Independent market research and
competitive analysis of next-generation
business and technology solutions for
service providers and vendors

HEAVY READING
WHITE PAPER

Domain-Specific Accelerators: What They Are & Why They Should Matter to Cloud & Network Providers

A Heavy Reading white paper produced for Netronome

NETRONOME

AUTHOR: SIMON STANLEY, ANALYST AT LARGE, HEAVY READING

EXECUTIVE SUMMARY

The bandwidth demands on server processors have significantly outgrown the increases in central processing unit (CPU) performance over the past few years. This is already having an impact on cloud and network service providers, which are installing a growing number of servers and data centers to meet the demand from customers. Virtualization is key to delivering many of these services but also increases the workload on server CPUs. Server performance can be a critical factor in providing cost-effective cloud and networking services.

Domain-specific accelerators such as SmartNICs, and machine learning and inference coprocessors can offload processing from CPU cores, significantly increasing application performance and releasing additional CPU cores for other revenue earning workloads. To deliver the best benefits these accelerators should integrate best-of-breed components such as processors, hardware engines, memory and I/O peripherals.

Domain-specific accelerators are programmable application-specific integrated circuits (ASICs) usually implemented as a monolithic system-on-chip (SoC). The shift to smaller silicon geometries has enabled larger SoC devices with big caches and many cores but each new silicon technology generation requires huge investment and results in significantly higher development and die costs. Developing domain-specific accelerators for many applications is not commercially attractive due to the high development costs for each device.

Chiplets and multi-chip modules (MCM) with multi-dies on a single substrate provide an attractive alternative to monolithic designs with a single die. Chiplets have been widely used where an integrated device uses components developed using different silicon technology and are increasingly being used to provide modular solutions. Notable uses of chiplets have been High Bandwidth Memory (HBM), mobile phone processors and the AMD EPYC processors. The use of chiplets offers a cost-effective approach to domain-specific accelerators that should enable a wider range of solutions to become available.

Most chiplets so far have been used in a closed development environment, where all the chiplets and the integrated system have been defined by a single company or group working on a single design. An open development environment where chiplets from different companies have common interfaces and support a common protocol stack will enable the wider use of chiplets.

The Open Domain-Specific Architecture (ODSA) Workgroup was launched in October 2018 by seven companies to develop an open architecture and related specifications for developing chiplets. The group is working on a reference design and a complete protocol stack to support domain-specific accelerator development. This white paper is based on inputs from the ODSA Workgroup and individual member companies including Achronix, Avera Semiconductor, Aquantia, ESnet, Kandou, Netronome, NXP, SamTec, Sarcina and SiFive.

WHY USE OPEN DOMAIN-SPECIFIC ACCELERATORS?

