done but why SiC?
High thermal conductivity (>350 W/mK, on par with Cu!), high Young's modulus (>370 GPa), dielectric constant (6.5-10)



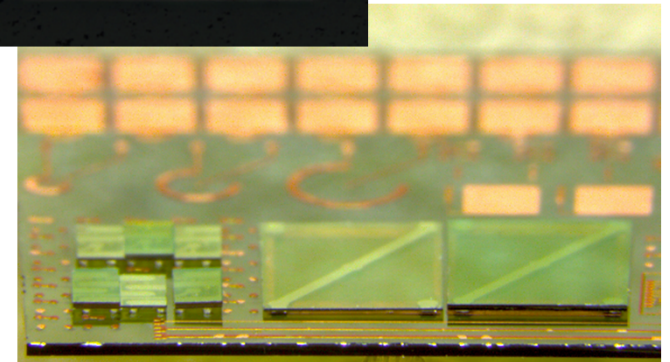
Interposers

Silicon, Glass and Silicon Carbide

- **Si-interposers** w/ polymer based RDL and TSV → established technology, approved by industry
- **BEOL Si-interposers** back side finishing → established technology, approved by industry
- **Glass interposers** → technology path explored, various TGV and cavity options, interest from industry
- **SiC interposers** → technology setup is done but why SiC?
High thermal conductivity (>350 W/mK, on par with Cu!), high Young's modulus (>370 GPa), dielectric constant (6.5-10)



SiC interposer
w/ test dies

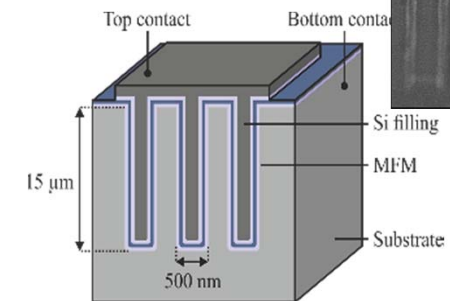
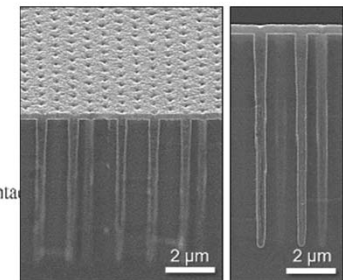
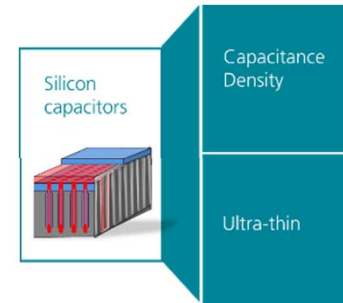
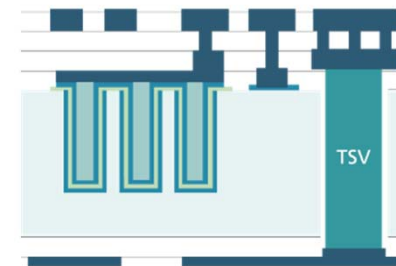


High-Density Si-Interposer

Passive/active

- Currently available
 - Flexible and high resolution wafer interconnect fabric by e-beam lithography for e.g. die-die connection
 - Integrated Silicon Capacitors:
 - Capacitance density > 500 nF/mm²
 - Capacitance stability over voltage and temperature
 - Low profile (< 50 μ m) and low temperature fabrication
- Under development:
 - Post bond Cu damascene metallization (starting 2023)
 - Security key integration through embedded NVM (starting 2024)
 - Coils and ferromagnetic materials for voltage regulators (2025)
 - W2W bonding with focus shift to (self-assembled) D2W bonding
 - Massive parallel transfer bonding for chiplet integration

Interposer integration

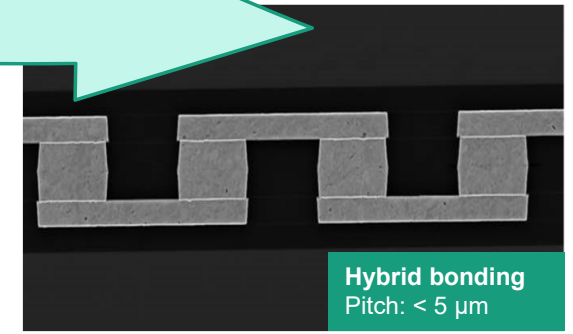
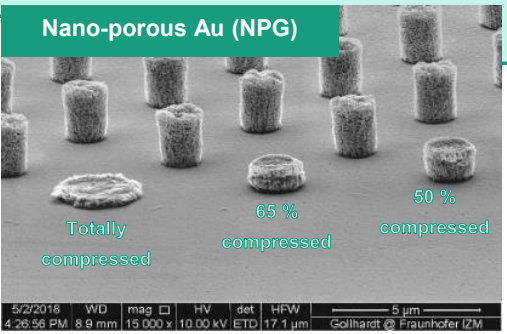
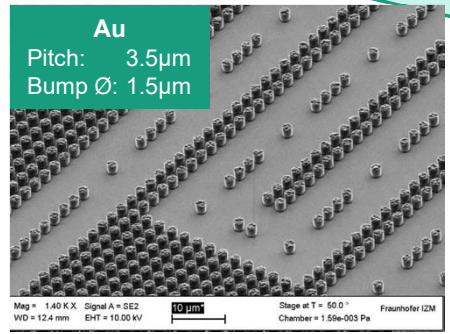
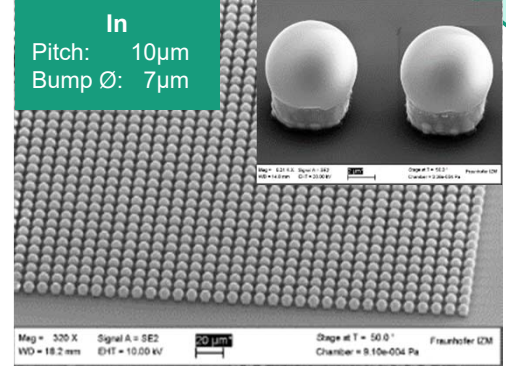
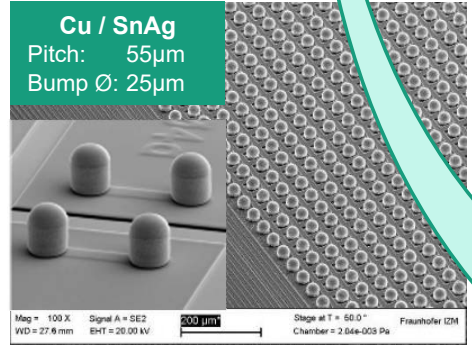


