Developing an open chiplets marketplace: Challenges and opportunities

EDITOR(S):
Anu Ramamurthy, Microchip
Arvind Kumar, IBM
Dharmesh Jani, Meta
Eric Ouyang, JCET
Javi DeLaCruz, ARM
Kevin Yee, Samsung
Mudasir Ahmad, Google
Ravi Agarwal, Meta
Roberto Rotili, Teradyne
Yin Yang, Meta

Introduction

The Open Domain Specific Architecture (ODSA) workgroup within the Open Compute Project aims to define and facilitate creation of open D2D interfaces for the chiplets. Conceived in Oct-2018, the ODSA project has grown substantially over the past 2 years and today has a community of over 50 companies spanning IP providers, semiconductor producers, equipment suppliers, EDA tool partners, system vendors and end users spread across 8 workstreams.

The ODSA focuses on common workload driven use-cases which broadly fall under following two categories:-

- Disaggregation of System on Chip into System in Package
- Aggregation of a board to a package

Primary charter of the ODSA has following 3 main pillars:-

1. Develop open die to die interfaces
2. Build reference designs
3. Standardize reference workflows

Business and End user workgroups within ODSA are focused on solving business challenges to enable robust and open chiplet marketplace.

This whitepaper summarizes some of the key activities within these groups that have identified challenges and opportunities to drive successful development of the open chiplet ecosystem.

Business Workgroup Background

The ultimate goal of Open Domain Specific Architecture group in OCP is to establish a thriving and open chiplet marketplace. This requires collaboration on multiple dimensions. Ensuring that companies are able to interact in an efficient and scalable manner is critical to fast development. Creating a sustainable business model requires the development of revenue and cost models that yield via business models for the product managers within semiconductor companies to lean on.

This has been the primary motivation for creation of the business workgroup: To enable product managers across the semiconductor industry to build product business cases on open interfaces developed in OCP. There are many key considerations such as sales enablement, workflow visibility across supply chains, cost modeling, and funding to develop test chips for interface

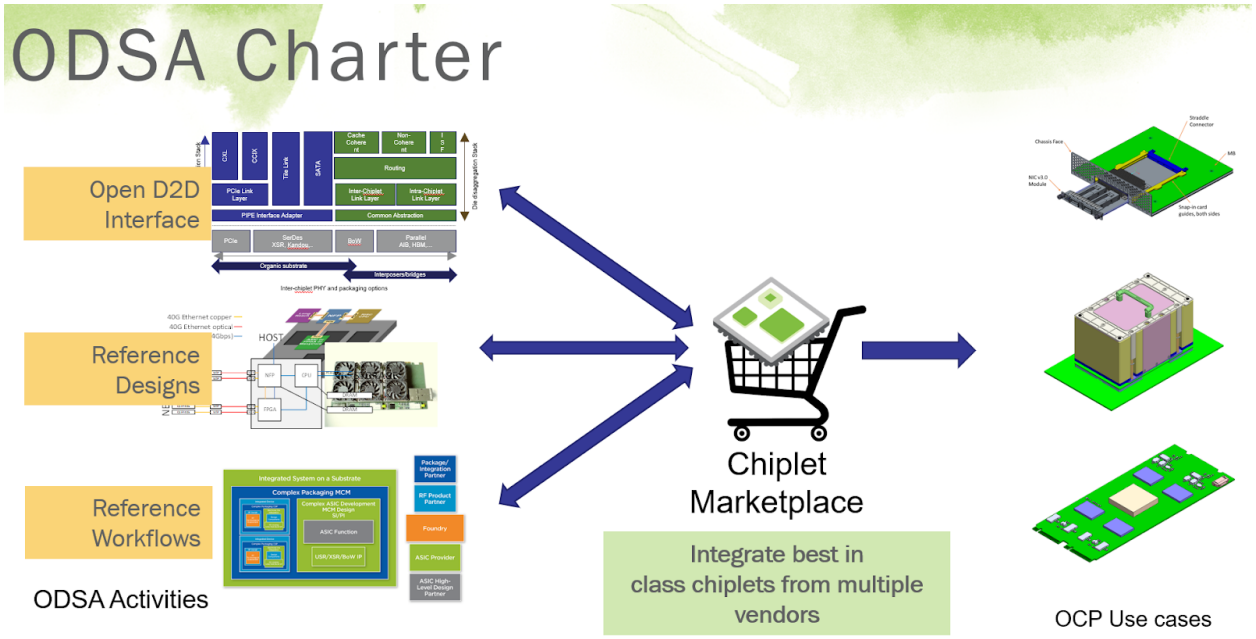


Figure 1: ODSA Goals

characterization and channel model development that can be shared among many parties and delivery and handling of “hardened IP” in the form of chiplet for further downstream integration.

