HETEROGENEOUS
INTEGRATION ROADMAP

2019 Edition

Chapter 22: Interconnects for 2D and 3D Architectures

http://eps.ieee.org/hir

We acknowledge with gratitude the use of material and figures in this Roadmap that are excerpted from original sources.
Figures & tables should be re-used only with the permission of the original source.

IEEE ELECTRONICS PACKAGING SOCIETY
ELECTRON DEVICES SOCIETY
IEEE Photonics Society
ASME SETTING THE STANDARD
semi

Chapter 22: Interconnects for 2D and 3D Architectures

Executive Summary

With increasing interest in on-package integration, there is a need to describe package architectures and their interconnect capabilities in a simple and consistent manner. This chapter has two primary objectives: to (a) define and proliferate a new standardized nomenclature for package architectures covering and clearly demarcating both 2D and 3D[1] constructions and to (b) define and proliferate key metrics driving the evolution of the physical interconnects in these architectures.

Key contributors:

Raja Swaminathan (chair)	Ravi Mahajan (chair)	Michael Alfano	Adeel A. Bajwa
Rozalia Beica	Tom DeBonis	Takafumi Fukushima	Paul Franzon
Kanad Ghose	John Hunt	Subramanian Iyer	Steffen Kroehnert
Debendra Mallik	Kaushik Mysore	Kwok Ng	Peter Ramm
Venky Sundaram	Jan Vardaman	Markus Wimplinger	Thom Gregorich
Sam Karikalan	Dave Armstrong		

1. Introduction

Moore's Law Scaling has paced growth of the microelectronics industry for the last 50 years by providing a template for silicon scaling and homogeneous SoC (**S**ystem **o**n **C**hip) integration of different IP. Moving forward, heterogeneous integration, enabled by package and microsystems scaling, is expected to increasingly complement Moore's Law scaling and continue to provide improved functionality. Scaling of current and new package architectures (including 3D architectures [1, 2] and architectures currently designated as 2.1D, 2.3D or 2.5D architectures [3, 4, 5]) are projected to be major enablers to sustain and enhance growth in the microelectronics industry. These architectures enable novel heterogeneous SiP (**S**ystem **i**n **P**ackage) integration and represent key innovations needed for cost-performance optimized microelectronics systems. Historically, the primary purpose of the package was to provide mechanical protection for the die, and space transformation for silicon features. Over the past few decades, packaging technologies have scaled to act as cost-effective space transformers for silicon devices to enable transistor scaling and to support SoC integration. Innovations in packaging have focused on minimizing impact to the power, performance and latency of silicon and maximizing performance opportunities made possible by silicon scaling. While mainly focused on supporting homogeneous integration, the semiconductor packaging industry has also been producing MCPs (**M**ulti-**C**hip **P**ackages) for a few decades, primarily for improved time-to-market and for critical heterogeneous integration needs (e.g. DRAM integration). Today's industry trends indicate an increasing need for heterogeneous integration driven by a need to add diverse functionality (often realized on different IP on silicon nodes from multiple different suppliers) [6], improved silicon yield resiliency and the continued need for heterogeneous integration. 2D and 3D package architectures are ideal heterogeneous integration platforms because they provide short, power efficient, high-bandwidth connections between components in compact form factors.

Heterogeneous packaging technologies:

- Deliver power-efficient, high-bandwidth on-package IO links that employ differing communication protocols;
- Enable a diversity of off-package IO protocols;
- Deliver noise isolation for single ended and differential signals;
- Manage increasing cooling demands;
- Support complex power delivery architectures;
- Meet diverse application functionality ranging from high performance servers to flexible, wearable electronics;
- Meet a broad spectrum of reliability requirements for different market segments and applications;
- Provide cost effective, high precision quick turn assembly.

[1] Scope of this chapter is restricted to electrical interconnects between one or more semiconductor devices.

Developing products using advanced packaging requires an integrated approach involving collaboration with product architects, system architects, process engineers, materials engineers, and reliability engineers, and a detailed understanding of the fundamental thermal, mechanical and electrical characteristics of the various architectures.