A cooperation of

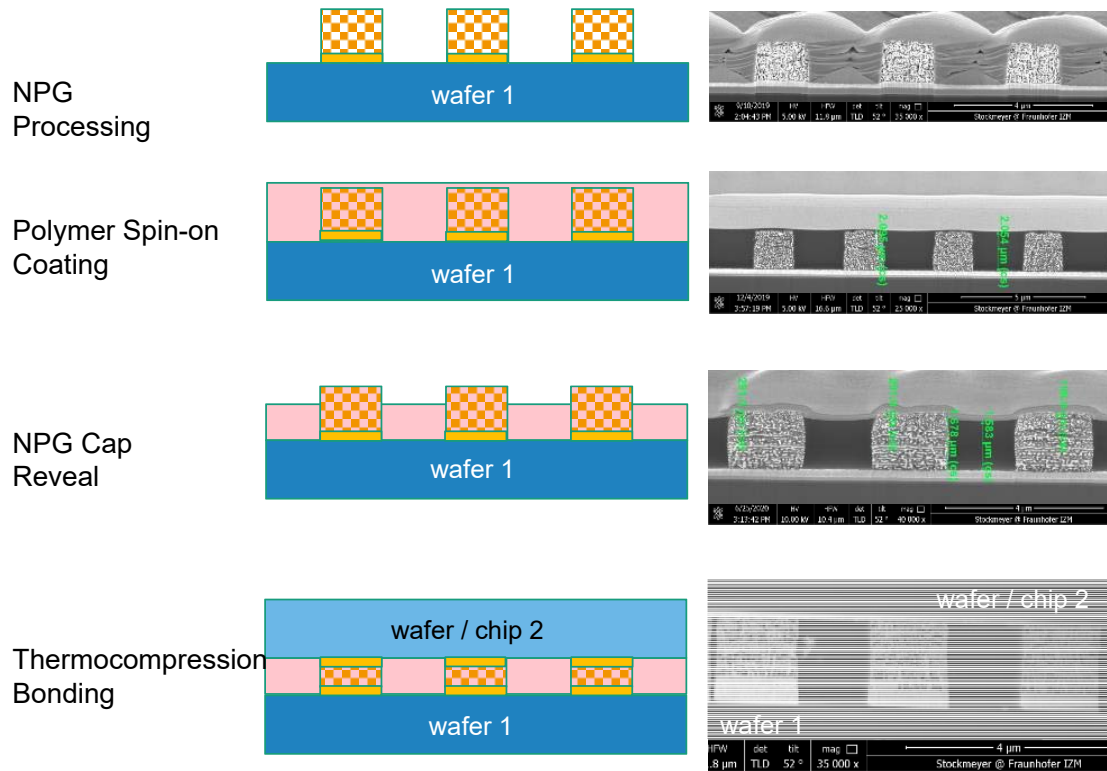
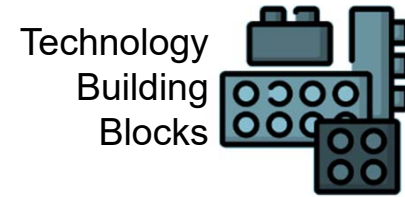


