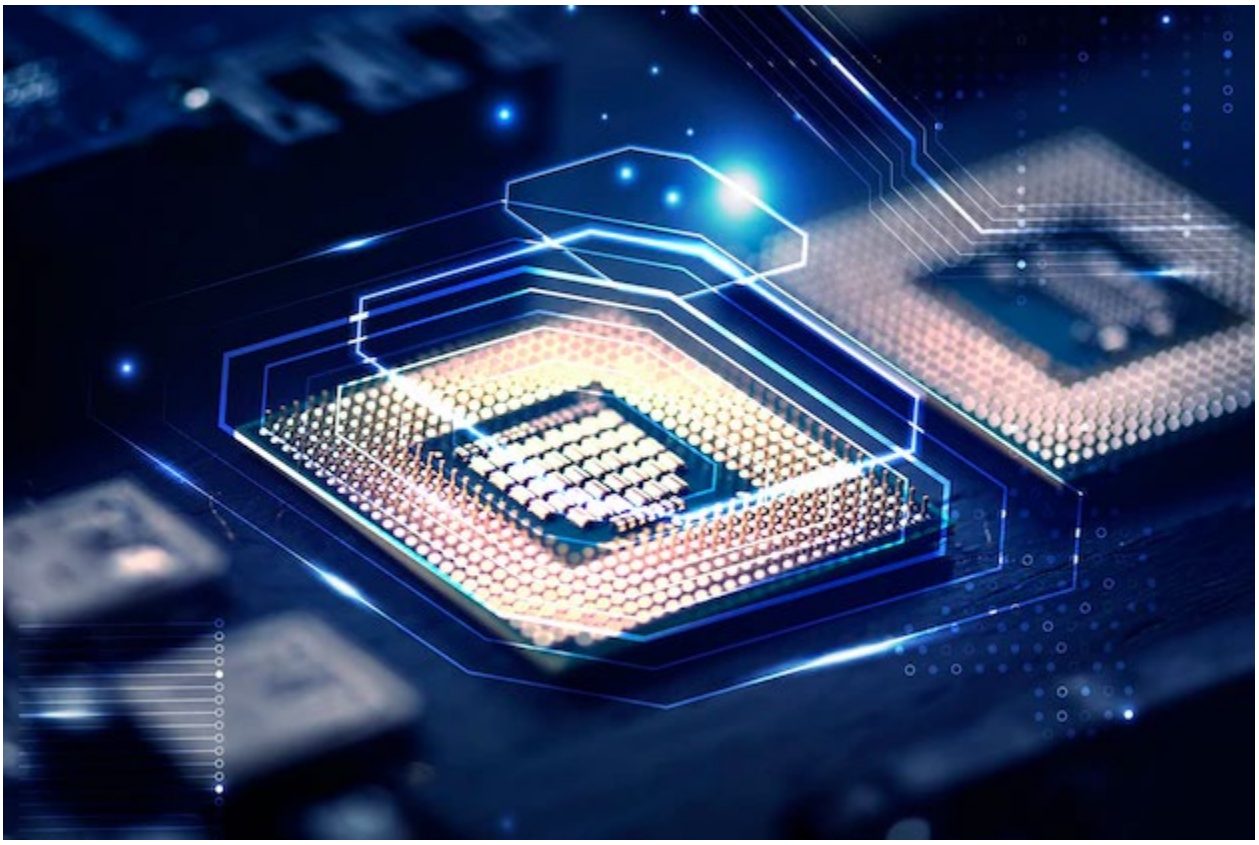
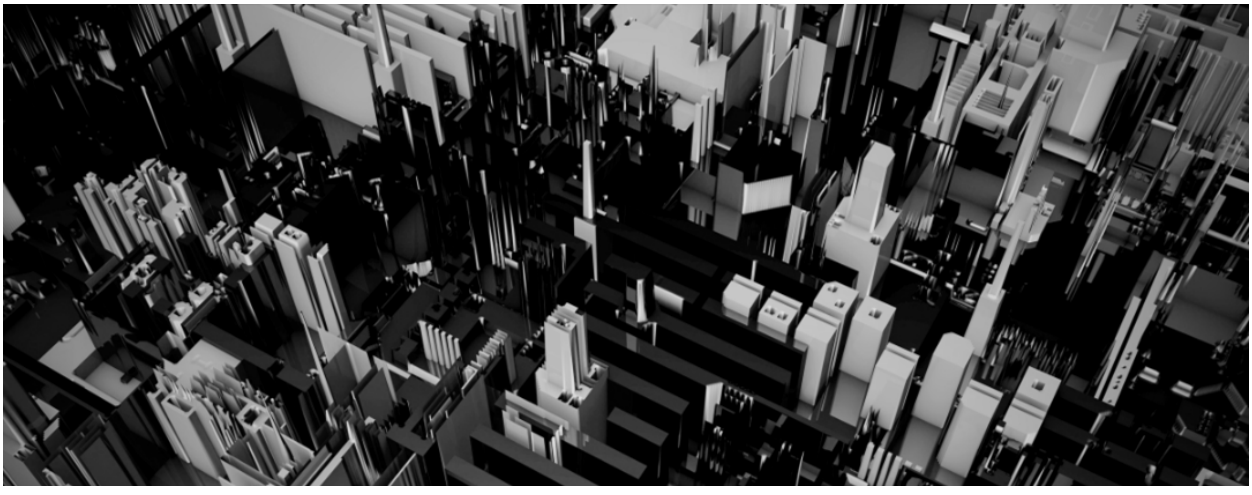
WHITEPAPER FEBRUARY 2023

