Chapter 22: Interconnects for 2D and 3D Architectures

For updates, visit eps.ieee.org/hir

The HIR is devised and intended for technology assessment only and is without regard to any commercial considerations pertaining to individual products or equipment.

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.
Chapter 22: Interconnects for 2D and 3D Architectures

Executive Summary

With increasing interest in on-package Heterogeneous Integration (HI), 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 standardized nomenclature for package architectures covering and clearly demarcating both 2D and 3D constructions and to (b) define and proliferate key metrics driving the evolution of the physical interconnects in these architectures.

Key Contributors:

<table>
<thead>
<tr>
<th>Name</th>
<th>Name</th>
<th>Name</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ravi Mahajan (Chair)</td>
<td>Raja Swaminathan</td>
<td>Subramanian Iyer</td>
<td>Kemal Aygun</td>
</tr>
<tr>
<td>Jan Vardaman</td>
<td>Zhiguo Qian</td>
<td>Kaladhar Radhakrishnan</td>
<td>Srikant Nekkanty</td>
</tr>
<tr>
<td>Tom DeBonis</td>
<td>Dishit Parekh</td>
<td>Vidya Jayaram</td>
<td>Debendra Mallik</td>
</tr>
<tr>
<td>Kanad Ghose</td>
<td>Krutikesh Sahoo</td>
<td>Adeel A. Bajwa</td>
<td>Paul Franzon</td>
</tr>
<tr>
<td>Gamal Refai-Ahmed</td>
<td>Kaushik Mysore</td>
<td>Peter Ramm</td>
<td>Steffen Kroehnert</td>
</tr>
<tr>
<td>Venky Sundaram</td>
<td>Markus Wimlinger</td>
<td>Thom Gregorich</td>
<td>Sam Karikalan</td>
</tr>
<tr>
<td>Dave Armstrong</td>
<td>Chintan Buch</td>
<td>SB Park</td>
<td>Takafumi Fukushima</td>
</tr>
<tr>
<td>Chandrasekhar Mandalapu</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Introduction

Moore’s Law has paced the growth of the microelectronics industry for the last 50 years by providing a template for silicon scaling and homogeneous SoC (System on a Chip) integration of different IP (Intellectual Property) circuits. Moving forward, HI, enabled by changes in the physical, electrical, thermal, and thermo-mechanical attributes of packages and micro-systems, increasingly complements Moore’s Law to provide improved functionality [1-7]. Current and new advanced package architectures are major enablers in sustaining and enhancing growth in the micro-electronics industry [8-22]. These architectures enable novel heterogeneous SIP (System in Package) configurations for cost-performance optimized micro-electronics systems. A number of products that use advanced HI have been announced in recent years attesting to the importance of this field [23-28].

Historically, the primary purpose of the package for homogenous integration was to provide mechanical protection for the die, space transformation for silicon features, form factor scaling, low parasitic power-delivery, efficient power removal and low loss, high bandwidth signaling. Innovations in packaging for homogeneous SoCs, focused on enabling silicon size scaling, power, performance, and latency while maximizing performance opportunities made possible by Moore’s Law. During the period where the primary focus was homogeneous integration, MCPs (Multi-Chip Packages) were used primarily for improved time-to-market and for critical HI needs (e.g., DRAM integration) [29].

1 Scope of this chapter is restricted to electrical interconnects between one or more semiconductor devices.
Industry trends today show an increasing need for HI driven by a need to add diverse functionality (often realized with different IP on silicon nodes from multiple different suppliers), improved silicon yield resiliency and the continued need for rapid time to market. Additionally, compact HI sub-systems using (mostly) advanced packages substrates today enable significantly larger die area (>3x reticle equivalent). 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. As bandwidth increases, the power consumed transferring the data and the impact of delay time in transit both benefit from the shorter path lengths enabled by advanced 2D and 3D interconnects. Today’s heterogeneous packaging technologies:

- Deliver power-efficient, high-bandwidth on-package IO links using different communication protocols
- Enable a diversity of off-package IO protocols
- Deliver noise isolation for single ended and differential on-package and off-package signaling
- Manage increasing cooling demands
- Support complex power delivery architectures
- Meet diverse application functionality, form factor and weight constraints 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 and quick turn assembly to meet fast production ramps

In this respect they differ from packaging for homogeneous integration in terms of increased complexity and the increased focus on on-package bandwidth. 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, electrical and materials characteristics of the various architectures.