2. Scope

This roadmap chapter has a two-fold purpose:

- Define and proliferate a new standardized nomenclature for package architectures covering, and clearly demarcating, both 2D and 3D constructions. Currently there are a number of intermediate definitions between 2D and 3D constructions, referred to as 2.xD architectures. Experts in this road-mapping effort, representing a wide spectrum of industry, academia and consultants, agree that the current nomenclature (e.g. 2.1D, 2.3D, 2.5D architectures) does not have a common rational basis and that there is a need to provide a comprehensive classification framework based on a common set of assumptions. The objective of this chapter is to drive clarity and provide a nomenclature framework that will house different architectures.
- Define and proliferate key metrics driving the evolution of the physical interconnects in these architectures. This chapter will list their current values (based on the state of the art) and projections for the next generations.

The chapter is organized into 4 primary areas:

- Converged Nomenclature Framework for 2D and 3D Architectures
- Key Metrics:[2]
  - Design Attributes
  - Electrical Attributes including Signaling and Power Delivery
- Difficult Challenges
- Discussion

3. Converged Nomenclature Framework for 2D & 3D Architectures

a. A 2D architecture is defined as an architecture where two or more active silicon devices are placed side-by-side on a package and are interconnected on the package. If the interconnect is "enhanced" i.e. has higher interconnect density than mainstream organic packages, and isaccomplished using an organic medium, the architecture is further sub-categorized as a 2DO (2D Organic) architecture and similarly, if the enhanced architecture uses an inorganic medium (e.g. a silicon/glass/ceramic interposer or bridge) the architecture is further sub-categorized as a 2DS architecture. Architectures that include enhancements over and above traditional 2D architectures (typically 2 or more die flip-chip attached on a traditional organic package) are variously referred to as 2.x architectures to emphasize their specialness. These nomenclatures do not have any particular technical basis. It is proposed here that they all be broadly categorized as 2D enhanced architectures.

b. A 3D architecture is defined as an architecture where two or more active silicon devices are stacked and interconnected **without** the agency of the package.

The ideas described by this nomenclature are schematically shown in Figure 1.

[2] Other key attributes such as thermal and process attributes are covered in different chapters in this roadmap

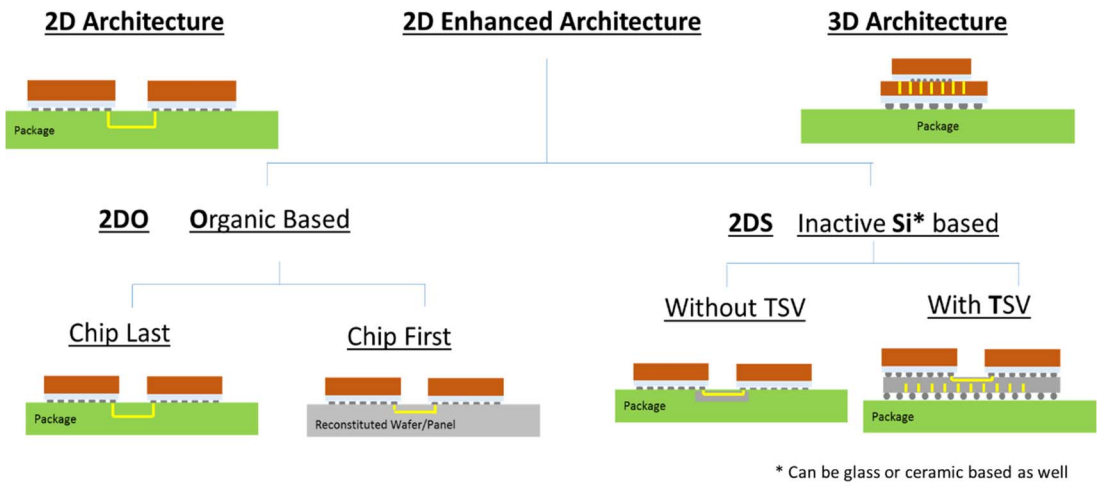

Figure 1: Schematic describing the Converged Nomenclature Framework for 2D & 3D Architectures

4. Interconnect Nomenclature

Package interconnects may be classified as:

(a) **Die-Die Interconnects:** Interconnects between stacked die that enable vertical interconnects between multiple die in a 3-D stack. These may be further sub-categorized using the process these interconnects are created with, which can lead to different physical attributes, such as Die-Die interconnects created using a:

a. Wafer-to-Wafer attach process
b. Die-to-Wafer attach process
c. Die-Die attach process

The roadmap for these interconnects is described in Section 5.1.1.