There are numerous business and commercial implications around multi-sourcing from many suppliers. Ecosystem enablement via funding of test chips, seeking foundry support, ensuring health of the die (known good die or KGD), supply chain guarantees, OSAT complexity in inventory management, packaging with predictable end performance and having ability to test for holding SLAs are some of the areas on which business and end user workgroups are focussing.

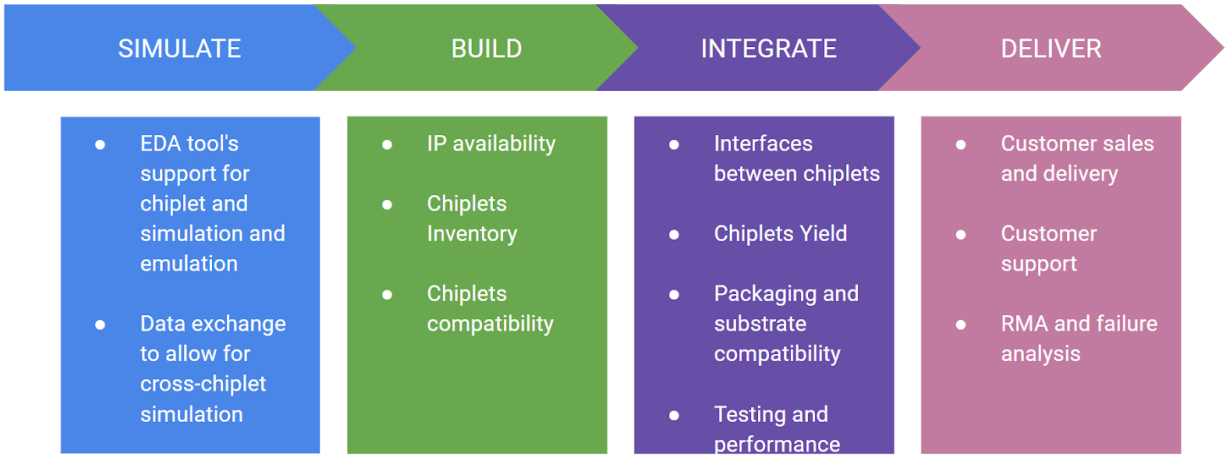


Figure 2: Delivery of a chiplet requires multi-vendor coordination across several activities

Major Business Workgroup activities for 2021

Open Compute Project (OCP) workstream Open Domain Specific Architecture (ODSA) organized the first ODSA Chiplet Business Workshop on March 12th, 2021. The goal of the workshop was to address many of the business-related questions the Community is facing today. Industry experts from 22 companies and government agencies in the Chiplet value chain shared their views during 24 presentations and 3 panels, first of a kind in the industry on Chiplet addressing business challenges with the Chiplet open ecosystem.

The workshop accomplished one of its core objectives, which was to start a dialogue within the industry around the challenges and gaps which exist on the business side that have to be addressed for open chiplets to become a reality. As a result, the workshop attracted significant industry attention with more than 400 attendees from 80 companies, the largest ODSA workshop or meeting ever and one of the most heavily attended workshops in the history of OCP! During this interactive workshop, industry experts identified and debated about value chain gaps that exist in the ecosystem today under the framework of Design, Build and Deliver to the End Users.

Some of the Industry experts who shared their views on how to enable the OCP Chiplet marketplace include:

- End-users such as Alibaba, Facebook, Google, Microsoft, and Foundry/IDM (Samsung & Intel)
- System Integrators such as Broadcom, Marvell, Microchip, Nvidia, Intel
- Packaging and Test houses such as ASE, JCET and Teradyne
- IP/EDA companies such as ARM, Cadence and Synopsys

Participants heard business challenges and potential solutions to enable Chiplet SOCs within an open ecosystem.

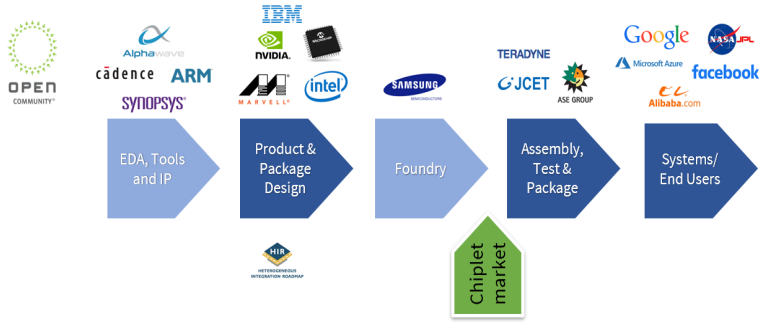


Figure 3: Chiplet value chain participating in the ODSA business workgroup activities