2. Scope

This roadmap chapter has a two-fold purpose:

- Define and proliferate a 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

² Other key attributes such as thermal and process attributes are covered in different chapters in this roadmap
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 is accomplished using an organic medium, the architecture is further sub-categorized as a 2DO (2D Organic)³ 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 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 nomenclature⁴ are schematically shown in Figure 1.

---

³ Accomplished by using a monolithic high-density substrate or more likely a high density organic interposer [10, 11, 16]
⁴ Figure 1 only describes a nomenclature and technology equivalence between architectures is not intended or implied.
⁵ Figure 1 is expected to undergo considerable changes in an upcoming revision, likely in the second half of 2024, to reflect recent technology developments. E.g., a 2021 paper points to three categories i.e. Die-First, Die-Middle and Die-Last [30].
4. Interconnect Nomenclature

Package interconnects can be classified as:

(a) **Die-Die Interconnects**: Interconnects between stacked die for vertical connectivity between multiple dies 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 (W2W) attach process (Wafers could be reconstituted e.g. rW2rW or W2rW)
   b. Die-to-Wafer (D2W) attach process (Wafer could be reconstituted i.e. D2rW)
   c. Die-Die (D2D) attach process

The roadmap for these interconnects is described in Section 5.1.1.

(b) **On-package Die-Die Interconnects**: i.e., 2D and 2D Enhanced Interconnects: Interconnects between die (and/or die stacks/pre-packaged die) within the package for lateral connectivity. 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).

The schematic in Figure 2 only shows area-array interconnects. Wire-bond interconnects are also an important die-to-package interconnect. For details on the wire-bonding technology including discussions on multi-tier stacking that allows for considerable innovations on heterogeneous integration, the reader is referred to the chapter on Single Chip and Multi-Chip Integration in this roadmap [32].

Another key metric is the flip-chip pitch for area array interconnects. Table 1 shows a 7-year trend for the traditional flip-chip pitch. Given that the pace of change is flat, and that the utility of finer bump pitch is limited by feature sizes within the substrate (line/space, via pad, etc.) it is reasonable to assume that 2D flip-chip pitch will stay at a minimum bound of 90µm (this does not cover the fine pitch scaling available in 2D Enhanced and 3D architectures).

---

6 An in-depth analysis for die-to-die and die-to-wafer interconnects is also presented in a recently published report on Manufacturing Roadmap for Heterogeneous Integration and Electronics Packaging (MRHIEP) [31]. The reader is referred to chapters 1 and 2.

7 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.

8 Statement limited to full area array interconnects since there have been cases (e.g. Qualcomm® Snapdragon™ 820) where 80µm pitch on embedded FLI pad has been shown for an essentially 3 rows deep peripheral interconnect.
Table 1: Die-Package Interconnect Pitch Roadmap

<table>
<thead>
<tr>
<th>Year of Production</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
<th>2024</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flip chip array, low end &amp; consumer (µm)</td>
<td>150</td>
<td>150</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>Flip chip – cost performance (µm)</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>Flip chip – high performance (µm)</td>
<td>110</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

(d) **Within-Package Interconnects:** Interconnects within the package that enable lateral connections between two or more die. Roadmap 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 ball grid array (BGA) and may be combined with on package cabling\(^9\). The 2015 ITRS roadmap projections for socket pin counts are reproduced below [33] in Table 2a\(^10\). Figure 3 shows a trend graph based on how sockets have evolved, thus far with an exponential pin count growth. While 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 seem to be under-projecting significantly after 2021. This is likely because the pin-count increase trend in the 2015 projections was assumed to be linear. Figure 4 shows an updated projection for the high-performance segment. The exponential pin-count growth is expected to continue for high performance compute segment due to heterogeneous integration and especially with the growth of artificial intelligence and data center applications demanding high bandwidth memory and I/O signaling with more than 15K pins projected before the end of this decade [34].