(b) **On-package Die-Die Interconnects:** i.e. 2D and Enhanced-2D Interconnects: Interconnects between die within the package that enable lateral connections. The roadmap for these interconnects is described in Section 5.1.1.

(c) **Die-to-Package Interconnects:** Interconnects between the die and the package (Figure 2), typically known as the first level interconnect (FLI).

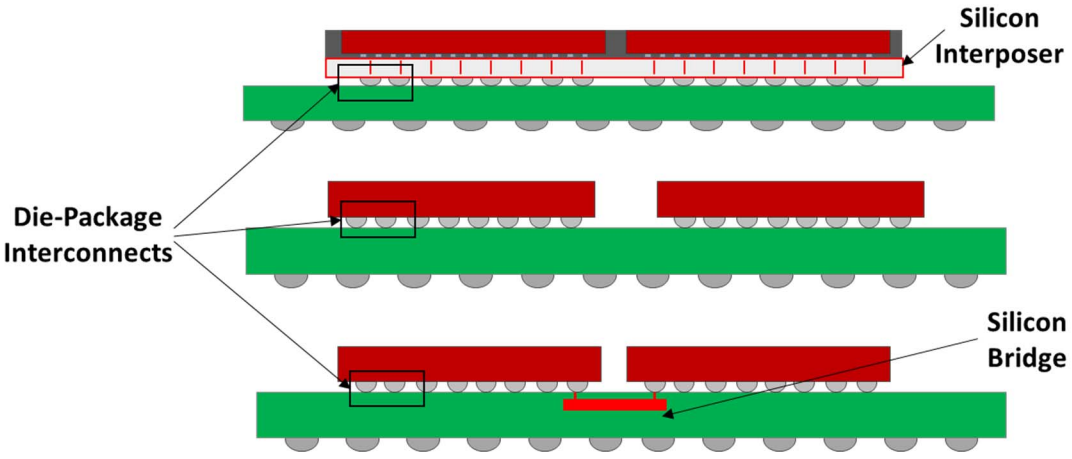

Figure 2: Schematic showing the die-package interconnects[3].

[3] Note that the values discussed in this section do not include the case where the organic substrate is scaled to accept fine pitch die stacks such as HBM @ 55µm, with and without EMIB. Since instances such as these are more relevant to die-die interconnects, they are discussed in Section 5.1.1.

The schematic in Figure 2 only shows area-array interconnects. Wire-bond interconnects are also an important die-to-package interconnect. Three types of wire bonding technologies, Au, Cu and Ag wire-based technologies, are widely used today. The finest in-line wire-bond pitch currently seen in high volume manufacturing (HVM) remains at 40µm inline pitch and has been that way for the last few years. Wire-bonders are capable of supporting a minimum inline pitch of 35µm or 40µm staggered (dual row) pitch in HVM (Au, Cu, or Ag wire). Process advances in recent years have brought Cu wire bonding capabilities just about on par with Au wire bonding capabilities. Additionally, current bonders have successfully demonstrated 30µm inline pitch capability to make sure we stay ahead of packaging requirements. Table 1 shows the best-judged 5-year roadmap from leading wire-bond experts.

Another key metric is the flip-chip pitch for area array interconnects. Table 1 shows a 5-year roadmap for the traditional flip-chip pitch. Given that the pace of change is flat, it is reasonable to assume that the flip-chip pitch will stay at a minimum bound of 90µm. This pitch does not cover the fine pitch scaling available in enhanced 2D and 3D architectures.

Year of Production	2018	2019	2020	2021	2022	2023	2024
Au Wire bond - Single in-line (µm)	40	35	35	30	30	30	30
Cu wire – single inline (µm)	40	35	35	30	30	30	30
Flip chip array, low end & consumer	150	150	130	130	130	130	130
Flip chip – cost performance	110	110	110	100	100	100	90
Flip chip – high performance	110	100	100	90	90	90	90

Table 1: Die-Package Interconnect Pitch Roadmap

(d) **Within-Package Interconnects:** Interconnects within the package that enable lateral connections between two nodes or electrodes. Scaling projections of within-package interconnects are not discussed in this chapter. The reader is referred to the chapter on package substrate technologies (chapter 8).