Post workshop, chiplet business workgroup brainstormed with industry partners and identified top 10 business challenges. A framework was developed to identify and gain consensus on potential chiplet business problems across the community. The following areas were identified by the team for focus in 2H 2021:

A. Open chiplet cost model
B. Known good die (KGD) financial model
C. Chiplet Thermal model

Open chiplet cost model

This topic was the #1 area of concern post the Business Working Group workshop held in April 2021. The goal of this task was to create a cost model for a disaggregated SoC and understand the implications of technology, process, assembly, NRE and other factors on the unit and total cost of the product. There has been a lot of collaboration between Foundry, OSATs, Silicon Vendors, IP vendors and end users to make this work possible.

Phase 1 of the work is a spreadsheet based model taking multiple factors into consideration and providing a summary of the comparative costs of a single SoC versus a disaggregated SoC. This model was presented at the OCP Summit in Nov 2021. There has been considerable interest from industry and we hope to develop this model such that it is useful to all the players of the chip making ecosystem.

Utilizing the model, we present two examples which go through two different cases for looking at chiplet cost models. Neither case considers logic self-repair through redundancy.

Example 1: Die disaggregation

Starting with a single 5nm SoC and running through the cost model to understand the implications of disaggregating it into two 5nm chiplets on the same substrate.

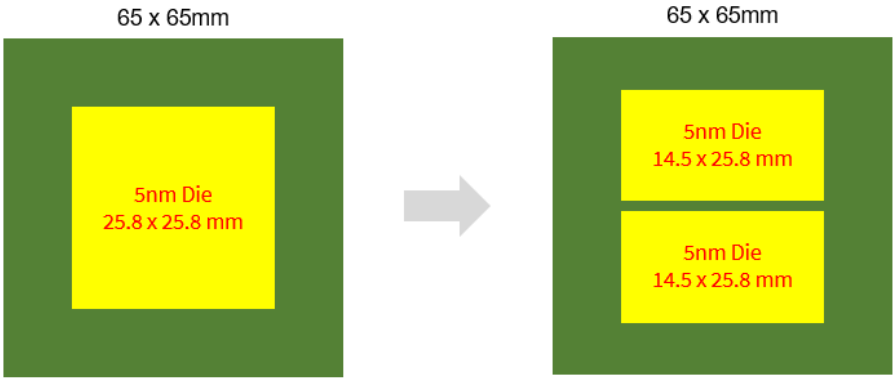


Figure 4: Monolithic to chiplet based solution

Unit cost shows that the chiplet based option is 32% less expensive than the monolithic option. This is despite the extra cost of the die-to-die IP for the chiplet version. The charts below show the unit cost over time (5 years) with forecasted volumes.