Table 2a: Socket pin count projections from 2015 ITRS [33]

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

---

\(^9\) This chapter does not address on package cabling solutions or provide a roadmap since this class of solutions is still maturing.

\(^10\) The ball count for Mobile packages has been removed from this table. The 2015 projections show a flat trend, and this trend needs additional study in light of the current evolution of mobile products.
Table 2b: Updated projections for socket pin count in the Cost-Performance and High-Performance segments

<table>
<thead>
<tr>
<th>Year</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2025</th>
<th>2028</th>
<th>2031</th>
<th>2034</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost performance</td>
<td>3200</td>
<td>3300</td>
<td>3400</td>
<td>3500</td>
<td>3800</td>
<td>4100</td>
<td>4400</td>
<td>4700</td>
</tr>
</tbody>
</table>

Figure 3: Socket pin count trend till 2024 (Source: Intel).

Figure 4: Historical high performance compute socket pin growth extrapolated out to the end of the decade.
As described in [35], off-package bandwidth, electrical lane speeds and ASIC IO continue to scale steadily. In addition to pin-count changes, socket constructions that minimize signaling losses should be developed. 2015 ITRS projections for BGA\(^\text{11}\) pitch continue to be valid.

**(f) POP (Package-on-Package) Interconnects:** The PoP construction [36] 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 5: Typical Package-on-Package Architecture.](image)

The VI pitch (i.e. pitch of interconnects between top and bottom package) 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.

<table>
<thead>
<tr>
<th>PoP Architectures</th>
<th>VI Pitch (mm)</th>
<th>Maximum Bottom Package Height (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare Die PoP</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td>Bare Die PoP with 2-Step solder resist (SR) + solder on pad (SOP)</td>
<td>0.4</td>
<td>0.75</td>
</tr>
<tr>
<td>TMV (Through Mold Via) PoP</td>
<td>0.4</td>
<td>0.78</td>
</tr>
<tr>
<td>Exposed Die TMV PoP</td>
<td>0.35 → 0.27</td>
<td>0.69</td>
</tr>
<tr>
<td>Interposer PoP</td>
<td>0.27 → 0.20</td>
<td>0.67</td>
</tr>
<tr>
<td>FOWLP PoP</td>
<td>0.30 → 0.20</td>
<td>0.50 → 0.30</td>
</tr>
</tbody>
</table>

*Table 3: State-of-the-Art Pitches and Package Heights and their projected targets for PoP Architectures (Source: TechSearch International)*

\(^{11}\) The focus in BGA interconnects and more broadly in other solder joints is on low temperature solders to provide additional options within the reflow hierarchy and opportunities for reduced power usage during manufacturing (hence greater environmental friendliness). There is extensive literature in this area in recent years, however it is not covered in this chapter.
5. Key Metrics

5.1 Design Attributes

The physical attributes and signaling speeds needed to enable generation over generation Bandwidth (BW) doubling are described in this section.

5.1.1 Peripheral Interconnects for 2D and 2D Enhanced Architectures (see Figure 1)

A key role of packaging is to provide physical interconnects. Two design metrics that describe capability of these interconnects are linear escape density and areal escape density. These two metrics are shown in Figure 6A. Note that these two can be combined into a single metric by multiplying the two (Figure 6B). The same metric has also been described by other researchers [37]. Tables 4-6 show possible ways of physical interconnect scaling instead of only signaling speed scaling to achieve raw bandwidth doubling (or greater) generation over generation. Such scaling can achieve aggressive bandwidth targets while keeping the signaling speeds relatively low, so that it is feasible to achieve a very low raw bit error rate (BER) such as $10^{-18}$. This avoids the complicated error correction circuits and minimizes the I/O latency. It is important to emphasize that the physical IO dimensions listed are not bounding parameters and depend on specific business and technical demands i.e., the same bandwidth doubling can be achieved through more or less aggressive choices of the physical IO dimensions.

![Figure 6A: Two Key Physical Design Attributes: (a) IO/mm (of die edge) – Linear Escape Density and (b) IO/mm2 (of die) – Areal Escape Density. Note that the term IO here refers to physical bumps and wires.](image)

---

12 2D to 3D packaging architectures provide the physical construction architecture to enable signaling and power delivery. 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.