Pitch / Material Progression to Ultra-Fine Pitch Bumping and Assembly Technology

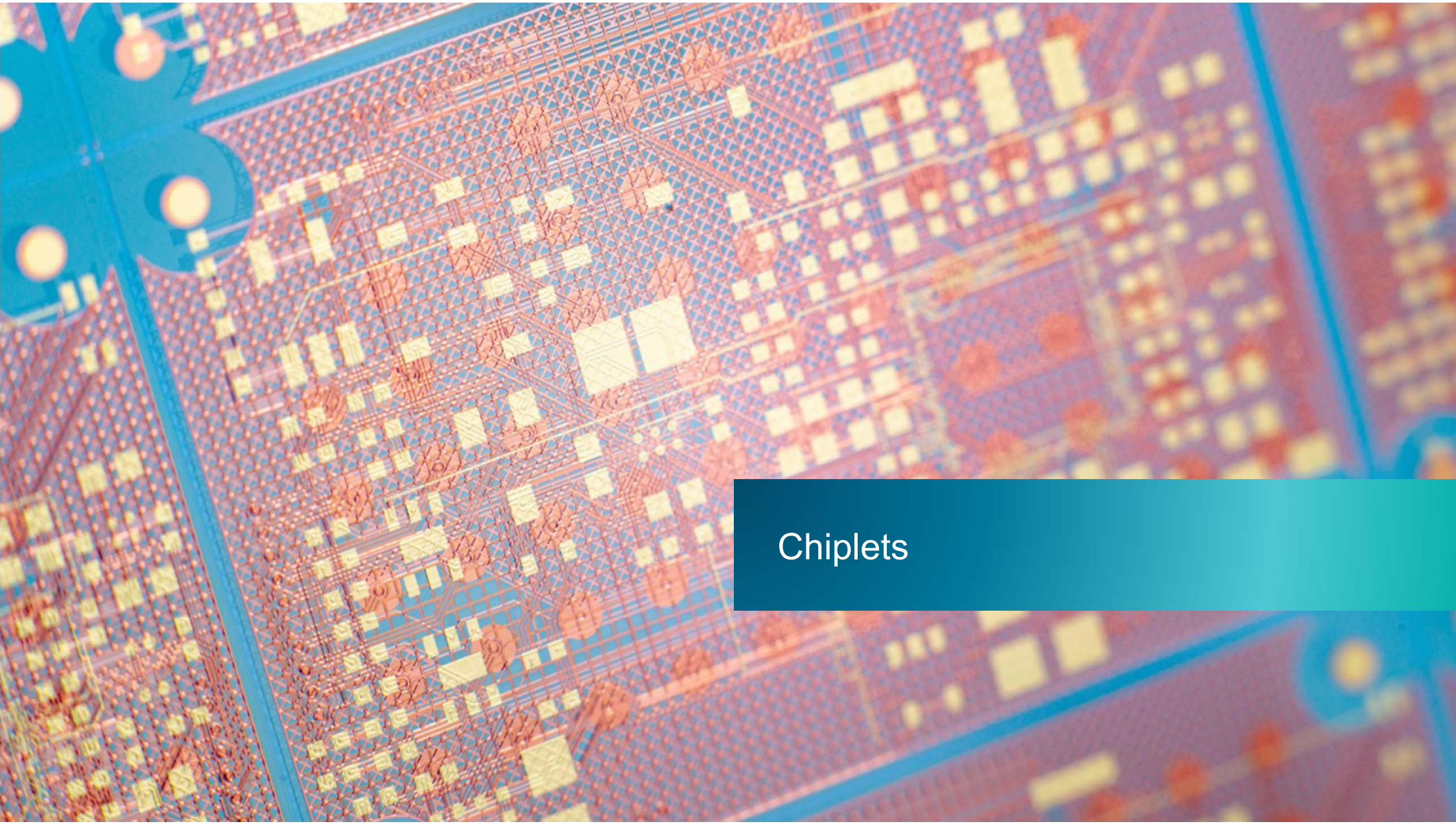
Interface	Bonding Method / Interconnect	Material
Solder	Reflow Soldering, Self-Alignment	SAC, SnAg, AuSn
Solder	TC Bonding: Cu-Pillar/Sn, PAUF	Cu-Pillar/Sn
IMC	Transient Liquid Phase Bonding („SLID“)	Cu/Sn, Au/Sn
IMC	Solder Diffusion Bonding (solid/solid)	Cu/Sn, Ni/In, Ag/In
Metal	Metal TC Bonding (solid/solid)	Cu-Cu, Au-Au, In-In
Metal	Nanoporous Gold (NPG)	NPG
Dielectric/Metal	Hybrid Bonding	Cu/SiO ₂ , NPG/Polymer
Bump Pitch, μm	1 3 5 7 10 20 30 50 70 100	



Spotlight: Assembly Progress to Ultra-Fine Pitch Polymer Hybrid Bonding Technology



D2W Bonding		
Spec	Hybrid SiO ₂ /Cu	Hybrid Polymer/NPG
Design Space	Very restrictive, one size only	Any size, mixed sizes
Topography	nm-scale (CMP), control BEoL topology	µm-scale, any wafer source possible
Step Height Pad/Bump	Dishing depth of nm-scale	20...40% of bump height
Dicing	Stealth/plasma dicing & surface protection	Standard dicing
Storage Time after Preparation	Several days	> 6 months
Clean Room Class	ISO 3 no particles	ISO 4-5 (partly) particle tolerant
Bonding	Pick & place	Thermode bonder
Throughput	Limited by alignment (if no collective bond possible)	



Chiplets



Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

19th April 1965

[...] It may prove to be more economical **to build large systems out of smaller functions, which are separately packaged and interconnected.** The availability of large functions, combined with functional design and construction, should allow the manufacturer of large systems to design and construct a considerable variety of equipment both rapidly and economically. [...]

Gordon E. Moore, 1965

Chiplet-Technology

Definition

A chiplet is an integrated circuit block specifically designed to work with other chiplets to form a larger more complex system that often makes use of reusable IP blocks

- A chiplet can be created by partitioning a die into functions that are more cost effectively fabricated (smaller die, higher yield, and less advanced nodes)
- A chiplet is a hard IP block
- Functions with other chiplets, so design must be co-optimized and silicon cannot be designed isolated
- Made possible by communication using chiplet interface (proprietary today)

Differs from SiP or MCM → New system design, not just a combination of different “off-the-shelf” chips

Chiplet is not the package, it's the design philosophy

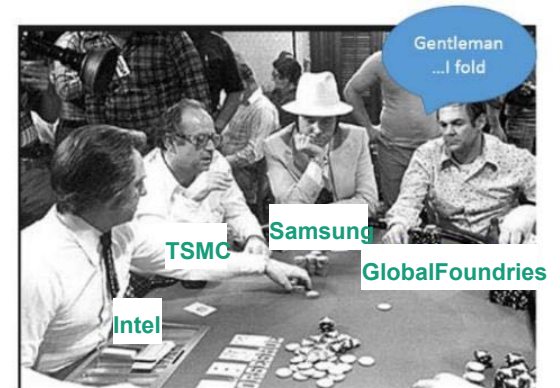
- Change from “silicon centric thinking” to “system-level planning” and “co-design of IC and package”
- The industry has to think about chip design in a new way
- Same impact as when the industry moved from a peripheral chip layout to area array!