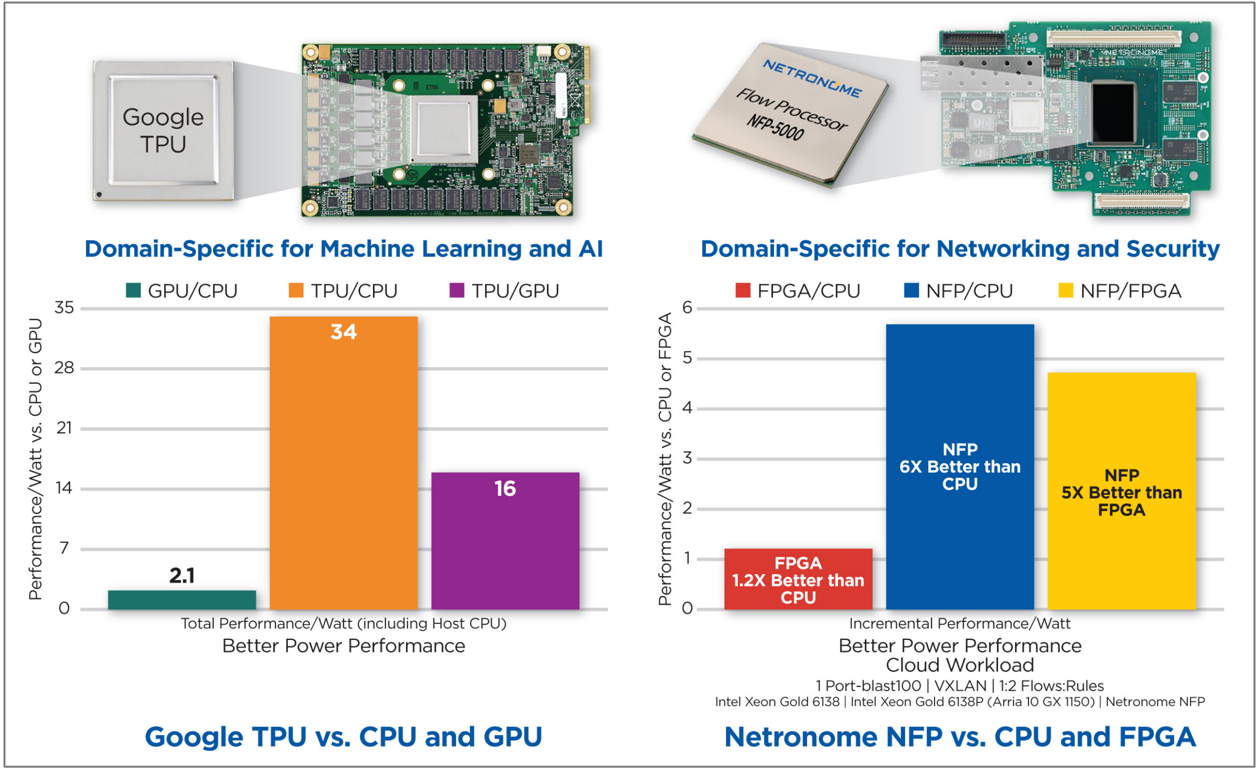
Key Market Areas

General-purpose CPUs alone cannot sustain the demands of server workloads in many key markets. The demise of Moore's Law (shrinking length of transistors and number of transistors bought per dollar) only exacerbates the situation. Silicon devices based on domain-specific architectures have come to the rescue. Key applications for these domain-specific accelerators are networking and security coprocessors as used in SmartNICs, and machine learning and inferencing coprocessors as used in PCIe adapters or appliances. Other applications include cryptocurrency/blockchain, database acceleration and streaming data processing for the Internet of Things (IoT).

Domain-specific architectures, as the name implies, are tailored to a class of workloads that share a common characteristic. The devices are programmable, not hardwired as are traditional ASICs. Domain-specific architectures integrate application and deployment-aware functions and support domain-specific languages for ease of use. Key attributes of a domain-specific architecture are parallelized data processing, function-specific logic, application-aware data management and control.

Two domain-specific silicon-based examples are shown in **Figure 1**, namely Google's Tensor Processing Unit (TPU) for machine learning and AI, and Netronome's Network Flow Processor (NFP) for networking and security. The charts show significantly better performance/watt for these solutions relative to CPUs and FPGAs.

Figure 1: Domain-Specific Silicon Delivers Higher Performance/Watt

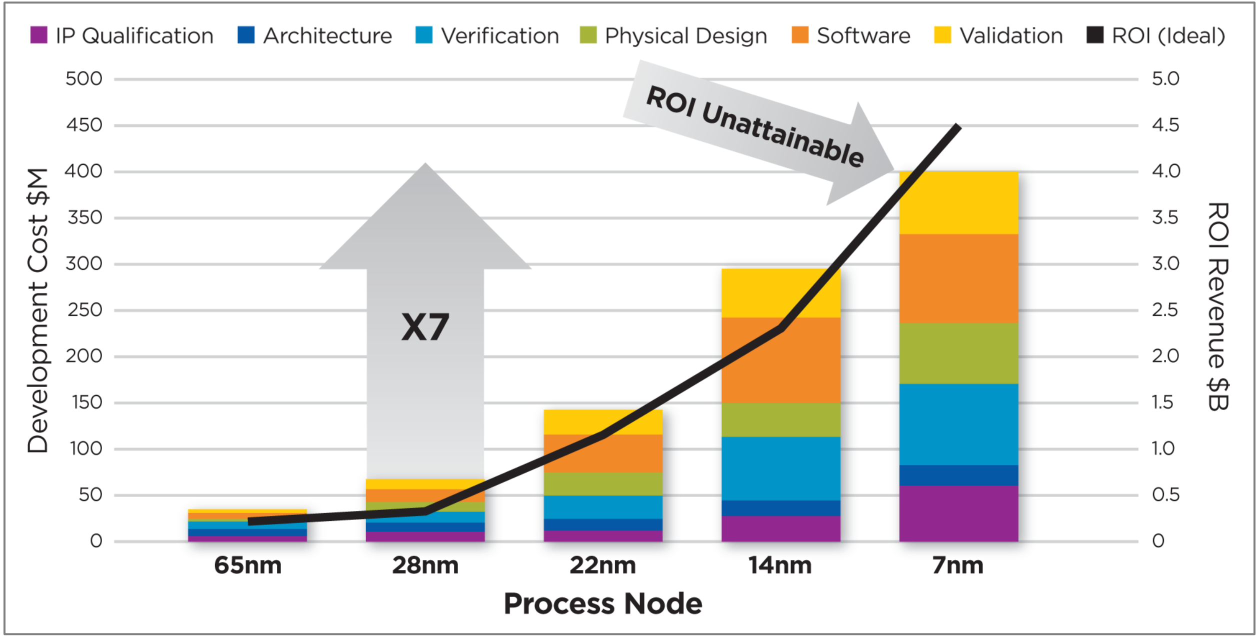


Sources: "An in-depth look at Google's first Tensor Processing Unit (TPU)," Google Cloud, May 2017 (left); Netronome, based on internal benchmarks and industry reports related to Xeon CPUs and Arris FPGAs (right)

Silicon Development Challenges

Domain-specific accelerators require a specific combination of processing cores and hardware engines. As can be seen in **Figure 2**, as technology becomes more complex, the cost for SoC development is rising exponentially from node to node. The ROI revenue required to embark on silicon development at smaller, more advanced process nodes is astounding.