(e) **Package-to-Board Interconnects:** Interconnects between the package and the next level, which is typically the motherboard, are referred to as the second level interconnect (SLI). SLI connections are either socketed or BGA. The 2015 ITRS roadmap projections for socket pin counts are reproduced below [7] in Table 2a. Figure 3 shows a trend graph based on how sockets have actually evolved. The 2015 ITRS projections are reasonable extrapolations for the cost-performance segment (minor changes are shown in Table 2b). For the high-performance segments, the projections look reasonable until ~2021 but seem to be under-projecting significantly after that. This is likely because the pin-count increase trend in the 2015 projections was assumed to be linear. Table 2b shows an updated projection for the high performance segment using a combination of exponential and polynomial fits.

	2019	*2020*	*2021*	*2022*	*2023*	*2024*	*2025*	*2026*	*2027*	*2028*	*2029*	*2030*
Low-end, Low-cost package	550	550	550	600	600	600	600	600	650	650	650	650
Mobile Package	1500	1600	1600	1600	1600	1600	1600	1600	1600	1600	1600	1600
Memory (MCP)	260	280	280	280	280	280	280	280	280	280	280	280
Cost-performance	3200	3300	3400	3500	3600	3700	3800	3900	4000	4100	4200	4300
Harsh	693	728	764	803	843	860	877	894	911	928	945	962
High performance	5394	5651	5934	6231	6543	6855	7167	7479	7791	8103	8415	8727

Table 2a: Table HI-14 from the 2015 ITRS [7]

	2019	2020	2021	2022	2025	2028	2031	2034
Cost performance	3200	3300	3400	3500	3800	4100	4400	4700
High performance	5125	5694	6302	6946	9105	11601	14434	17604

Table 2b: Updated projections for the Cost-Performance and High-Performance segments

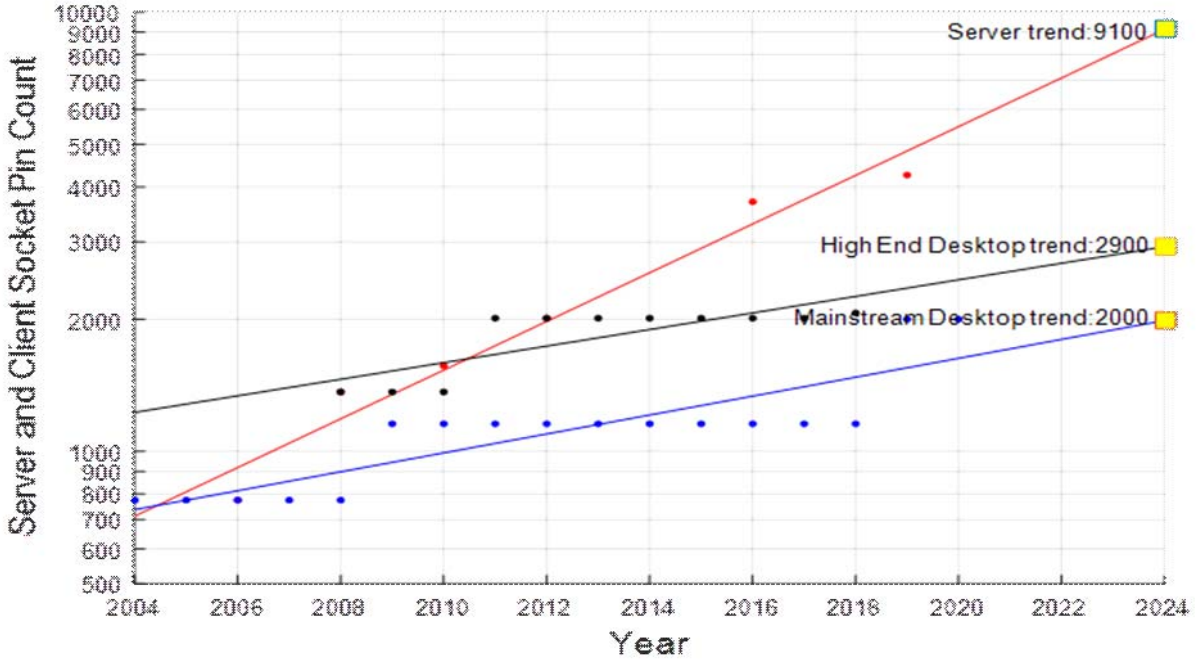

Figure 3: Near term trend graph based on actual pin count evolution. (Source: Intel)