Chiplet-Technology

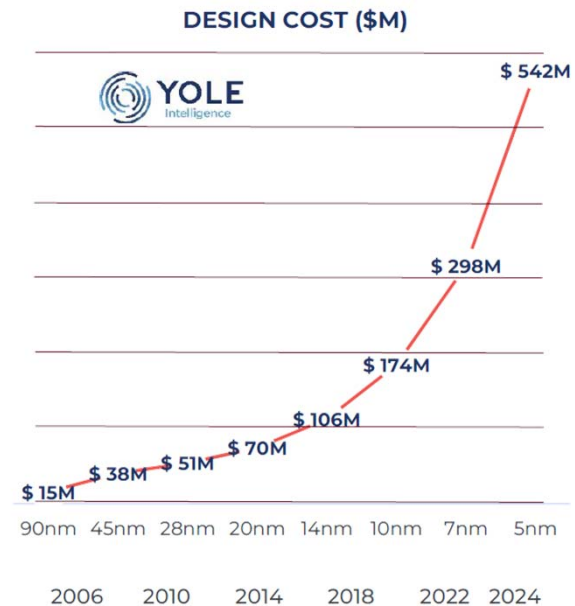
Motivation

Moore's Law and its classic node scalability has no economic advantages anymore:

- **High design cost and equipment invest** für next node generation (significantly caused by lithography)
- **High production cost vs. Yield** (Samsung has 60 % and TSMC at ca. 55% yield at 3 nm, according to KMIB News, 07/2023)



Phil Garrou, 2018 (7nm)

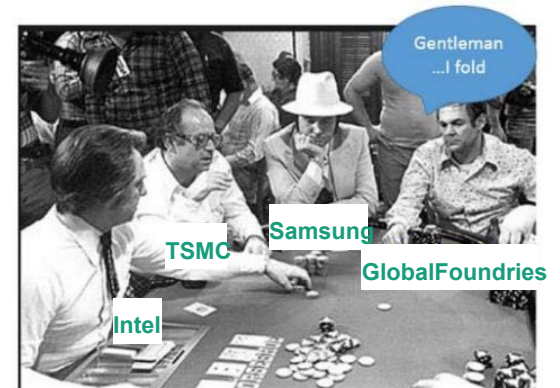


Chiplet-Technology

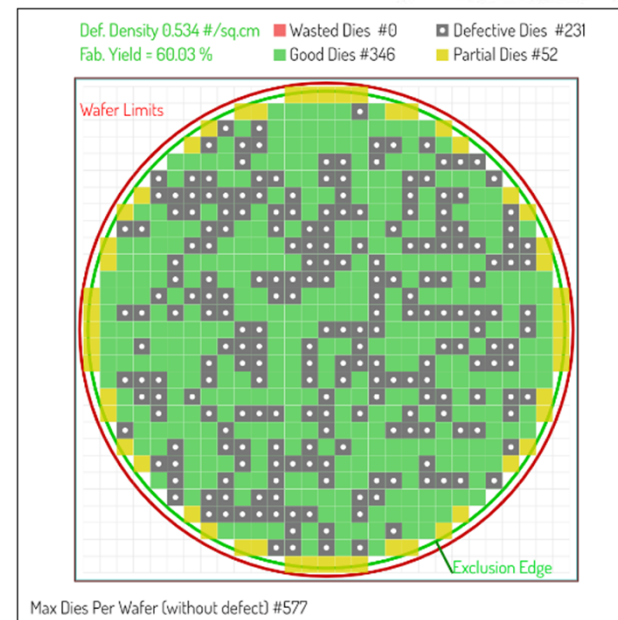
Motivation

Moore's Law and its classic node scalability has no economic advantages anymore:

- **High design cost and equipment invest** für next node generation (significantly caused by lithography)
- **High production cost vs. Yield** (Samsung has 60 % and TSMC at ca. 55% yield at 3 nm, according to KMIB News, 07/2023)



Phil Garrou, 2018 (7nm)



Chiplet-Technology

Motivation

Moore's Law and its classic node scalability has no economic advantages anymore:

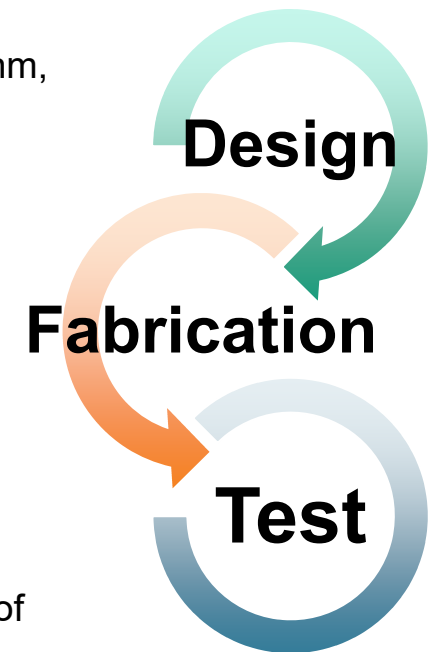
- **High design cost and equipment invest** für next node generation (significantly caused by lithography)
- **High production cost vs. Yield** (Samsung has 60 % and TSMC at ca. 55% yield at 3 nm, according to KMIB News, 07/2023)

Heterogeneous integration offers an economic turning point:

- 2.5D / 3D integration
- Embedded bridge
- Chiplet-disaggregation
- HD Fan-Out
- ...

