Heterogeneous Integration: 2D - 3D Interconnects 2021 Chapter

Presenter : Ravi Mahajan (Intel Corporation)

Ravi Mahajan (Chair)	Raja Swaminathan	Subramanian Iyer (Co-Chair)	Kemal Aygun
Jan Vardaman	Zhiguo Qian	Kaladhar Radhakrishnan	Srikant Nekkanty
Rozalia Beica	Tom DeBonis	Dishit Parekh	Vidya Jayaram
Kanad Ghose	John Hunt	Adeel A. Bajwa	Paul Franzon
Debendra Mallik	Kaushik Mysore	Peter Ramm	Steffen Kroehnert
Venky Sundaram	Markus Wimplinger	Thom Gregorich	Sam Karikalan
Dave Armstrong	Chintan Buch	Vaibhav Agrawal	Michael Alfano
Gamal Refai-Ahmed	SB Park	Takafumi Fukushima	Ivy Qin

Ravi Mahajan

- Intel Fellow responsible for Assembly and Packaging Technology Pathfinding for future silicon nodes
- Interests:
  - Packaging Architectures, Thermal Management, Stress Analysis, Reliability, Assembly Process and Materials
  - Industry-Academia/Research Institute engagements

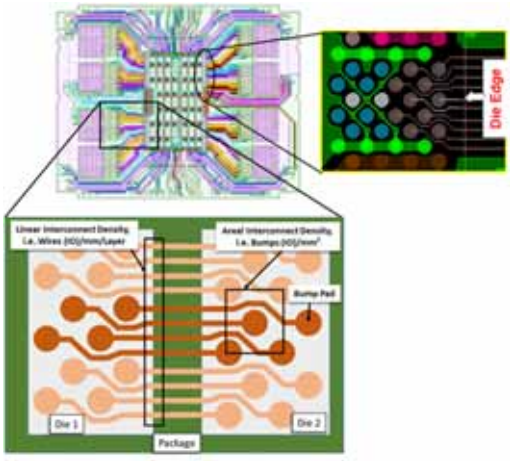
Subu Iyer

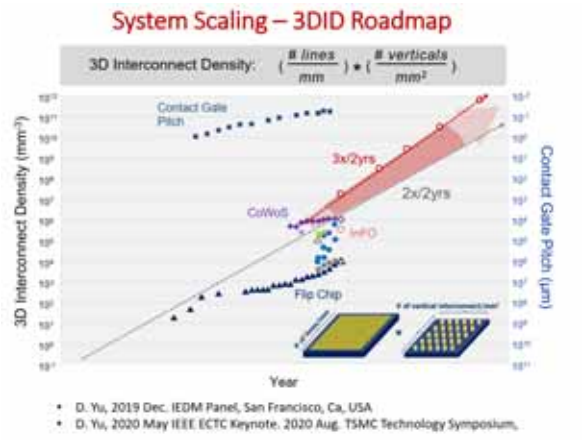
- Professor ECE and MSE at UCLA Charles P. Reames Endowed Chair, UCLA
- Interests:
  - New packaging constructs and their system applications in high performance computing and medical Engineering applications
  - In-memory compute and post-fab circuit modification

Chapter Objectives

- Define and proliferate a standardized nomenclature for package architectures covering and clearly identifying, 2D and 3D and hybrid packaging constructs
  - Comprehend all HI announcements including Wafer-scale constructions.
- Define and disseminate key metrics driving the evolution of the physical interconnects in these architectures
  - The chapter lists their current values and projections for the next generations
- Chapter is organized into 4 primary areas:
  - Converged Nomenclature Framework for 2D , 3D and hybrid Architectures
  - Key Metrics:
    - Design Attributes
    - Electrical Attributes including Signaling and Power Delivery
  - Challenges
  - Discussion

Physical Metrics





- Identical primary physical density metrics from three separate sources (Intel, TSMC and IMEC (https://www.imec-int.com/en/articles/view-3d-technology-landscape)

Interconnects

- **Die-Die Interconnects:** Interconnects between stacked die that enable vertical interconnects between multiple die in a 3-D stack *Covered in this chapter*
- **On-package Die-Die Interconnects** i.e., 2D and Enhanced 2D Interconnects: Interconnects between die within the package that enable lateral connections *Covered in this chapter*
- **Die-to-Package (*or to substrate*) Interconnects:** Interconnects between the die and the package typically known as the first level interconnect (FLI) *Covered in this chapter*
- **Within Package Interconnects:** Interconnects within the package that enable lateral connections *Covered in the Substrate Section of the Single Chip and Multi-Chip Chapter*
- **Package to Board Interconnects:** Interconnects between the package and the next level, which is typically the motherboard, referred to as the second level interconnect (SLI) *Covered in this chapter*
- **Package on Package Interconnects** *Covered in this chapter*