As described in [8], off-package bandwidth, electrical lane speeds and ASIC IO continue to scale steadily. In addition to pin-count scaling, socket constructions that minimize signaling losses should be developed. 2015 ITRS projections for BGA pitch continue to be valid.

(f) **POP (Package-on-Package) Interconnects:** The PoP construction [9] allows for packages to be placed on top of other packages using peripheral package interconnects, also referred to as VI (Vertical Interconnects). It is typically used to stack memory packages on logic to create compact form factors. One such typical construction is shown in Figure 4.

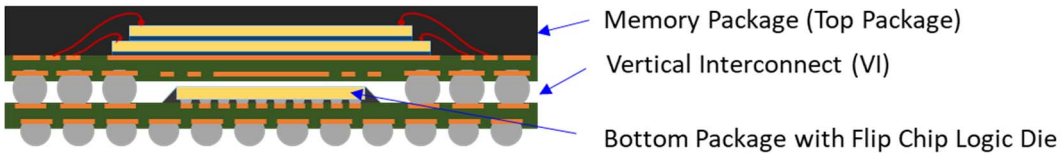

Figure 4: Typical Package-on-Package Architecture.

The VI pitch and the overall height of the package are two key characteristics for this architecture. Currently there is no methodology to project a roadmap for these architectures and in lieu of such a roadmap, the state-of-the-art pitches and package heights, along with their projected changes, are listed in Table 3.

PoP Architectures	VI Pitch (mm)	Maximum Bottom Package Height (mm)
Bare Die PoP	0.5	0.75
Bare Die PoP with 2-Step solder resist (SR) + solder on pad (SOP)	0.4	0.75
TMV PoP	0.4	0.78
Exposed Die TMV PoP	0.35 → 0.27	0.69
Interposer PoP	0.27 → 0.20	0.67
FOWLP PoP	0.30 → 0.20	0.50 → 0.30

Table 3: State-of-the-Art Pitches and Package Heights and their projected targets for PoP Architectures

5. Key Metrics

5.1 Design Attributes[4]

5.1.1 Peripheral Interconnects for 2D and Enhanced-2D Architectures

A key role of packaging is to provide physical interconnects.[5] The two design metrics that describe capability of these interconnects are linear escape and areal escape. These two metrics are shown in Figure 5.

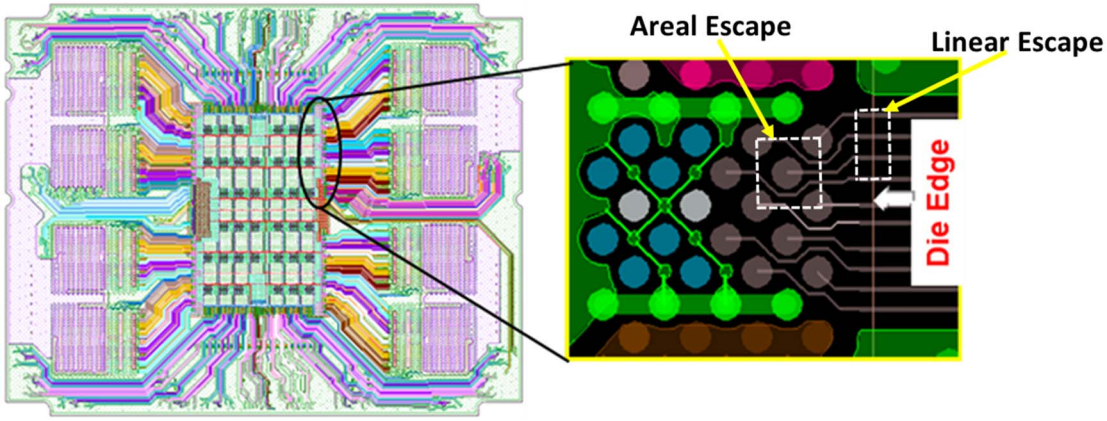

Figure 5: Two Key Physical Design Attributes: (a) IO/mm (of die edge) - Linear Escape Density and (b) IO/mm2 (of die edge) - Areal Escape Density. Note that the term IO here refers to physical bumps and wires