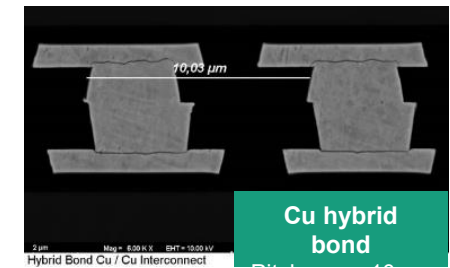
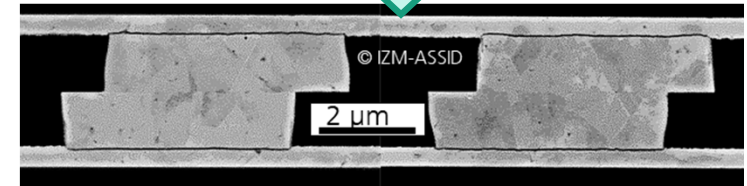
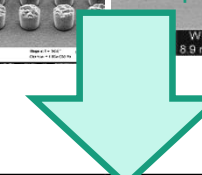
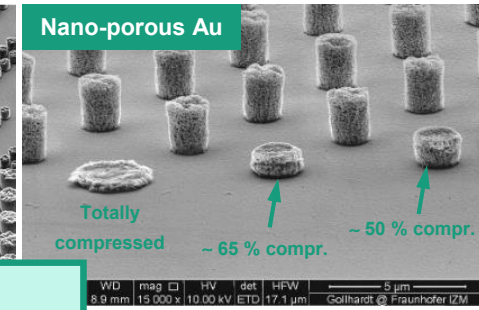
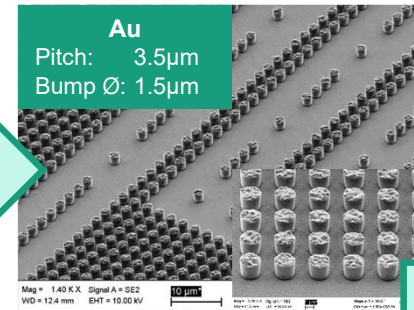
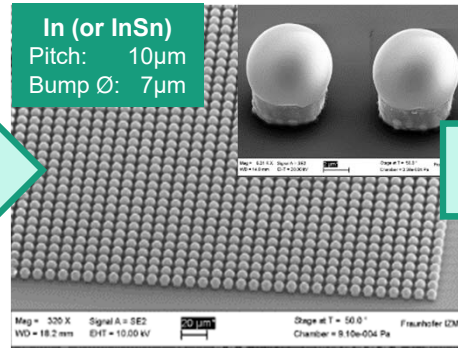
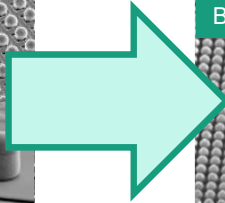
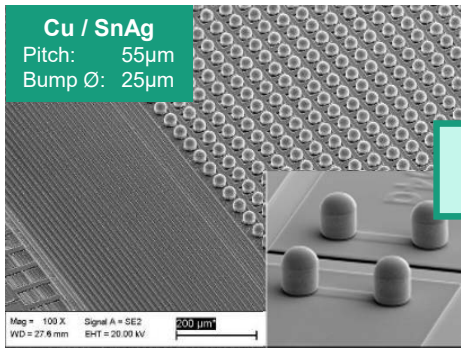
Challenges and at the same time the requirements would be:

- System Technology Co-Optimization (STCO) on system level is essential, combination of different materials, parts of classic assembly (D2W) probably are integrated in the fab processing
- Union of the triumvirate of design / fabrication / test



Wafer-Level Hetero-Integration

Ultra-Fine Pitch Bump/Bonding Technology



Interface	Bonding Method / Interconnect	Parameter
solder	Reflow Soldering, Self-alignment	no pressure
solder	TC Bonding: Cu-Pillar/Sn, PAUF	low pressure
IMC	TLPB („SLID“), Cu/Sn, Au/Sn	no or low pressure
IMC	Solder Diffusion Bonding (solid/solid), Cu/Sn, Ni/In, Ag/In	medium pressure
Met	Metal TC bonding (solid/solid), Cu-Cu, Au-Au, In-In	high pressure high temperature
Met	Nanoporous Gold (NPG) / Gold Nanosponge	low pressure low temperature
Ox/Met	Hybrid Bonding (oxide / metal)	very low pressure very low temperature

Process / material overview with respect to different bump pitches

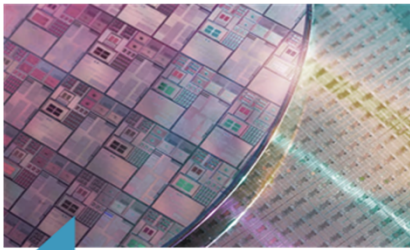
Wafer-and Panel-Level Hetero-Integration

Merging of Wafer- and PCB Technologies



Fraunhofer – Heterogeneous Integration

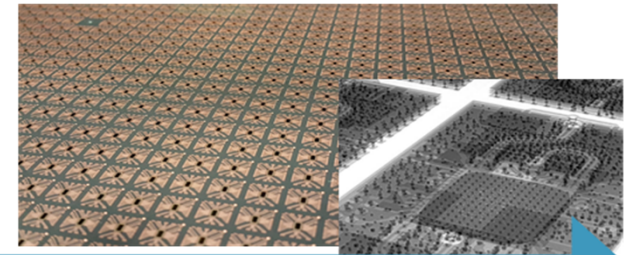
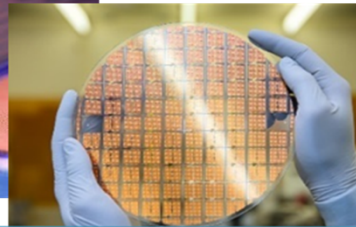
From Wafer Level System Integration to Panel Level System Integration



CMOS Feature Sizes
5 nm ... >100 nm



WLP Feature Sizes
0.75 μm ... >10 μm



PLP Feature Sizes
< 5 μm ... 100 μm

Wafer Level Packaging (WLP)