HETEROGENEOUS INTEGRATION - CHIPLETS

GLOBAL SEMICONDUCTOR ALLIANCE INTELLECTUAL PROPERTY INTEREST GROUP



CONTENTS



1. INTRODUCTION

Chiplets are a hot topic in the semiconductor industry, and to many represent a paradigm change for chip designers and chip consumers alike. Rather than a traditional single piece of monolithic silicon with many functions and features, a chiplet contains one or a very limited set of functions and features. Multiple chiplets are then assembled into a MCM (multi-chiplet module).

Chiplets can offer multiple advantages over monolithic silicon, including:

- Chiplets can be designed in the optimum process node (performance, power, cost) for the particular function and/or feature
- Size and yield – as monolithic silicon reaches reticle limitations, chiplets are much easier and cheaper to manufacture
- The ability to manufacture with and integrate diverse and potentially incompatible semiconductor materials, such as GaN, SiC, Ga_2O_3, and others
- Chiplets can be re-used over multiple products and projects
- Enables choice: allows chipmakers/OEMs to select the right combination of chiplets for their needs, rather than be forced to choose a potentially less-optimum monolithic chip
- Lower NRE (non-recurring engineering) costs for chiplet designs, potentially enabling smaller volume applications
- Lowers the entry barrier to designing and selling silicon
- Allows semiconductor IP to be sold in silicon form
- Quickens time to market

Homogenous chiplet MCMs, or MCMs containing chiplets from only a single manufacturer, have been mass-market deployed – an example would be the AMD Ryzen Zen 3 processors. Homogenous chiplets share many of the advantages mentioned above – however, as they remain the domain of a single chip supplier, they don't offer all the advantages (and issues) that *heterogenous* chiplets do. A heterogenous chiplet MCM contains chiplets from multiple manufacturers, often made in different facilities on different process nodes. Heterogenous chiplets might include a combination of CPU, GPU, NPU, FPGA, and/or special purpose chiplets. Chip consumers (OEMs, brands, etc.…) can assemble the optimum selection of chiplets into a heterogenous chiplet MCM for their needs, and chiplet suppliers can manufacture their chiplets in the ideal location/node for price, performance, and area.

While heterogenous chiplets seem to have multiple advantages over traditional monolithic silicon and even homogenous chiplets, they still have not been mass-market deployed. This white paper will explore the commercial, interface, packaging, and security issues heterogenous chiplets face, along with recommendations for the industry's successful deployment of heterogenous chiplets.

2. COMMERCIAL CHALLENGES

With monolithic silicon and homogenous chiplets, there are clear lines of ownership and responsibility when business and technical issues arise. However, how are these issues addressed with heterogenous chiplets?

Business

Monolithic silicon has always been 'owned' by a single party (the chipmaker). The same is true for homogenous chiplets. When a customer wishes to purchase a chip or monolithic MCM, they negotiate a contract with the chipmaker (or agent) and procure the silicon.