[4] An underlying premise of this work is that the 2D to 3D packaging architectures provide the physical construction architecture to enable signaling and power delivery. To the first order, these physical constructions are agnostic to the IO protocols for which they are used. Hence all attributes described here are independent of the IO protocol.

[5] These interconnects must be designed to minimize power dissipation and signal distortion in addition to provide effective interconnects.

Generations[6]		1	2	3	4	5
Raw Bandwidth Density (GBps/mm)[7,8,9]		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)	55	50	40	35	30
	IO/mm[10]	500	667	1000	1500	2000
	IO/mm^2	331	400	625	816	1111
Signaling Speed (Gbps)[11]		2	3	4	5.33	8

Table 4: Interconnect Pitch Roadmap for Solder based Interconnects

Generations		1	2	3	4	5
Raw Bandwidth Density (GBps/mm)		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)[12]	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

Table 5: Interconnect Pitch Roadmap. Interconnects will transition away from solder gradually as pitches scale to less than 30□m.

5.1.2 Area Interconnects for 3D Architectures

Generations		1	2	3	4	5
Raw Bandwidth Density (Gbps/mm^2)[13]		125	250	500	1000	2000
Package Technology	Minimum Bump Pitch (µm)[14]	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

Table 6: Interconnect Pitch Roadmap. Interconnects will transition away from solder gradually as pitches scale to less than 30□m.

5.3 Signaling Attributes

Generations		1	2	3	4	5
Bandwidth Density (GBps/mm)		125	250	500	1000	2000
Channel Performance	Channel Length (mm)	2	1.5-1.9	1.2-1.8	1.0-1.8	0.5-1.8
	Total Channel RC (ps)	100	75-90	50-70	45-70	20-60

Table 7: Channel Signaling Characteristics for 2D Architectures

[6] At present there is no universal understanding of the required gap between generations. The TWG judgment is that it will be a minimum of 2 years, and from a planning perspective we recommend a maximum of 3 years, to ensure that the interconnect roadmap is competitive.

[7] Per mm of die edge

[8] Starting value of 125GBps is estimated raw bandwidth possible in an AIB style implementation.

[9] Raw Bandwidth is essentially the product of # of connections and signaling speed per connection. Achieved bandwidth will be lower since not all connections are used for data transmission. The starting point of 125GBps is a judged value.

[10] Since multiple silicon back-end layers or package layers can be used to route the bumps, specific geometrical features of the layers are not described.

[11] Representative example showing how the BW goals are reached. These speeds are not unique.

[12] Starting value of 55µm is based on the most common current implementation i.e. HBM

[13] Per mm^2 of die area

[14] Starting value of 40µm bump pitch based on known implementations in HBM and WIO2 .

5.3.2 Area Interconnects

Generations	1	2	3	4	5
Bandwidth Density (Gbps/mm^2)	125	250	500	1000	2000
Bump Capacitance (fF)	30	22	15	8	4

Table 8: Channel Signaling Characteristics for 3D Architectures

5.4 Power Delivery Attributes

5.4.1 Area Interconnects for 2D and 3D Architectures

Generations		1	2	3	4	5
Core Power Density (W/sqmm)		5	80	12	18	25
	On-die Capacitance Density (nF/mm^2)	20	30	45	67	100
	VR Power Density (W/mm^2)	0.4	0.6	1	1.5	2
	Ceramic Cap Density (µF/mm^2)	60	90	140	200	300
	Bump Current Carrying Capability (A/mm^2)	6	9	14	20	30

Table 9: Power delivery Attributes for 2D and 3D Architectures. It should be noted that power delivery attributes are agnostic to the architecture.