Technology: Based on thin film materials & equipment

Wafer size: 100 mm ... up to 300 mm

Input: CMOS - III/V - WBG wafers

Output: 2.5D/3D integrated systems or system components

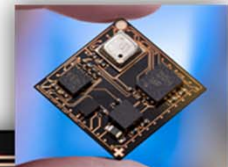
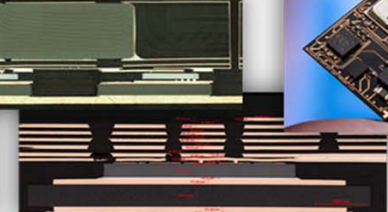
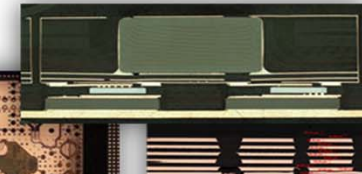
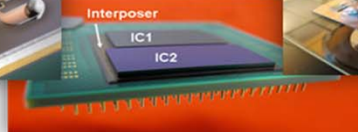
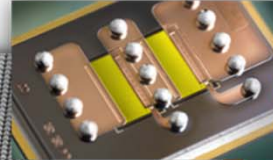
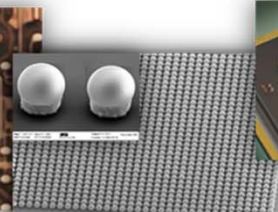
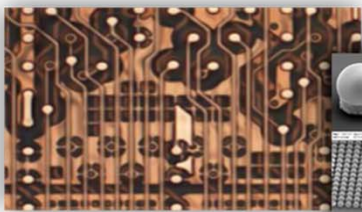
Panel Level Packaging (PLP)

Technology: Based on PCB materials & equipment

Panel size: up to 610 x 456 mm²

Input: CMOS / III/V / wide bandgap dies (w/ bumping)

Output: Packaged/embedded modules



A cooperation of



Process and Material Level

- Hybrid bonding/TSV: contact integrity, bonding strength, cyclic protrusion
- Smaller size & pitch (μm ...sub- μm): dielectric breakdown, migration, ESD
- New material interaction: Si/SiC/GaN...Metals like Cu, W, Al...barriers...dielectric

→ Model upgrades needed (size, process, time / age, temp. ... dependences)

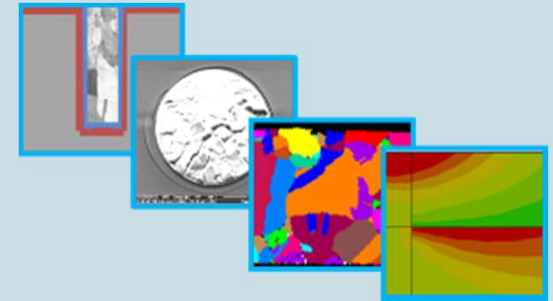
Assembly Level

- Thermal aspects: heat path solutions, thermo-mechanical interactions
- Design aspects: modularity, speed, cost, BIST & monitors, feedback loop
- Modeling aspects: digital twin, compact digital twin, health monitoring

Use Case / Application Level

- Flexibility & performance of QMS technology □ Useful in many different markets
- Much increased robustness & safety needed for markets like automotive

Analysis of Cu protrusion and TSV stress



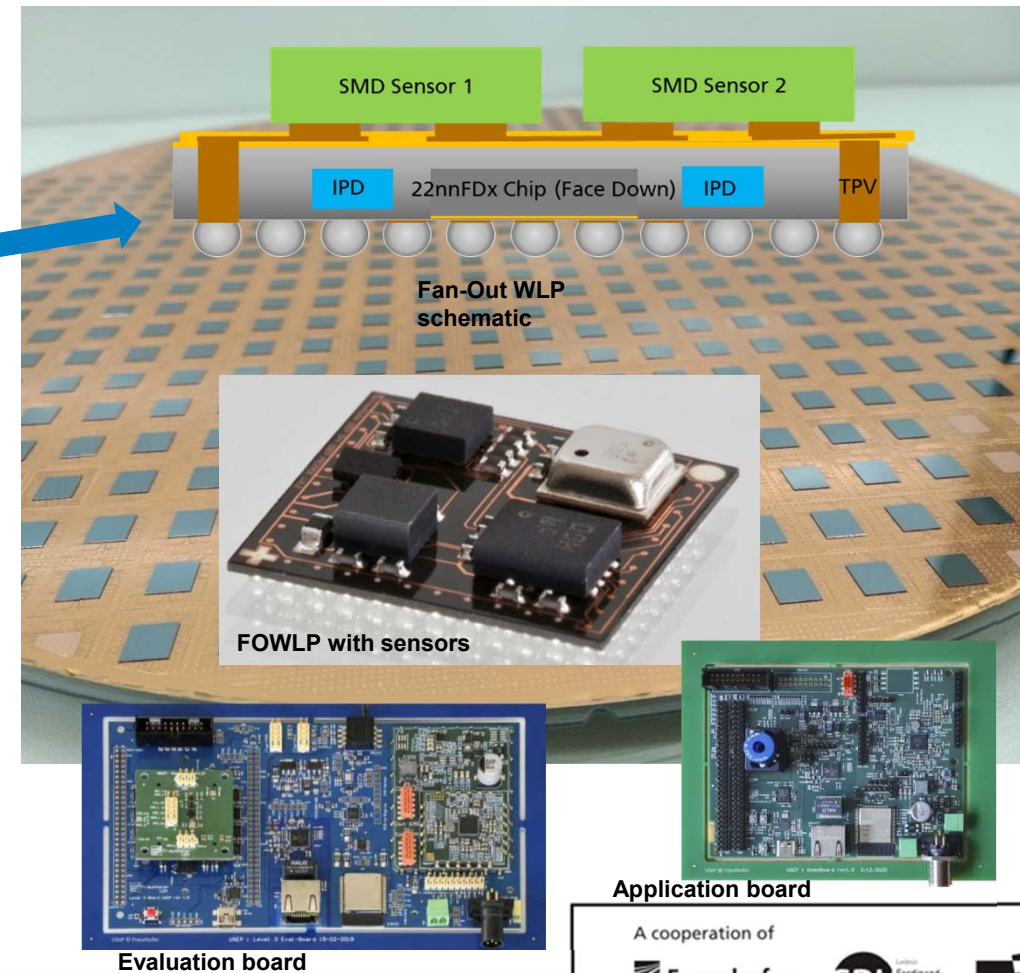
Hybrid system model (grey box model)