However, what happens with a heterogenous chiplet MCM? As chiplets will likely be sold on the open market as raw silicon, rather than assembled in MCM form – what happens if an OEM specifies which chiplets are to be part

of an MCM, and contracts the assembly and test of the MCM to a third party? Does the OEM own that design, or does the third party? Who warrants the design and fitness? Who owns the supply chain? Certainly, this can be solved if a SoC (System on Chip) maker is taking ownership of the MCM design, sourcing chiplets from various manufacturers, and sells the MCM as a finished product

Many potential business challenges arise with heterogenous chiplets which likely must be addressed contractually prior to any MCM design start.

Technical

Business issues aside, what happens if there are technical issues with heterogenous chiplet MCMs? With monolithic silicon and homogenous chiplets, the chip maker has always been the responsible party. What happens with heterogenous MCMs?

Given that there will be multiple suppliers of chiplets within the MCM, who takes ownership in root cause analysis and fixes? And how does the market prevent echoes of the latter 1990s/early 2000s "WinTel" technical support boondoggle (CPU maker points finger at OS maker, who then points finger at computer maker, who in turn points finger at CPU maker)?

3. INTERFACE CHALLENGES

Design engineers are all familiar with the C2C (chip-to-chip) communications that are used between various components with electronic systems today -- common examples include PCIe, LVDS, and CXL. Very similar to C2C communications, a D2D (die-to-die) interface is a functional block that provides the data interface between two silicon dies (chiplets) within a package. The D2D creates a reliable link and is generally described as the PHY and the controller. The D2D interface is a key enabler in the move from monolithic SoC to multi-die SoC in the same package. There have been several recent developments with D2D communications that are used between chiplets.

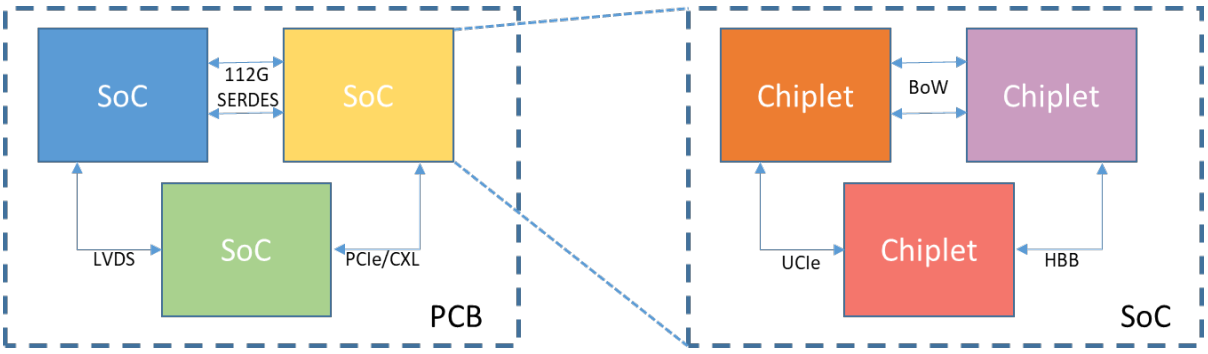


Figure 1: C2C versus D2D

The comparison between C2C and D2D is very similar. Where, in the case of C2C communications, the channel is usually over a PCB board. The PCB can have a variety of different materials and thus different channel characteristics. Similarly, for D2D communications, there are a variety of channels or substrates that can be used to communicate, generally an organic substrate (2D) or an interposer (2.5D). This will impact the particular D2D interface IP since the channel will be different and will likely require modifications to the interface, for example, the bump pitch or the signal loss.

D2D	C2C
Substrate	PCB
Short reach (mm)	Long reach (cm)
Small low-power drivers	High current drivers
Raw/stream	Protocol or packets

Table 1: D2D vs C2C Comparison

These D2D communications can be serial or parallel, be source synchronous or clock forwarded, single-ended or differential. The Serial D2D interface is fundamentally a SerDes (serializer/deserializer), which includes the parallel-to-serial (serial-to-parallel) conversion, impedance matching, CDR (clock data recovery), or clock forwarding. The greatest value of a serial interface is the minimization of the number of IOs required. Parallel interfaces basically take a large number of IO pins (100' to 1000's) that drive a single-ended signal from die to die. The data rates can be from a few gigabits to up to 40+ Gbps and the distance is generally up to 10mm.

4. D2D INTERFACE TYPES

D2D interfaces can be characterized in a variety of ways-- serial vs. parallel or channel/package type or throughput. In this white paper, we will take a simple approach and discuss this in 2 ways a) Serial vs Parallel and b) Proprietary standards and an Open standard.