6. Difficult Challenges

The high IO/mm values listed in Tables 4 and 5 are achieved using silicon back-end technologies to create thin, closely spaced wires (Figure 6). This roadmap projects the need for increasing density, i.e. reduced line pitch. When combined with increasing signal speeds, there will be greater concerns about signal quality due to increased cross-talk caused by the reduced line spacing. The packaging community will be challenged to develop solutions that minimize impact to signal integrity and provide physical links with improved power efficiency.[15]

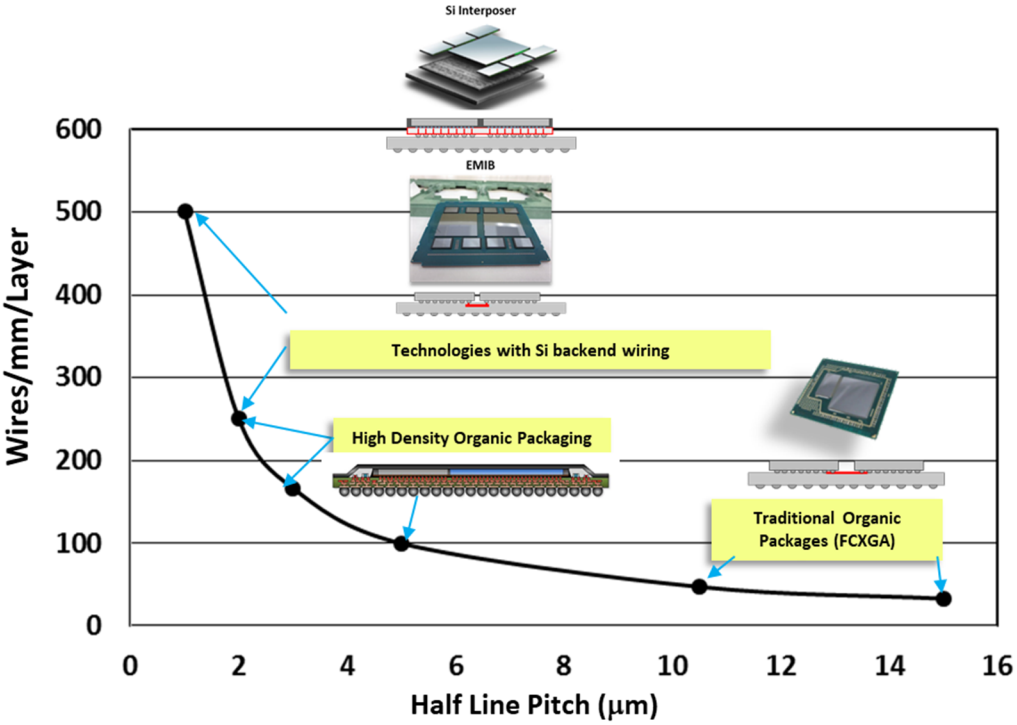

Figure 6: Technologies for different wiring features. L is the width of the line in µm, S is the minimum space between lines in µm; half line pitch is (L+S)/2. Technologies that use silicon backend wiring can achieve wiring densities of greater than 1000 with L & S ≤ 0.5µm.

[15] Power efficiency (measured in pJ/bit) is a sum of Tx, Rx and link power efficiency. The die-die links need to provide reducing RC (Table 6) to ensure improved power efficiency.

There will be greater need to enable novel assembly technologies for ultra-fine pitch enhanced-2D and 3D architectures using both solder and non-solder based approaches. A number of researchers have demonstrated the reduced bump pitches described in Table 4 and there is a fairly good understanding of the technologies needed to transition from solder-based interconnects to solderless interconnects [10, 11]. Key challenges for stacked-die architectures will continue to be in fine pitch sort/test, thermal management, power delivery network development, design process co-design, and equipment readiness for high volume.

7. Discussion

The primary driver for advanced 2D and 3D technologies is the need for increased interconnect densities to support heterogeneous integration and deliver increasing bandwidth. Physical interconnects (i.e. wires, bumps) and link RC characteristics are completely under the control of packaging technologists and it is relatively easy to establish a scaling roadmap. Of these, the bump pitch roadmap can be unambiguously stated whereas the IO/mm roadmap can be achieved in more ways than one, and hence the specification of IO/mm scaling instead of wire feature scaling. We anticipate that moving forward, this chapter will spur discussion among product architects and will help develop clarity on various use cases that will drive the pace of technology innovation [12-15].