Interconnect Scaling to Enable Bandwidth Scaling

- Peripheral Interconnects for 2D Architectures with Aggressive pitch Scaling

Generations		1	2	3	4	5
Raw Bandwidth Density (GBps/mm)		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)	55	40	30	20	10
Signaling Speed (Gbps)		2	3	4	5.33	8

Calibrating these numbers w/ UCIe, HBM and BOW etc., will help define the generation length

- Area Interconnects for 3D Architectures

Generations		1	2	3	4	5
Raw Bandwidth Density (Gbps/mm^2)		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)	40	30	20	15	10
	IO/mm^2	625	1111	2500	4444	10000
Signaling Speed (Gbps)		1.6	1.8	1.6	1.8	1.6

- In synchronous signaling, the data rate , BW etc., is limited to even multiples of chip clock rate
- Should latency be added as a metric ?

Interconnect Scaling to Enable Bandwidth Scaling

- Peripheral Interconnects for 2D Architectures with Aggressive pitch Scaling

Generations		1	2	3	4	5
Raw Bandwidth Density (GBps/mm)		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)	55	40	30	20	10
	IO/mm	500	667	1000	1500	2000
	IO/mm^2	331	625	1111	2500	10000
Signaling Speed (Gbps)		2	3	4	5.33	8

- Area Interconnects for 3D Architectures

Generations		1	2	3	4	5
Raw Bandwidth Density (Gbps/mm^2)		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)	40	30	20	15	10
	IO/mm^2	625	1111	2500	4444	10000
Signaling Speed (Gbps)		1.6	1.8	1.6	1.8	1.6

Signal Integrity Attributes

Generation Number →		1	2	3	4	5
Linear Bandwidth Density (GBps/mm)		125	250	500	1000	2000
Channel Performance	Channel Length (mm)	<2	<1.7	<1.4	<1.1	<0.8
	Bump-to-Bump Channel RC (ps)	<10	<10	<10	<10	<10

Channel Signaling Characteristics for 2D and Enhanced-2D Architectures (RC Dominated)

Generation Number →	1	2	3	4	5
Areal Bandwidth Density (GBps/mm^2)	125	250	500	1000	2000
Bump Capacitance (fF)	<30	<22	<15	<10	<7

Channel Signaling Characteristics for 3D Architectures (Capacitance Dominated)

Power Delivery Attributes: Area Interconnects for 2D and 3D Architectures

Generation Number →		1	2	3	4	5
Core Power Density (W/mm²)		5	8	12	18	25
	On-die MIM Capacitance Density (nF/mm²)	50	100	200	400	750
	Silicon Trench Capacitance Density (nF/mm²)	300	500	1000	1750	3000
	VR Power Density (W/mm²)	0.4	0.6	1	1.5	2.5
	Ceramic Cap Density (µF/mm²)	40	50	70	100	150
	Bump Current Carrying Capability (A/mm²)	6	9	14	20	30

Power delivery Attributes for 2D, Enhanced-2D and 3D Architectures. It should be noted that power delivery attributes are agnostic to the architecture (*Errata: Ceramic cap density values in the 2020 version of this chapter had an error i.e., an extra "0" was included in Gens 4 and 5*)

- Current carrying capacity is limited by the capture level connection current crowding – this is not directly captured by the average, and we may need to figure out how to account for this effect

Power Delivery Attributes: Area Interconnects for 2D and 3D Architectures

Generation Number →		1	2	3	4	5
Core Power Density (W/mm²)		5	8	12	18	25
	On-die MIM Capacitance Density (nF/mm²)	50	100	200	400	750
	Silicon Trench Capacitance Density (nF/mm²)	300	500	1000	1750	3000
	VR Power Density (W/mm²)	0.4	0.6	1	1.5	2.5
	Ceramic Cap Density (µF/mm²)	40	50	70	100	150
	Bump Current Carrying Capability (A/mm²)	6	9	14	20	30

Power delivery Attributes for 2D, Enhanced-2D and 3D Architectures. It should be noted that power delivery attributes are agnostic to the architecture (*Errata: Ceramic cap density values in the 2020 version of this chapter had an error i.e., an extra "0" was included in Gens 4 and 5*)