Serial vs. Parallel

"Serial" D2D is essentially a SerDes, while "parallel" D2D is basically a point-to-point connection. While both can be used for D2D interfaces, the strengths, and costs of each have generally dictated which is used. Serial has been predominately used in high-end networking and optical chiplet applications, while parallel has been more generally adopted. Both are characterized by the following below.

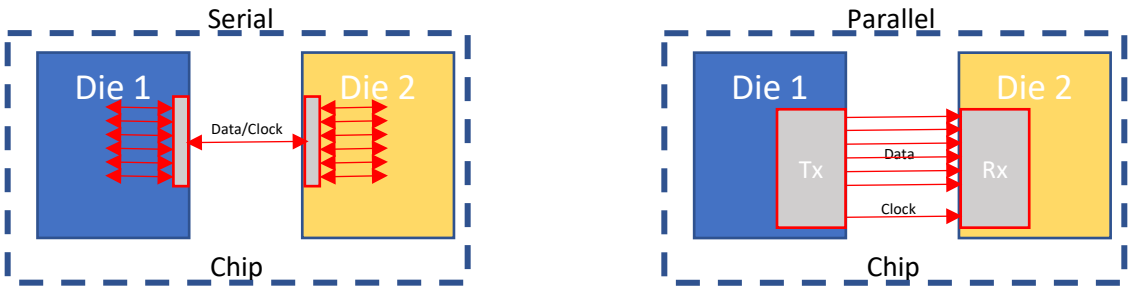


Figure 2: Serial and Parallel Interfaces

Serial	Parallel
High data/line rate	Lower data/line rate
Lower IO count	High IO count
Higher power	Lower Power
Higher latency	Lower latency
Organic substrate	Organic or Interposer

Table 2: Serial vs Parallel

Proprietary vs. Open D2D Standards

Specific to chiplets, the industry to date has utilized many proprietary interfaces as part of homogenous chiplet MCM deployments. The homogenous nature of these MCMs has allowed the use of proprietary D2D interfaces. Table 3 illustrates some, but not all, of the proprietary, interconnects proposed or used on homogenous chiplets.

	HBB-M/HBB-I	UltraLink	G-Link 1/2/3	HBI	M-Link
Org:	Samsung	Cadence	GUC	Synopsys	Mediatek
Type:	Parallel	Parallel	parallel	parallel	parallel
Line Rate	32G	40G	8/16/32G	8G	20G
Clocking:	Forwarding	Forwarding	Forwarding	Forwarding	forwarding
Signaling	DDR	DDR	DDR	DDR	DDR
Reach:	3mm	50mm	.5	5mm	1mm
Channel:	Organic Interposer	Organic	interposer	organic	interposer
Package:	2D, 2.5D	2D	2.5D	2D	2.5D
Latency: (ns)	~3ns	4ns tx, 5.2ns Rx			
Power: (pJ/bit)	.6/.3	1.9	.5/.3/.26 pJ	.55	.46

Table 3: Survey of Proprietary D2D

With the use of proprietary interfaces in homogenous chiplets, chip designers have design control over both sides of any chiplet link.

However, as the market moves to heterogenous chiplets, the market must move to an open standards. These allow for different chiplet makers to build discrete chiplets which interoperate seamlessly. Several open standards have been developed to usher in the age of heterogenous chiplets. Examples of some of the open standards developed to date are shown in Table 4.

	BoW: Bunch of Wires	OpenHBI	AIB	UCIe	XSR
Org:	OCP	OCP	Chips Alliance	UCIe	OIF
Type:	parallel	parallel	Parallel	parallel	Serial
Line Rate:	8/16G	8/16G	6G	16/32G	112/224G
Clocking:	forwarding	forwarding	forwarding	Forwarding	recovered
Signaling:	DDR	DDR	DDR	DDR	differential
Reach:	5mm (unterminated) 50mm (terminated)	4mm	<10mm	2mm,10-25mm	~50mm (terminated)
Channel:	Organic Interposer	Interposer Organic	EMIB/interposer Organic	Interposer Organic	Organic
Package:	2D, 2.5D	2D, 2.5D, 3D	Bridge	2D, 2.5D, Bridge	2D
Latency:	3ns			2ns	
Beachfront:	1.78 Tbps/mm	3.34	1.64	2.24 /13.17 I/M	
Power:	.7 .5 Syn	.5 .4-.5 Syn	1.2 .5 syn	.3 to 1.25	~1

Table 4: Survey of Open D2D Standards

How to Select the "Right" D2D Interface?