8. Acknowledgments

The TWG members would like to acknowledge Zhiguo Qian, Kemal Aygun, Kaladhar Radhakrishnan, Srikant Nekkanty and Bob Sankman from Intel Corporation for their help in describing some of the key attributes and their numerical values in this chapter. We also acknowledge the contributions of other TWGs.

9. References

1. https://www.engineering.com/Hardware/ArticleID/16894/TSMCs-New-Wafer-on-Wafer-Process-to-Empower-NVIDIA-and-AMD-GPU-Designs.aspx
2. https://www.anandtech.com/show/13699/intel-architecture-day-2018-core-future-hybrid-x86/6
3. http://www1.semi.org/eu/sites/semi.org/files/docs/7_Amkor_NaniumSymposium_Jun26th2013.pdf
4. K. Oi, et al, "Development of New 2.5D Package with Novel Integrated Organic Interposer Substrate with Ultra-Fine Wiring and High Density Bumps," Proceedings of the 2014 IEEE 64th Electronic Components and Technology Conference, Orlando, pp. 348-353.
5. Embedded Multi-Die Interconnect Bridge (EMIB) – A High Density, High Bandwidth Packaging Interconnect, R.Mahajan et.al, 2016 IEEE 66th Electronic Components and Technology Conference, DOI 10.1109/ECTC.2016.201
6. https://www.darpa.mil/program/common-heterogeneous-integration-and-ip-reuse-strategies
7. https://www.semiconductors.org/resources/2015-international-technology-roadmap-for-semiconductors-itrs/
8. http://www.ieee802.org/3/cfi/1117_3/CFI_03_1117.pdf
9. R. Co et al., "High-Volume-Manufacturing (HMV) of BVA Enabled Advanced Package-on-Package (PoP)," ICEP, April 14-17, 2015.
10. Adeel A. Bajwa, SivaChandra Jangam, Saptadeep Pal, Niteesh Marathe, Tingyu Bai, Takafumi Fukushima, Mark Goorsky, Subramanian S. Iyer, "Heterogeneous Integration at Fine Pitch 10 µm) using Thermal Compression Bonding", Proc. ECTC 2017, pp. 1276-1283.
11. Eric Beyne, Soon-Wook Kim, Lan Peng, Nancy Heylen, Joke De Messemaeker, Oguzhan Orkut Okudur, Alain Phommahaxay, Tae-Gon Kim, Michele Stucchi, Dimitrios Velenis, Andy Miller, and Gerald Beyer imec, Leuven, Belgium, "Scalable, sub 2µm Pitch, Cu/SiCN to Cu/SiCN Hybrid Wafer-to-Wafer Bonding Technology, IEDM Tech. Dig., 2017, pp.729-732.
12. K. Cho et al., "Design optimization of high bandwidth memory (HBM) interposer considering signal integrity," 2015 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS), Seoul, 2015, pp. 15-18.
13. S. Jangam, A. A. Bajwa, K. K. Thankkappan, P. Kittur and S. S. Iyer, "Electrical Characterization of High Performance Fine Pitch Interconnects in Silicon-Interconnect Fabric," 2018 IEEE 68th Electronic Components and Technology Conference (ECTC), San Diego, CA, 2018, pp. 1283-1288.
14. Ahmet C. Durgun, Zhiguo Qian, Kemal Aygun, Ravi Mahajan, Tim Tri Hoang, Sergey Yuryevich Shumarayev, "Electrical Performance Limits of Fine Pitch Interconnects for Heterogeneous Integration," accepted for publication, ECTC Las Vegas, 2019.
15. Sergey Y. Shumarayev, Conor O'Keeffe*, Tim T. Hoang, David Kehlet, Sangeeta Raman, Benjamin Esposito, A SiP Standard for Reusable Chiplet Enabled Platforms, GOMACTech Conference, March 25-28, 2019 - Albuquerque, NM, http://www.gomactech.net/.

Edited by Paul Wesling