Review of Different Packaging Architectures

Architecture Type (2D/2D Enhanced/3D)	Packaging Materials	Tightest Pitch	Typical Interconnection Process	Typical Equipment	Typical Applications	Key Concerns (Partial List)	Challenges (Partial List)	Advantages (Partial List)
2D [30]	Wire-Bonding	25µm	Wire-bonding	Wire-Bonder	Automotive, LCD drivers, Sensors, ASICs, Controllers	- Oxidation of Cu bonds - Wire-bond lift-off due to CTE mismatch	- Limitations with bonding temperatures - Corrosion / bond integrity	- Low cost-of-ownership - Flexible process - Easy to test / re-work wire-bonds
2D & 2D Enhanced [31]	Micro-bumps, C4s, TSVs, and passive Si Interposer	20 µm	TCB / mass-reflow processes	Thermo-compression bonding tools	CPUs, GPUs, FPGA, Network servers, Gaming Console Servers	- Warpage issues due to large package size	- Interposer testing - Handling thinned wafers	- Packaging enables high performance - Multifunction heterogeneous integration e.g. ASIC + HBM
3D Die-to-Die, Die-to-Wafer [32]	Micro-bumps, TSVs, Cu-Cu Bonding	5 µm	F2F or F2B Direct Cu-Cu bonding interconnection	Custom bonding equipment	AI, HPC, AR/VR, 5G applications	- Ensuring known good dies (KGD) - Test coverage limited during wafer probing	- High cost of ownership - Misalignment / FM particles during die placement	- Pitch Scalability
3D Wafer-to-Wafer [33]	Solder Bumps, Cu-Cu Bonding	0.9 µm	F2F or F2B Direct Cu-Cu bonding interconnection	Custom bonding equipment	AI, HPC, AR/VR, 5G applications	- Ensuring known good dies/stacks - Test coverage limited during wafer probing	High cost of ownership - Misalignment / FM particles during die placement	High 3D interconnect density with ultra-low bonding latency

Examples of applications from literature as a function of the different packaging architecture and process/material attributes (*Reference envelope values. Values listed in Table in this chapter represent the broader mainstream envelopes*)

Challenges

- Impact to Signal Integrity with shrinking features
- Novel Assembly for both Solder and Non-Solder Interconnects
- Fine Pitch Sort and Test
- Thermal Management
- Efficient Power Delivery Networks
- Design-Process Co-Design
- Equipment Readiness

2021 Chapter Changes

- Improve Images (No Changes)
- Continue Cross-referencing with other chapters (On-going)
- Re-calibrate on the projections (too aggressive? Too conservative?) – No Changes in numbers but captions and numbers have been clarified)
- Industry announcements (Included)
- Include work done at Universities (Included)
- Process & Materials attributes added

Metrology Focus Areas*

- Characterization Metrologies for Model development and Validation
  - Warpage as a function of temperature
  - Properties (Mechanical, Thermal, Electrical, Interface)
- In-Situ (Fast & Accurate) Process Metrologies
  - Defects (Voids, Cracks, Delamination, Residue)
  - Dimensional metrologies
- Reliability
  - Defects (Voids, Cracks, Delamination, Residue)
  - FA/FI Techniques/Enablers

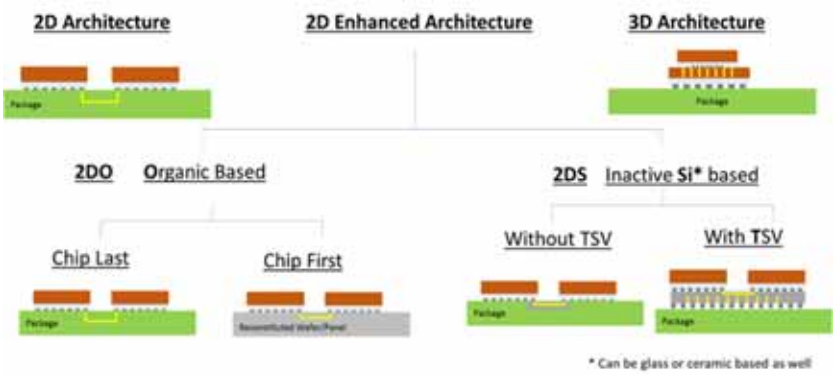
* Partial List. See refs (e.g., SRC Needs (https://www.src.org/program/grc/research-needs/) for a more detailed discussion

Cross-TWG Collaboration Opportunities

- Substrates and Power Delivery are key areas of Interaction
- This chapter (with appropriate calibration) should be the basis of driving on-package signaling definition

Back-up

Nomenclature Framework



Schematic describing the Converged Nomenclature Framework for 2D & 3D Architectures. Note that the difference between the 2D and 2DO Chip Last schematic in this figure is in the interconnect density in the die-die links. The latter has increased interconnect density enabled by finer lines and spaces along with reduced sized vias and via pads. In the 2DS images the grey color is used for silicon (or glass)