When designing chiplets or multi-die SoCs, system designers are faced with many design choices and trade-offs. Perhaps the most difficult of these is in selecting the "right" D2D interface. But is there such a thing as the "right" interface?

For chip designers and system architects, the first step in determining the chosen D2D interface may come from the application or the functions of chiplets. This is where the priorities, trade-offs, and considerations are made. Here are some key trade-offs and considerations:

- **Bandwidth:** The D2D interface must be able to support the required bandwidth of the application or system that it is talking to. This can be achieved by having fewer PHY lanes at a very high line rate or having a greater number of PHY lanes to increase throughput. Bandwidth requirements are generally determined by the chiplet usage.
- **Beachfront:** Refers to the throughput at the edge of a die, or the density of I/O at the edge. This can potentially determine the size of a chiplet.
- **Latency:** This can have an impact depending upon architecture. In the case of partitioning a unified architecture or memory architecture that requires uniformity, designers will want to prevent non-unified delays or access times to data or memory.
- **Channel or Loss:** The channel or loss can determine the kind of package used and how far or long the channel can be. These criteria can also determine the robustness or SI/PI of the interface a designer might select.
- **Power/Energy:** Power is also an issue in chiplets, and a likely design goal is to achieve power targets on par with monolithic levels. D2D interfaces utilize short-reach and low-loss channels, but there is still a significant difference in the various implementations and number of lanes used.
- **Package Type:** If a given package is already determined, it may also dictate which D2D interfaces are available to the designers. Some D2D interfaces are more optimized for MCM-type packages and will utilize C-bumps whereas interposer-based packages will utilize micro-bump architecture.
- **Cost:** Continually a leading design factor that must be considered. This can include but is not limited to IP, implementation, package, and testing costs.

However, one non-technical factor that needs to be considered is **market acceptance**. Just because a D2D interface may be superior does not mean it is necessarily the right choice. It is imperative that system architects examine what D2D interface open standards are present in their target markets, and make choices not only based on technical reasons but also based on market reasons.

Moore's law has driven the semiconductor industry for decades and the chiplet era is clearly a way to preserve the essence of Moore's Law. With that said, the D2D interface will become even more and more critical in the years to come. It will determine how innovative the next generation of chiplets can be and truly democratize multi-die SoCs.

5. PACKAGING INNOVATIONS FOR CHIPLETS

What is the role of packaging as we move into the Chiplets era? The industry is moving toward desegregated SOC models that bring together heterogeneous chiplets components within a single package. As such, the packaging community is responding by developing advanced packaging technology innovations to address urgent and increasingly complex industry needs.

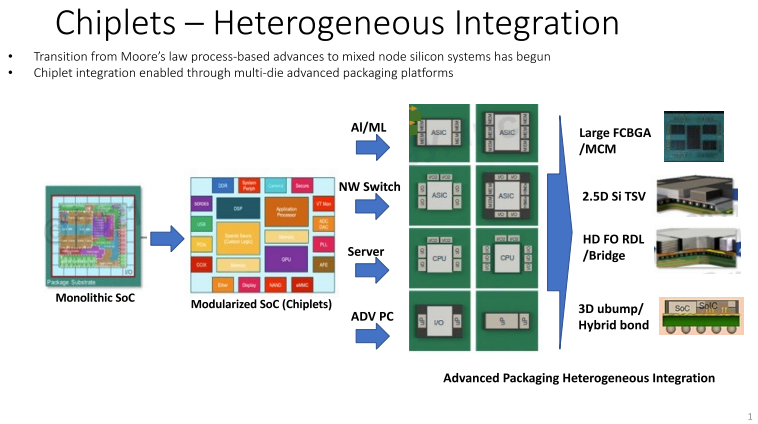


Figure 3: Chiplets – Heterogeneous Integration **Source: SEMICON Europa Nov 16, 2022**

Figure 3 illustrates the traditional monolithic System on Chip (SoC) die, desegregated by design into a modularized SoC (Chiplets) format, that is then integrated into a single package. Applications include artificial intelligence (AI) and machine learning (ML) network switches for high-performance servers within data centers and advanced PCs. The packaging platform for chiplet integration includes large flip chip BGA (multichip module) on advanced build-up substrate, 2.5D (TSV) silicon interposer, high-density fanout RDl, silicon bridge, and 3D micro-bump stacking, and hybrid bond. Market applications, as well as packaging integration technologies, will continue to evolve for performance and efficiency, from system and packaging architecture to co-design and co-design tools, assembly, test, and manufacturing processes.