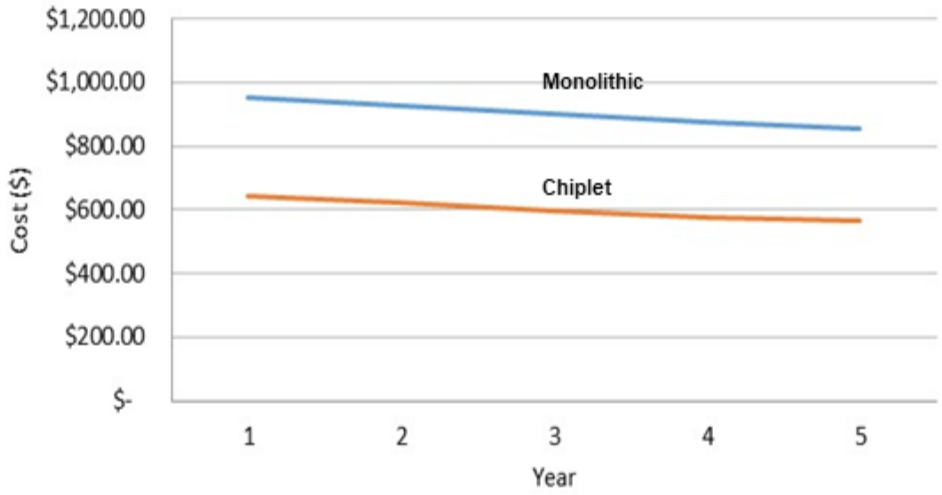
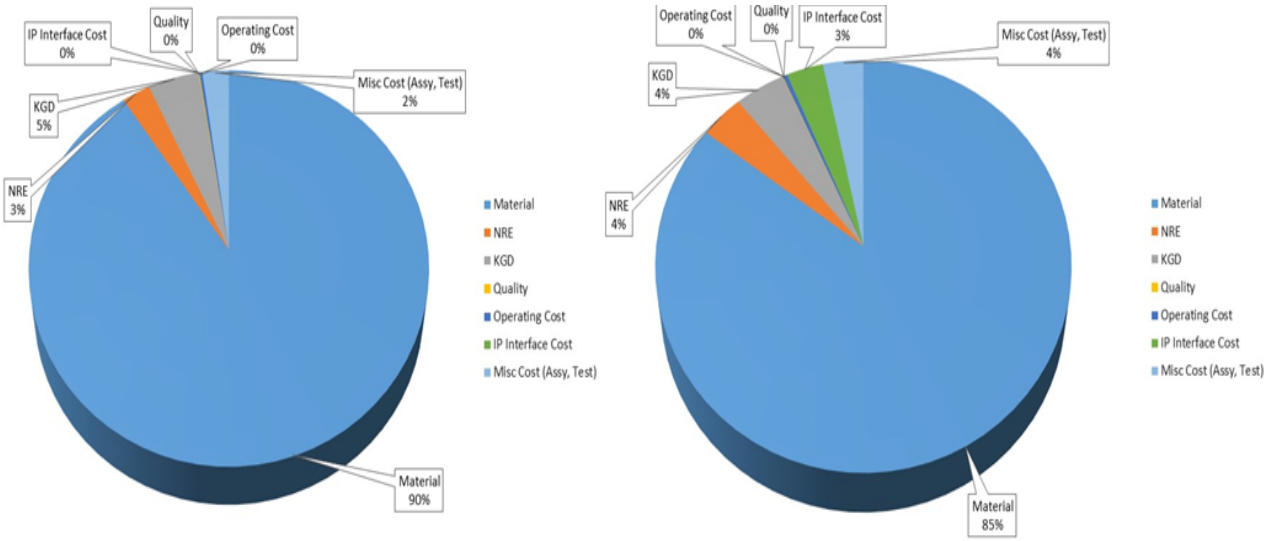


Figure 5: Unit cost comparison: Monolithic versus chiplet based

Figure 6 outlines the various cost contributions that make up the total cost of each solution. The rank order of the cost contributor for both example approaches is:

Material > KGD > NRE > Misc Cost > IP Interface > Operating Cost > Quality



Monolithic Chiplet based

Figure 6: Total cost breakdown: Monolithic versus chiplet based

The ordering of these contributors may change depending on the use case and supply chain. The charts below show the total cost over 5 years with forecasted volumes. The impact of shipment volume is shown in the figure below. As the figure shows, the higher the shipment volume, the more significant the benefits of chiplets.

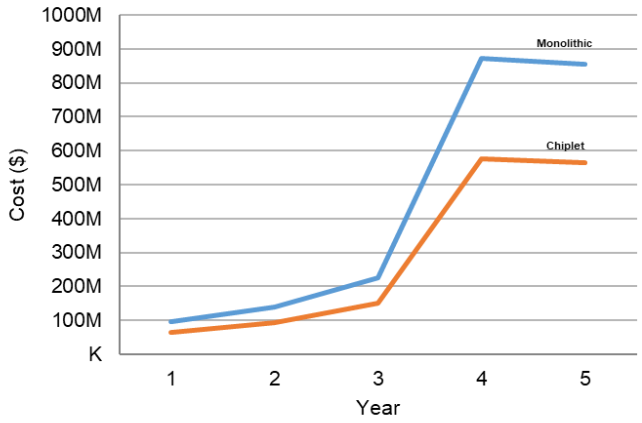
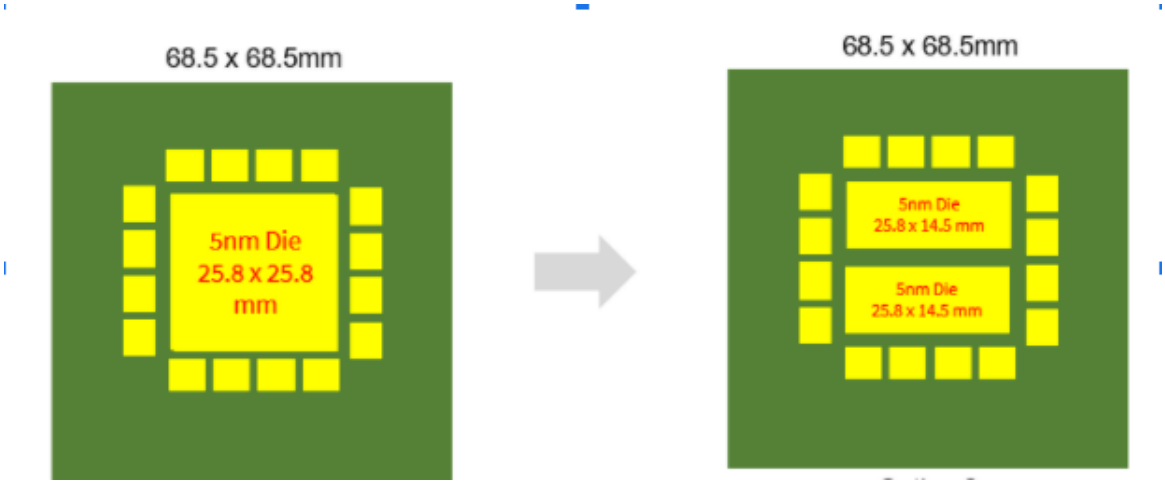