Application: Sensor platform for next generation electronics

Universal Sensor Platform (USEP)

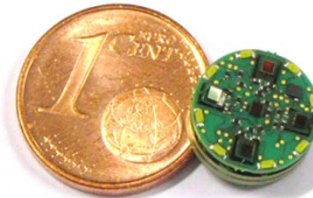
- Development of universal sensor platform (USEP), with which even smaller system providers can shoulder the growing development and production effort for next generation electronics
- Embedded chips
- Package-size: 10mm x 10mm x 0.2mm
- 4 layer thinfilm redistribution
- SMD-assembly of different sensors on top
- Assembly of complete sensor system on evaluation and application boards
- System level validation and functionality demonstration



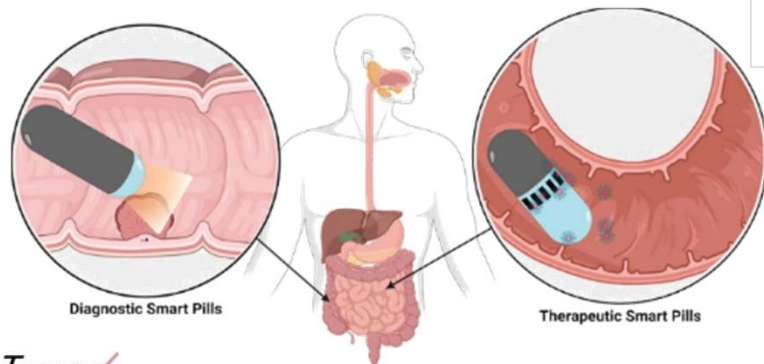
Application: Modular Capsule Endoscopy System

Project EndoTrace

- Use of different miniaturization techniques (e.g. Module stacking, embedding, semi-flex)
- Reduce number of images during the passage through the body (approx. 1/10)
- Onboard image capturing and storage, no external devices needed

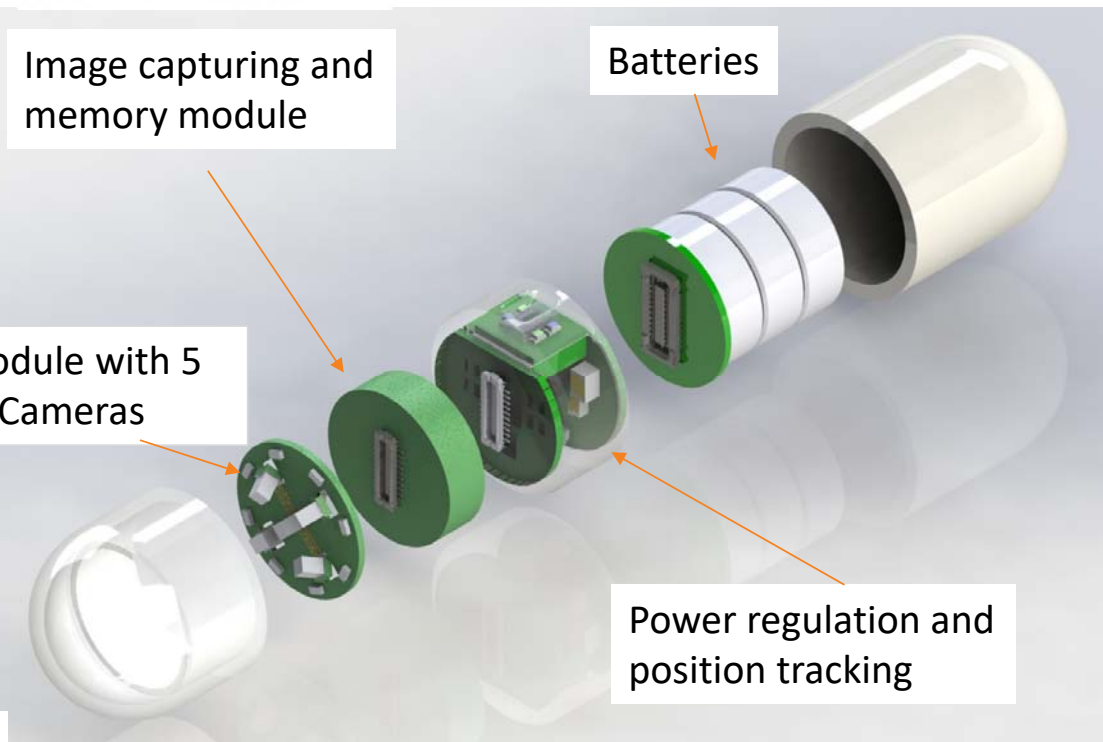


Memory / controller module with embedded SMDs



endoTrace

small intestine



GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

amin

o v e s c o
innovation in scope

A cooperation of

Fraunhofer
MIKROELEKTRONIK

FBH
Ferdinand
Braun
Institut

ihp

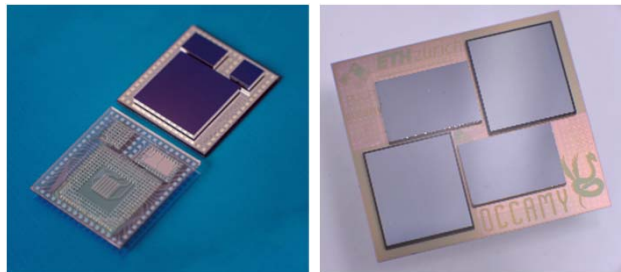
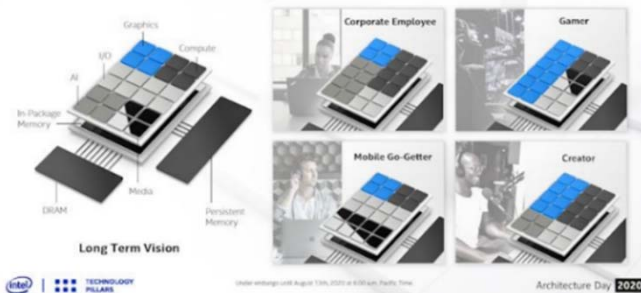


Summary

Key Elements for HPC Module Integration



Source: Intel.com



- Optimized system design and package technology for superior signal- und power integrity (active interposer, back side power supply, ...)
- Thermal management crucial for scalability of package
- Purpose built client (tailoring of chiplet mix for specific applications)
- Handling of thin wafers + TSV integration
- Thermo-mechanical stability → warpage / scalability
- Interconnect scalability → bump size/pitch and assembly technology (extreme placement accuracy specs!)