The electronics and semiconductor communities are highly innovative, and the packaging community is no exception. The fanout reconstitution process, materials, and equipment were originally developed for low-end WLCSP applications. Some years ago, packaging engineers used the materials and process equipment to integrate two die into a reconstituted format for flip chip assembly onto an organic substrate, shown in figure 4. This innovative concept, the so-called high-density Fan Out RDL, has become an essential advanced packaging platform across our industry for multiple die integration. Together with TSV silicon interposer and bridge technologies, Fan Out RDL innovations will form the advanced packaging technology base for the chiplet revolution as we move full steam ahead.

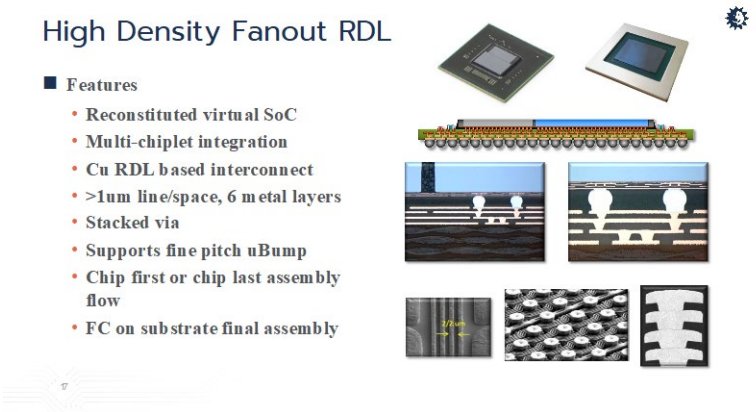


Figure 4. High Density Fanout RDL application **Source SEMICON Europa Nov 16, 2022**

While the chiplet revolution may have already started with the concept of desegregated and "modularized" SoC, innovations in advanced packaging have expanded this original concept to reach many more market applications. The design and integration of multiple die through advanced packaging into a single package is enabling the business of bringing electronics products to the marketplace in the most effective and efficient ways possible. An interesting article on this topic by Ed Sperling and Karen Heyman may be found in reference [2]

6. CHIPLET SECURITY CHALLENGES

For all the advantages chiplets offer, they have distinct drawbacks from a security point of view. Due to the nature of heterogenous MCMs, there is a particular vulnerability to malicious counterfeit systems snuck into the manufacturing flow of any one of the chiplets being integrated. In addition to counterfeits, since chiplets involve an expanded number of supply-chain entities, each with their own design teams and EDA tool flows, there is an increased opportunity for hardware trojans to be injected into one of the many chiplets, and therefore into the aggregate MCM. Finally, another risk emerges from the new attack surfaces, in the form of the chiplet-to-chiplet interfaces within the MCM, which are not present (or are much more difficult to electrically contact) in a monolithic chip. In this situation, a reworked MCM could be susceptible to a man-in-the-middle attack (MITM), which can be used to spy, sabotage communications, and/or corrupt data, as well as steal personal information.

Proving and Securing Authenticity

Authenticity is best achieved with cryptographic challenge/response authentication. This technique is not dissimilar to how online passwords are verified or how ATM withdrawals are authenticated with the account holder's PIN number. In general: the verifier issues a "challenge" (e.g., a random number) to the prover, who then calculates a cryptographic "response" involving the challenge and some pre-shared secret. In this way, the prover can securely and reliability demonstrate to the verifier that "it knows a secret" which only authentic components should know.

While the challenge/response protocol aspect is straightforward enough, the challenges to doing it well are significant. Specifically:

- How is secret data provisioned to the component (e.g., a chiplet) during manufacture, and how is this provisioning process itself protected from theft or misuse?

- How is secret authentication data stored and utilized within an MCM's chiplets – both the prover and the verifier chiplets – such that it is not easily recovered by an adversary?

Silicon-based Countermeasures

The first domain of authenticity resides with the silicon devices themselves. Specifically, how does the prover chiplet protect the secrets it's been given?

Generally speaking, public-key encryption techniques are the most straightforward approaches to measuring authenticity. The major drawback to these techniques is the computational effort it requires to perform a challenge/response exchange – symmetric encryption techniques are much lighter-weight computationally, but more difficult to deploy.