Figure 7: Total cost comparison: Monolithic versus Chiplet-based

Example 2: Die disaggregation with a system of chiplets

Starting with a large 5nm SoC surrounded by sixteen, 16nm chiplets, the cost model is used to see if disaggregating the large 5nm SoC into 2 smaller 5nm die while keeping the surrounding chiplets as it would provide a more cost effective solution



Monolithic with chiplets Chiplets with Chiplets

Figure 8: Disaggregation of the large 5nm die while keeping the 16nm chiplets

Unit cost, as shown in Figure 9, and total cost of the two choices shows that going with the disaggregated solution for the 5nm die will result in a 40% cost savings over the monolithic 5nm die.

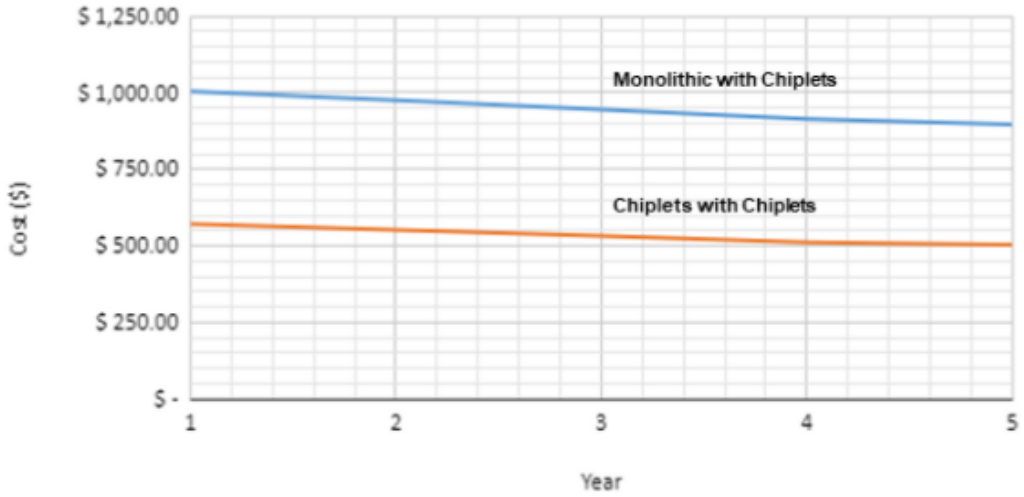


Figure 9: Unit costs of the two options shown in figure 8

The cost model also provides the percentage contribution of the various factors that make up the total cost of manufacturing the chip for the monolithic 5nm chip surrounded by 16nm chiplets. The pie chart shows at a glance the largest cost contributors.

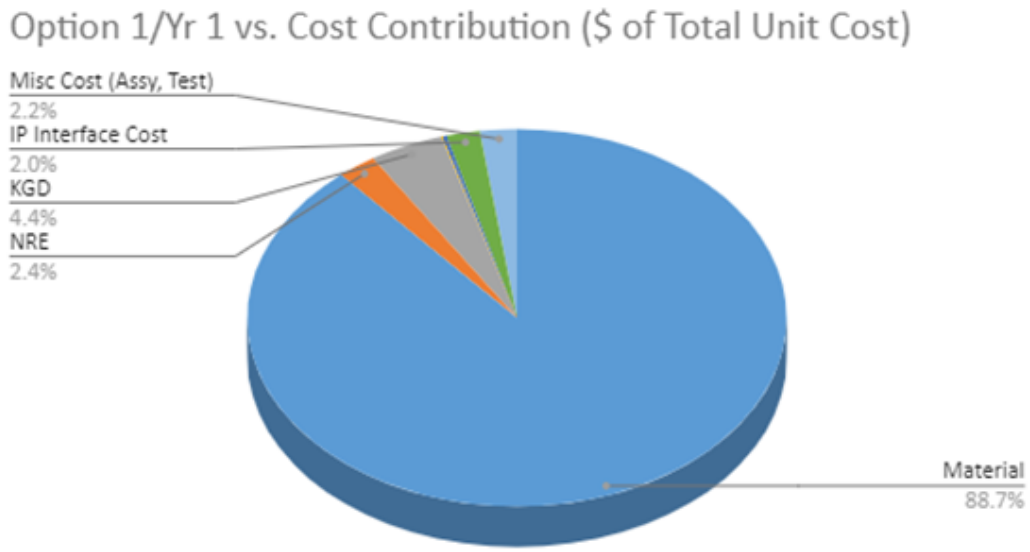


Figure 10: Pie chart showing the contributions of the various factors in the chip supply chain