13 These interconnects must be designed to minimize power consumption and signal distortion in addition to provide effective interconnects [38-40].

14 Previous revisions of the roadmap only showed bandwidth doubling. However, as this revision quantitatively shows, greater than raw bandwidth doubling is possible.
System Scaling – 3DID Roadmap

3D Interconnect Density: \( \frac{\text{# lines}}{\text{mm}} \times \frac{\text{# verticals}}{\text{mm}^2} \)

- D. Yu, 2019 Dec. IEDM Panel, San Francisco, Ca, USA
- D. Yu, 2020 May IEEE ECTC Keynote. 2020 Aug. TSMC Technology Symposium,

Figure 6B: Two Key Physical Design Attributes: (a) IO/mm (of die edge) – Linear Escape Density and (b) IO/mm² (of die) – Areal Escape Density are multiplied to create a single 3D Interconnect Density (3DID). Note that the term IO here refers to physical bumps and wires. (Ack: TSMC)

<table>
<thead>
<tr>
<th>Generation Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Linear Bandwidth Density (GBps/mm)</td>
<td>125</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td>Package Technology</td>
<td>Minimum Bump Pitch (µm)</td>
<td>55</td>
<td>50</td>
<td>40</td>
<td>35</td>
</tr>
</tbody>
</table>

16 Acronyms CoWoS (Chip-on-Wafer-on-Substrate) and InFO (Integrated Fan-Out) are well known in the industry.
17 At present there is no universal understanding/agreement of the required time 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. Both these numbers reflect the leading edge and mainstream implementations will depend on both economic and technical factors.
18 Per mm of die edge.
19 Starting value of 125 GBps/mm is estimated raw bandwidth possible in an AIB style implementation.
20 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 125 GBps/mm is a judged value.
21 Note that @ 32Gbps UClé-A theoretical BW@ 45um pitch is 2.2716 TBps & this roadmap states @ 3Gbps & @ 40µm pitch a theoretical BW of 0.25TBps is possible. There is no discrepancy in the numbers and modulating speed (hence power efficiency) is a knob for scaling BW.
<table>
<thead>
<tr>
<th>Generation Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Linear Bandwidth Density (GBps/mm)</td>
<td>125</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>2000</td>
<td>4000</td>
<td>8000</td>
</tr>
<tr>
<td>Package Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum Bump Pitch (µm)²³</td>
<td>55</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>Linear Escape Density (IO/mm)</td>
<td>500</td>
<td>667</td>
<td>1000</td>
<td>1333</td>
<td>2000</td>
<td>4000</td>
<td>8000</td>
</tr>
<tr>
<td>Areal Escape Density (IO/mm²)</td>
<td>331</td>
<td>625</td>
<td>1111</td>
<td>2500</td>
<td>10000</td>
<td>40000</td>
<td>160000</td>
</tr>
<tr>
<td>Signaling Speed (Gbps)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 5: Physical IO Scaling Roadmap for 2D and 2D Enhanced Architectures that use both solder and hybrid interconnects

21 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.

22 Representative example showing how the BW goals are reached. These speeds are not unique.

23 Starting value of 55µm is based on initial HBM pitch.
5.1.2 Area Interconnects for 3D Architectures (see Figure 1)

<table>
<thead>
<tr>
<th>Generation Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Areal Bandwidth Density (GBps/mm²)</td>
<td>125</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>2000</td>
<td>8000</td>
<td>32000</td>
<td>200000</td>
</tr>
<tr>
<td>Package Technology</td>
<td>Minimum Bump Pitch (µm)</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Areal Escape Density (IO/mm²)</td>
<td>625</td>
<td>1111</td>
<td>2500</td>
<td>4444</td>
<td>10000</td>
<td>40000</td>
<td>160000</td>
</tr>
<tr>
<td></td>
<td>Signaling Speed (Gbps)</td>
<td>1.6</td>
<td>1.8</td>
<td>1.6</td>
<td>1.8</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Table 6: Physical IO Scaling Roadmap for 3D architectures that use both solder and hybrid interconnects

5.2 Signal Integrity Attributes

5.2.1 Signal Integrity Attributes for 2D and 2D Enhanced Architectures

The short and high-density interconnects described in Section 5.1.1 become more and more RC dominated with the generational scaling. The interconnect inductance is expected to have a secondary effect on the channel performance. Table 7 shows the reduction of channel length with the bump pitch scaling. This does compensate the larger per-unit-length RC of denser routing interconnects, so that the total channel RC loss does not increase from generation to generation.