In the most straightforward public-key authentication system, the "prover" chiplet contains a secret private key, as well as the associated public key encapsulated within a signed certificate. During an authentication event, the verifying component (e.g., another chiplet containing the MCM's principal Root-of-Trust security processor) obtains the certificate from the prover chip, verifies the signature on the certificate, then challenges the prover device to prove that it has the secret key associated with the verified public key. For example, the verifier might deliver a 16-byte random number to the prover chip which takes that nonce as an input to a private-key signature operation. The prover chip then calculates and returns the signature value, which the verifier can then authenticate – if the prover chip completes the calculation correctly, it must know the secret private-key value.

Of course, while knowing the secret private-key value is an essential aspect of multifactor authentication, it shouldn't be the only one – counterfeit chips can (and do) contain entirely valid private key values stolen from other authentic chips. At least two multifactor components are recommended for improved security, so not just "something the chip knows" (e.g., the chip knows the secret key) but perhaps also "something the chip has" like a hardware-accelerated engine for performing the calculations. A hardware-accelerated authentic chip might complete the signature calculation in 10mS while an inauthentic chip with a stolen key might complete the same calculation in 50mS – a verifier could detect this as a second element of multifactor authentication.

As for how to best secure the secret data within a prover chip, the most critical aspect is to protect the calculations involving the data, not the data itself.

For example, one of the most tamper-resistant techniques for deriving secret key data is the use of a Physically Unclonable Function circuit (aka, a "PUF"). This is because, unlike NVM where the data exists in perpetuity, a true PUF circuit is a "mixed signal circuit" whose output value is inaccessible when the chip is powered off. Attacks against PUFs must therefore be staged while the chip – and other chip-level countermeasures – are powered on.

However, even with the most tamper-resistant PUF as the secret-key source, if the secret key value is cryptographically operated on without power-analysis countermeasures applied during that calculation, an adversary need only monitor power-supply consumption during execution of the challenge/response protocol to recover the secret key with relatively low effort.

Supply Chain Security & Provisioning

Ensuring the integrity of a chiplets supply chain is in some ways more challenging than insuring the integrity of the challenge/response algorithm, in that the protections are not as well known, nor as well standardized. Encryption technology does come into play, as do several other best practices. For example:

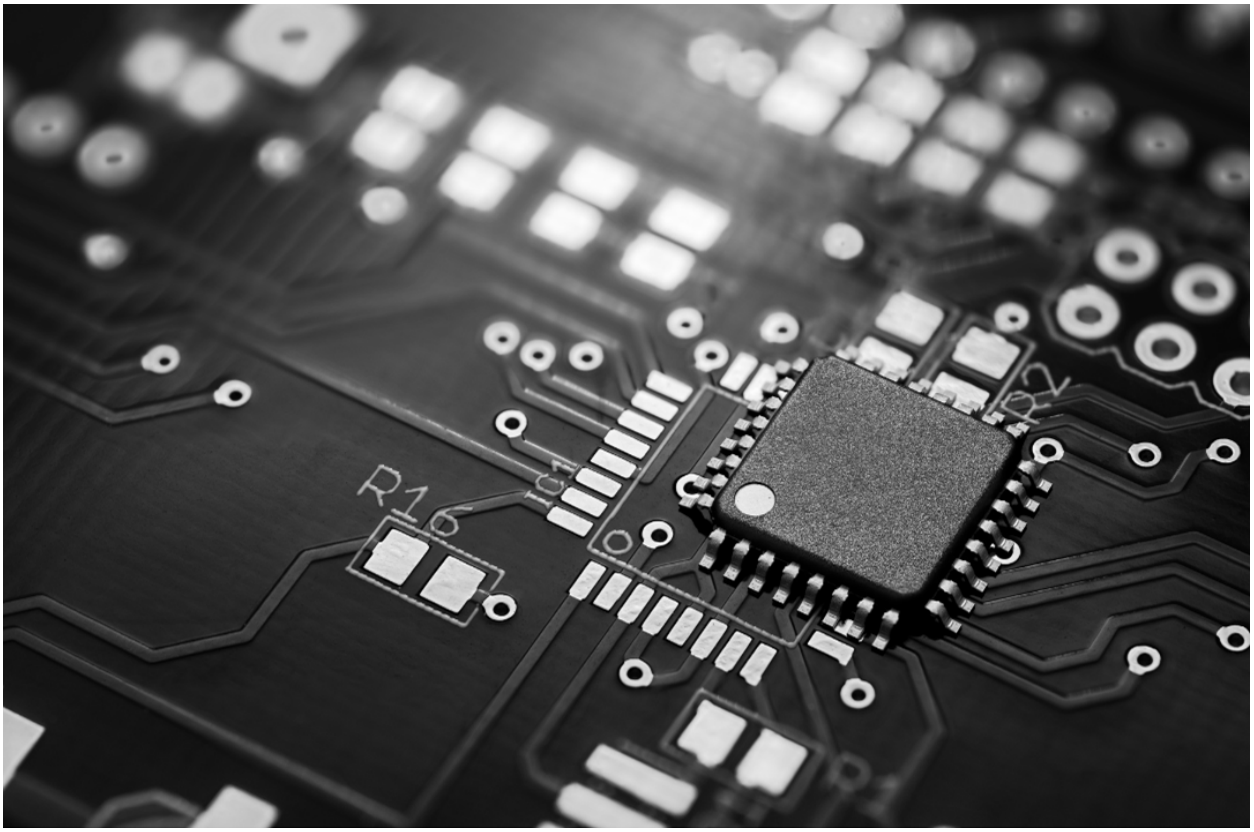
- Every chiplet in a given product line should be provisioned with a device-unique ID, and that ID should be included as part of any authentication certificate. Ideally, verifiers should be able to confirm the validity of an ID value – for example, it can be compared against an online block/allow list.