Phase 2 of this project is planned to be more automated using Python scripts and also allow the user to perform sensitivity analysis on certain parameters which have uncertainties. This will allow for more fine grained analysis and provide data to the user to enable a very justifiable selection of whether to build monolithic or disaggregate to chiplets.

If anyone needs access to the model, please contact the editor of this white paper. We will publish model 2H-2022 on Github (https://github.com/opencomputeproject/).

KGD Financial Model

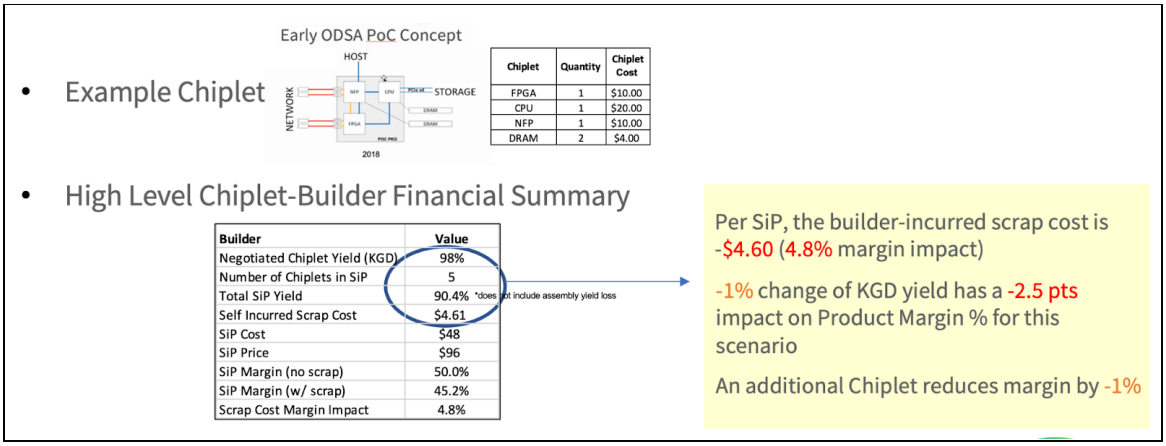
Another topic of interest that emerged from the Business Working Group workshop held in April 2021 was to understand the impact of KGD in a heterogeneous chiplet marketplace.

Sensitivity analysis was performed for the KGD taking into account chiplet-based-SiP Builder (Figure 11) and chiplet Provider (Figure 12). The model looked at 3 different Chiplet-based-SiP designs that vary between number of chiplets, cost/price/margins per chiplet, and yields.

Figure 11 below shows a simplified financial summary of a SiP builder who purchases chiplets. The example is a SiP with 5 purchased chiplets that each have a KGD of 98%. The SiP will have a yield of 90% (not including assembly yield loss) and a scrap cost of $4.60 (or -4.8 pts of product margin loss). For every 1 point of decrease of KGD per chiplet would cause a 2.5 point decrease in standard product margin.

Figure 12 below shows a simplified financial summary of a Chiplet Provider who sells chiplets to a SiP builder. This example assumes that the Chiplet Provider is financially responsible to the SiP builder (under a contractual obligation) for the total scrap cost of SiP if KGD drops below a pre-negotiated threshold. As noted, a -2 point decrease in KGD relative to the negotiated rate has a significant impact on the Chiplet provider's business. In the case of the DRAM chiplet provider there is a loss of -24 pts of standard product margins.