<table>
<thead>
<tr>
<th>Generation Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Bandwidth Density (GBps/mm)</td>
<td>125</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>2000</td>
<td>4000</td>
<td>8000</td>
</tr>
<tr>
<td>Channel Performance</td>
<td>Channel Length (mm)</td>
<td>&lt;2.0</td>
<td>&lt;1.7</td>
<td>&lt;1.4</td>
<td>&lt;0.8</td>
<td>&lt;0.50</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td></td>
<td>Bump-to-Bump Channel RC (ps)</td>
<td>&lt;10</td>
<td>&lt;10</td>
<td>&lt;10</td>
<td>&lt;10</td>
<td>&lt;10</td>
<td>&lt;10</td>
</tr>
</tbody>
</table>

Table 7: Channel Signaling Characteristics for 2D and 2D Enhanced Architectures

---

24 Per mm² of die area.
25 Note that previous versions of this chapter had shown a range of signaling speeds. The range has been removed to simplify the table.
5.2.2 Area Interconnects

The signal integrity performance of the extremely short area vertical interconnects is dominated by their capacitance. Their resistance and inductance are expected to have a secondary impact on the channel performance with the generational scaling. Table 8 shows the scaling of bump capacitance with the bump pitch reduction. The improved capacitance is an important enabler to avoid the need for stronger driver design.

<table>
<thead>
<tr>
<th>Generation Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Areal Bandwidth Density (GBps/mm²)</td>
<td>125</td>
<td>250</td>
<td>500</td>
<td>1000</td>
<td>2000</td>
<td>8000</td>
<td>32000</td>
<td>200000</td>
</tr>
<tr>
<td>Bump Capacitance (fF)</td>
<td>&lt;30</td>
<td>&lt;22</td>
<td>&lt;15</td>
<td>&lt;10</td>
<td>&lt;7</td>
<td>&lt;5</td>
<td>&lt;3</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

*Table 8: Channel Signaling Characteristics for 3D Architectures*

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

The main power consumption in microprocessors is from cores rather than the IO domains. As a result, the generational numbers used to represent interconnect scaling do not necessarily correlate to power consumption on the compute cores. As a result, we will scale the power delivery targets as a function of time. Table 9 reflects projected demand and packaging engineers will be challenged and will need to explore new materials and architectures to deliver capacitance and current carrying capabilities. Though not a focus of this chapter, it should be pointed out that with increasing HI, greater focus on power delivery is anticipated including (1) proliferation of IVR integration on package and on chip, (2) embedded inductors and capacitors displacing and or complementing board level components in some emerging architectures, (3) increased and urgent need for much more granular power delivery, and escalation of power rails in the package as "heterogenous integration" continues to increase on a single package, (4) accelerating drivers for vertical power delivery architectures in the package, and last but not the least, (5) impact of the imminent backside power delivery on chip at 2nm node and beyond. In this context, there will be a drive for both discrete passives complemented by more integrated passives.

<table>
<thead>
<tr>
<th>Year</th>
<th>2024</th>
<th>2026</th>
<th>2028</th>
<th>2030</th>
<th>2032</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Core Power Density (W/mm²)</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>Sustained Core Power Density (W/mm²)</td>
<td>1</td>
<td>1.4</td>
<td>2</td>
<td>2.8</td>
<td>4</td>
</tr>
<tr>
<td>On-die MIM Capacitance Density (nF/mm²)</td>
<td>140</td>
<td>210</td>
<td>320</td>
<td>520</td>
<td>800</td>
</tr>
<tr>
<td>VR Power Density (W/mm²)</td>
<td>1</td>
<td>1.4</td>
<td>2</td>
<td>2.8</td>
<td>4</td>
</tr>
<tr>
<td>Ceramic Cap Density (µF/mm²)</td>
<td>10</td>
<td>14</td>
<td>20</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>Sustained Bump Current Carrying Capability (A/mm²)</td>
<td>2.8</td>
<td>4.3</td>
<td>6.1</td>
<td>9.3</td>
<td>13.3</td>
</tr>
<tr>
<td>Sustained TSV Current Carrying Capability (A/mm²)</td>
<td>2.8</td>
<td>4.3</td>
<td>6.1</td>
<td>9.3</td>
<td>13.3</td>
</tr>
</tbody>
</table>