System Technology Co-Optimization (STCO)

Heterogeneous Integration

Conclusion

❖ Co-Design

- ❖ Chip design needs a close link to the package design to guarantee high performance and reliability
- ❖ Multiple domains with different scaling properties have to be taken into account
- ❖ Different design libraries are necessary for product development
- ❖ Thermal, mechanical and electrical analysis are key for high yield

❖ New Materials

- ❖ Mechanical, electrical and thermal interactions of different materials not yet used as default

❖ Cost

- ❖ Complex systems require new packages
 - ➔ Therefore new package platform like embedding have to be considered

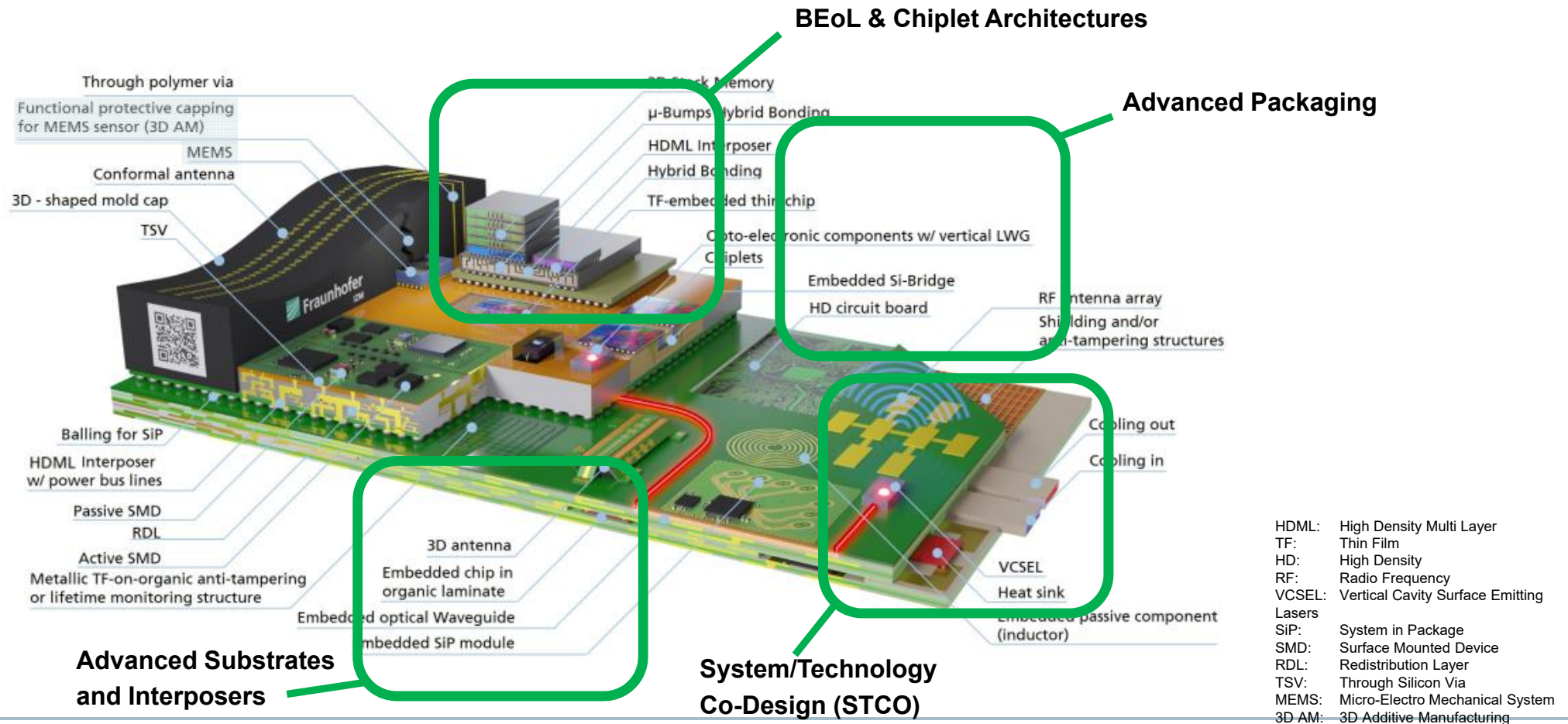
❖ Customer Requirements

- ❖ Reliability and application specific requirements
- ❖ High performance is mostly required in HI-applications
- ❖ Control of temperature requires advanced cooling concepts

❖ Test

- ❖ Application specific tests for complex packages (incl. mixed signal, media, etc.)
- ❖ Electrical, mechanical and thermal aspects are important

Complexity of Advanced Heterogenous System Integration



Contact

Dr. Andreas Middendorf
Strategic Business Development

Phone: +49 30 46403-135

E-Mail: andreas.middendorf@izm.fraunhofer.de

Fraunhofer IZM
Gustav-Meyer-Allee 25
13355 Berlin
www.izm.fraunhofer.de



Fraunhofer Institute for Reliability
and Microintegration IZM