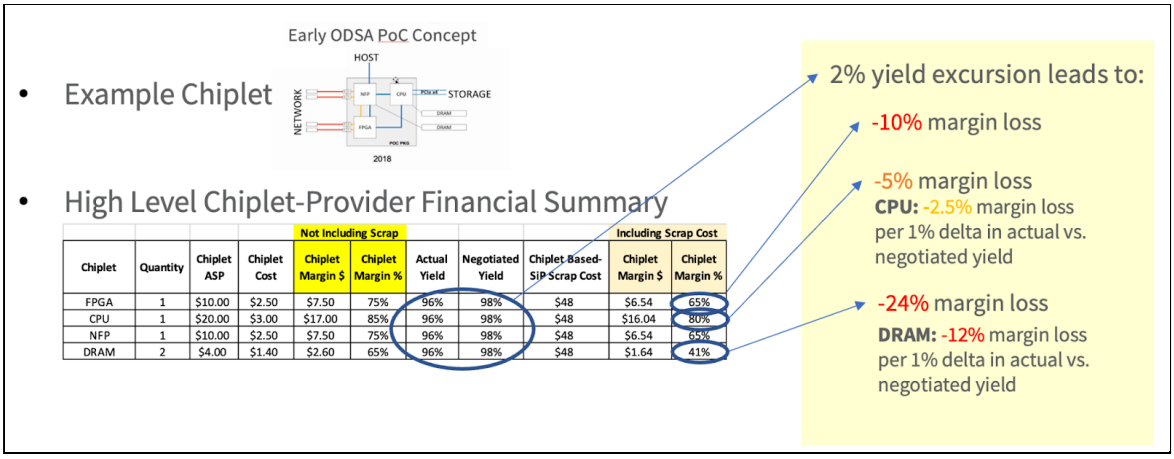
The summary is that both the SiP Builder and Chiplet Provider business model are very sensitive to any variations of KGD and in order for a healthy heterogeneous chiplet marketplace to exist KGD needs to be as high and predictable as possible. The presentation also explores the challenges and possible solutions of achieving a high and predictable KGD.



Chiplet	Quantity	Chiplet Cost
FPGA	1	$10.00
CPU	1	$20.00
NFP	1	$10.00
DRAM	2	$4.00

Builder	Value
Negotiated Chiplet Yield (KGD)	98%
Number of Chiplets in SiP	5
Total SiP Yield	90.4% *does not include assembly yield loss
Self Incurred Scrap Cost	$4.61
SiP Cost	$48
SiP Price	$96
SiP Margin (no scrap)	50.0%
SiP Margin (w/ scrap)	45.2%
Scrap Cost Margin Impact	4.8%

Figure 11: KGD sensitivity to chiplet SiP Builder



				Not Including Scrap					Including Scrap Cost	
Chiplet	Quantity	Chiplet ASP	Chiplet Cost	Chiplet Margin $	Chiplet Margin %	Actual Yield	Negotiated Yield	Chiplet Based-SiP Scrap Cost	Chiplet Margin $	Chiplet Margin %
FPGA	1	$10.00	$2.50	$7.50	75%	96%	98%	$48	$6.54	65%
CPU	1	$20.00	$3.00	$17.00	85%	96%	98%	$48	$16.04	80%
NFP	1	$10.00	$2.50	$7.50	75%	96%	98%	$48	$6.54	65%
DRAM	2	$4.00	$1.40	$2.60	65%	96%	98%	$48	$1.64	41%

Figure 12: KGD sensitivity to chiplet provider

Chiplet Thermal model

A thermal comparison study has been conducted to compare the thermal performance among various package options, including monolithic die (option 0), MCM (option 1), 2.5D (option 2) and 3D (option 3) (Figure 13). The study selects a logic die with 2 SRAM dies as an example, and assumes the same package size and TDP (logic die at 150W and SRAM die at 32W each). The package design consideration is based on the best practice and engineering experience. To maintain an apple-to-apple comparison, the platform boundary conditions such as ambient temperature, inlet airflow, and heatsink design are kept the same.

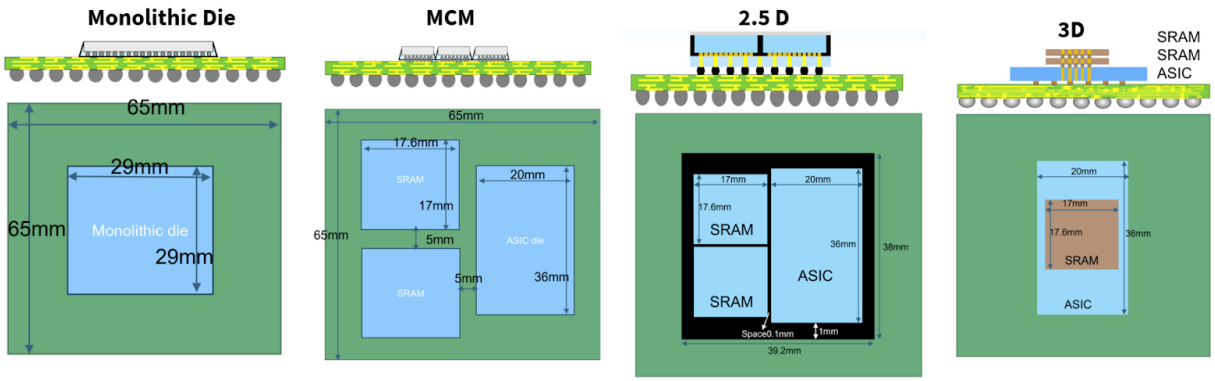


Figure 13 Different packaging technologies use cases for the thermal modeling

Figure 14 below shows the results from the study. According to the results, there are four observations:

1. Thermal density drives the junction/case temperature referring to option 0 vs. option 1 as option 0 (monolithic die) normally has a smaller die size.