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

26 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
6. Review of Different Packaging Architectures:

In this section, we tabulate the various examples seen in literature with their corresponding architecture type, interconnection materials used, tightest pitches observed and the required processes and equipment to enable them. In addition, we list the resulting applications, reliability concerns and any additional challenges resulting from the same. This is a broad overview intended to offer perspective on all the interconnect technologies.

Table 5. Examples of applications from literature as a function of the different packaging architecture and process/material attributes. [42, 43]

| Architecture Type (2D / 2D Enhanced / 3D) | Interconnect Technologies | Tightest Die to Package Pitch
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2D &amp; 2D Enhanced [45]</td>
<td>Micro-bumps, C4s, TSVs, and passive Si Interposer</td>
<td>20 µm</td>
</tr>
<tr>
<td>2D Enhanced, 3D die-to-die, die-to-wafer [46-48]</td>
<td>Cu pillars, passive and active Si coupons &amp; wafers</td>
<td>&lt; 10 µm</td>
</tr>
<tr>
<td>3D Die-to-Die, Die-to-Wafer [49-58]</td>
<td>Micro-bumps, TSVs, Cu-Cu Bonding</td>
<td>5 µm</td>
</tr>
<tr>
<td>3D Wafer-to-Wafer [50]</td>
<td>Cu-Cu Bonding</td>
<td>0.9 µm</td>
</tr>
</tbody>
</table>

Note 27: Reference envelope values. Values listed in Table in this chapter represent the broader mainstream envelopes.
7. 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 crosstalk, 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.28

![Figure 7: 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.](image)

There will be greater need to enable novel assembly technologies for ultra-fine pitch 2D Enhanced 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 [22-24, 42, 43, 45-58]. Key challenges for stacked-die architectures will continue to be in fine pitch sort/test, thermal management, power delivery network development, design process co-optimization, in-line process control and equipment readiness for high volume.

8. Discussion

The primary driver for advanced 2D and 3D packaging technologies is the need for increased interconnect densities to support HI and deliver increasing bandwidth in a power efficient manner while enabling efficient power delivery. An increasing number of innovative packaging architectures deliver significantly improved HI

---

28 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.
envelopes. In these HI architectures 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 non-unique scaling roadmap. Already a number of power efficient, high bandwidth (custom and standardized) IO links that take advantage of advanced packaging have been defined to further spur the proliferation of heterogeneous integration [59-71]. We anticipate that moving forward, this chapter will spur discussion among product architects and will help develop further clarity on various use cases that will drive the pace of technology innovation. Delivering higher interconnect densities will challenge packaging engineers to develop novel materials, assembly processes and metrologies to develop cost effective technology solutions that meet the performance demands of future HI architectures. Even though it is not the focus of this chapter, it is also important to note that the ability to integrate the right thermal features will define the physical envelope (i.e., form factor and number of die/die stacks that can be integrated on the package) and the warpage characteristics that will ensure manufacturability [72].

9. Acknowledgments

We would like to acknowledge the contributions of other TWGs.

10. References


29 Chapter 6 the MRHIEP report also discusses IO requirements and examples of the IO protocols (UCIe57,58, BoW59,60, and SuperCHIPS63,65) for within package interconnections in context of high performance computing applications [31].


25. <https://www.youtube.com/watch?v=j0z4FweCy4M&t=2814s>

26. <https://www.youtube.com/watch?v=eierRXlaDuzo>

27. <https://www.youtube.com/watch?v=ggAYMx34euU>


40. 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.
60. https://github.com/chipsalliance/AIB-specification
68. High-performance, power-efficient three-dimensional system-in-package designs with universal chiplet interconnect express | Research Communities by Springer Nature
69. High-performance, power-efficient three-dimensional system-in-package designs with universal chiplet interconnect express | Research Communities by Springer Nature