- During provisioning, the provisioning equipment should not only authenticate the chip (e.g., if a chiplet appears at final test asserting it corrected wafer-sorted, the provisioning equipment should verify the wafer sort provisioning contents before trusting the device), but the chip should authenticate the equipment (e.g., the provisioning equipment should prove it knows a netlist key before the chip unlocks itself for wafer-sort provisioning).

- All certificate-signing operations performed during manufacture should of course be performed within the security boundary of a tamper-resistant hardware security module. It is nearly impossible to distinguish authentic devices from counterfeit devices if an adversary can steal a certificate signing key.

- Auditing logs from various points of manufacture (wafer sort, final test, system attach, etc.) should be centralized so that strict auditing checks can be performed. For example, every device provisioned during wafer sort should appear at final test once and only once – anything else would indicate supply-chain theft (bad) or supply-chain injection (even worse).

- Lastly, and primarily in the context of chiplets, a PKI-assured technique for adding certificate chain-of-trust techniques to the authentication certificates should be considered. For example, chiplet vendors should deliver not only silicon to MCM integrators, but they should also deliver trusted certificates associated with their individual signing key used for their devices. The MCM integrator would then provision that collection of trusted certificates into the main verifier of the MCM (i.e., the chiplet containing the principal Root-of-Trust) so that the authenticity of each chiplet in the system can be reliably confirmed.



7. RECOMMENDATIONS

The value of heterogenous chiplets cannot be understated. The ability to create heterogenous MCMs featuring chiplets serving different functions, manufactured in the ideal method and with ideal materials, will usher in a new wave of semiconductor deployments and use cases.

However, for heterogenous chiplets to achieve mass market success, the semiconductor industry is going to have to align on the key challenges discussed in this white paper.

- The business and technical support owner of an MCM should be contractually established at the onset of a program.
- A very limited number of, if not a single, open D2D standard for chiplet interfaces must be agreed upon. A large number of competing D2D standards will only cause incompatibility and limit the effectiveness and deployment of heterogenous chiplets.
- The packaging community needs to respond by developing advanced packaging technology innovations to address urgent and increasingly complex industry needs for chiplets.
- To assure chiplet authenticity, security provisions must be standardized and put into place during chiplet design and followed through in production.

8. REFERENCES

1. "Advanced Packaging Creativity Enabling Silicon Systems" presentation by Yin Chang, SEMICON Europa, November 16, 2022

2. "The March Towards Chiplets", Ed Sperling and Karen Heyman, Semiconductor Engineering December 15, 2022. https://semiengineering.com/the-march-toward-chiplets/

9. CONTRIBUTORS

Source material for this white paper has been provided by:

Paul Karazuba (Editor)

Frankwell Lin

William Chen

Yoan Dupret

Scott Best

Volker Politz

Kevin Yee, Tony Luk