2. 3D stack has the biggest challenge to cool as the thermal path becomes much longer to the ambient fluid and heat sources are stacked together comparing to the 2D planer design (refer to option3)

3. Tcase might not always be a good indicator to reflect the junction temperature as comparing option 1 and option 2, both with similar junction temperature but the thermal behavior on the case temperature is very different.

4. Die placement plays a significant role in the package thermal optimization. For example, SRAM is usually more sensitive to the temperature compared to the logic die while placing them together creates a heat crosstalk challenge.

	Option 0	Option 1	Option 2	Option 3
Package Design	**Monolithic**	**MCM**	**2.5D Interposer**	**3D**
Ratio of die to package footprint	20%	32%	32% (die are placed closer to each other)	17%
Tcase, center [C]	82.9	73.4	78.3	89.6
Tj, ASIC [C]	**89.1**	**84.3**	**84.5**	**99**
Tj, SRAM1 [C]	-	73.4	75.8	98.8
Tj, SRAM2 [C]	-	72	74.7	98.5

Figure 14 Thermal modeling result

Accomplishments for chiplet activities for 2021

ODSA End User Work Group consisting of Google, Microsoft, Facebook and Alibaba have come together to put a test chip for BOW in 2022. Over the past year there are the following key achievements which the group has achieved via joining forces:-

1. ODSA has clearly developed momentum on building a standardized die to die interface that is open and aligned the industry behind an open interface not dominated by one vendor
2. Build momentum behind BOW test chip development effort via producing BOW Test Chip proposal which now is being considered by most of the companies to identify activities and areas they will support for BOW test chip development in 2022.
3. End User WG has committed to funding the BOW test chip and they are also influencing other industry partners to support the effort with funding.
4. Influencing industry players such as Marvel to develop and support BOW interface. Marvel has publicly announced support for the BOW interface.
5. Align TSMC and Samsung foundry to support all BOW related efforts within OCP ODSA. This is the first time a foundry like TSMC is agreeing to support development of an open

die to die interface and support all companies within OCP ODSA community by providing their PDK kit for 5nm node for realization of a BOW test chip.

Next steps to drive the ecosystem

The ODSA work group will be contributing a final version of the BoW spec into OCP in early 2022. After this progress, we plan to work on defining and releasing BOW channel models via OCP to allow for BOW IP creation across different process nodes. The goal for 2022 is to expand support for BOW across the major players on silicon design beyond Marvel and Broadcom who have already announced plans.

2022 will also be the year when ODSA would like to engage IP vendors as Synopsys and Cadence to make BOW IP generally available via their IP library. ODSA workstream in OCP is driving open, standard chiplet marketplace and business and end user workgroups are driving the effort with other end users to enable this. 2022 looks exciting and an industry defining year to realize an open ecosystem of chiplets.

License

As a contributor to this document, all members represent that they have the authority to grant the rights and licenses herein. They further represent and warrant that the Materials do not and will not violate the copyrights or misappropriate the trade secret rights of any third party, including without limitation rights in intellectual property. The contributor(s) also represent that, to the extent the Materials include materials protected by copyright or trade secret rights that are owned or created by any third-party, they have obtained permission for its use consistent with the foregoing. They will provide OCP evidence of such permission upon OCP's request. This document and any "Materials" are published on the respective project's wiki page and are open to the public in accordance with OCP's Bylaws and IP Policy. This can be found at http://www.opencompute.org/participate/legal-documents/.

If you have any questions please contact OCP.

Footer:

OCP will need a signed Copyright License Agreement so we can publish this on our website. Please insure you have signed this agreement.

About Open Compute Foundation

The Open Compute Project Foundation is a 501(c)(6) organization which was founded in 2011 by Facebook, Intel, and Rackspace. Our mission is to apply the benefits of open source to hardware and rapidly increase the pace of innovation in, near and around the data center and beyond. The Open Compute Project (OCP) is a collaborative community focused on redesigning hardware technology to efficiently support the growing demands on compute infrastructure. For more information about OCP, please visit us at http://www.opencompute.org