Figure 2: Skyrocketing Costs of Silicon Development



Source: Keith Flamm, "Measuring Moore's Law; Evidence from Price, Cost & Quality Indices," Global Foundries, semiengineering.com ("How much will that chip cost?") and Netronome

Due to the significant cost in developing ASICs in advanced nodes, only extraordinarily well funded projects can be brought to completion. Many accelerator developers, with applications that serve a limited market, are simply unable to justify the expense even though they can provide a significant cost and power advantage versus using processors and FPGAs to implement the needed function.

In the past, a designer could often combine two ASICs in their system into a single monolithic design in the next process node, with a potential increase in frequency as well. As multiple parts are combined into a single device interface power consumption is reduced, an additional benefit beyond the active power improvement of moving to a smaller geometry. While this has area and power benefits over having multiple devices, it often results in ASIC designs being larger than they actually need to be as they are designed with a superset of needed functions for a variety of applications.

Given the increasing investments required to build an ASIC, along with decreasing cost benefits of doing so, it seems that an obvious choice would be to build very large monolithic die in older technology nodes. Unfortunately, this has cost implications driven by defectivity of large die, technical limitations due to the limit of the reticle used in lithography tools and limits of reliable large die attached to laminates.

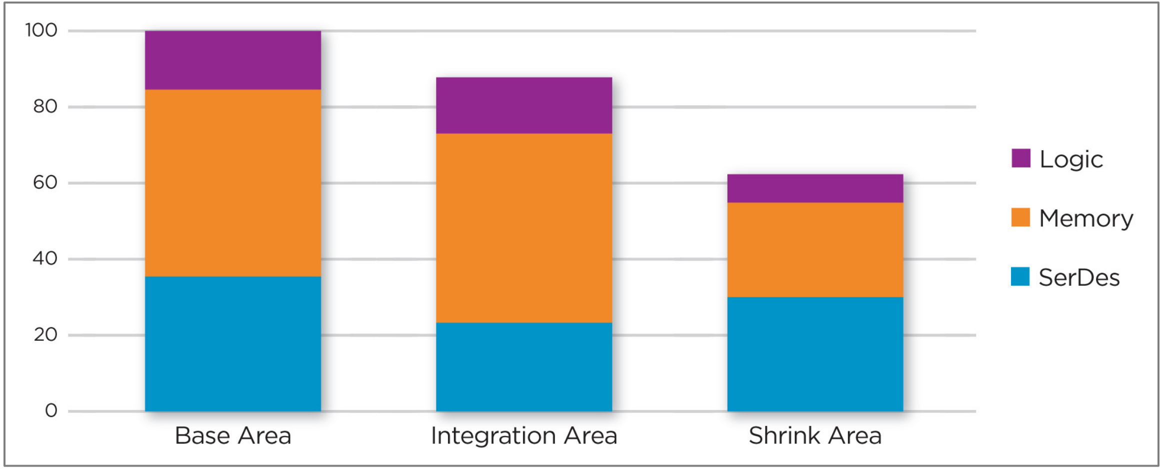
Benefits for End Users and Developers

The use of open domain-specific accelerators offers a third path with heterogeneous integration of multiple component die or chiplets using lower power interfaces such as USR, Bunch of Wires (BoW) or emerging 112G SiP standards. By layering a common protocol over different interfaces, a common "building block"-based approach can be leveraged to create systems on a substrate by simply varying the Bill of Materials (BoM) for an MCM. This chiplet approach can produce many of the benefits of integration at a fraction of the development cost. The integrated system on a substrate also provides a major savings in board real-estate and routing layers, significant system costs that can often dwarf any increased cost to design and integrate the multi-die package.

The economics of building silicon using chiplets with smaller die sizes has been established by multiple companies and products. High Bandwidth Memory (HBM) is an example of a chiplet-based product that uses open interfaces. AMD, with its successful EPYC CPUs, has shown how the use of chiplets with CPU cores can reduce silicon development and manufacturing costs of a 32-core CPU by up to 40 percent. The strategy enables a rapid go-to-market for other versions of CPUs with fewer cores.

One significant advantage of a chiplet-based approach is that unlike a monolithic design, all the components need not be ported to a new process node on each shrink. For example, interface component blocks (such as a Long Reach SerDes tile or Electrical to Optical Interface) may remain in cost-effective nodes to reduce overall investment. As shown in **Figure 3**, while not attaining the same area and power advantage of a technology shrink, this third path provides a considerable area and power savings over monolithic integration in more cost-effective nodes by reducing interface area and power significantly. While multi-chip systems typically have a higher cost than individual die, the incremental investment can be somewhat offset by these area and power savings.

Figure 3: Area for Multi-Die Integration vs. Technology Shrink



Source: Avera Semiconductor

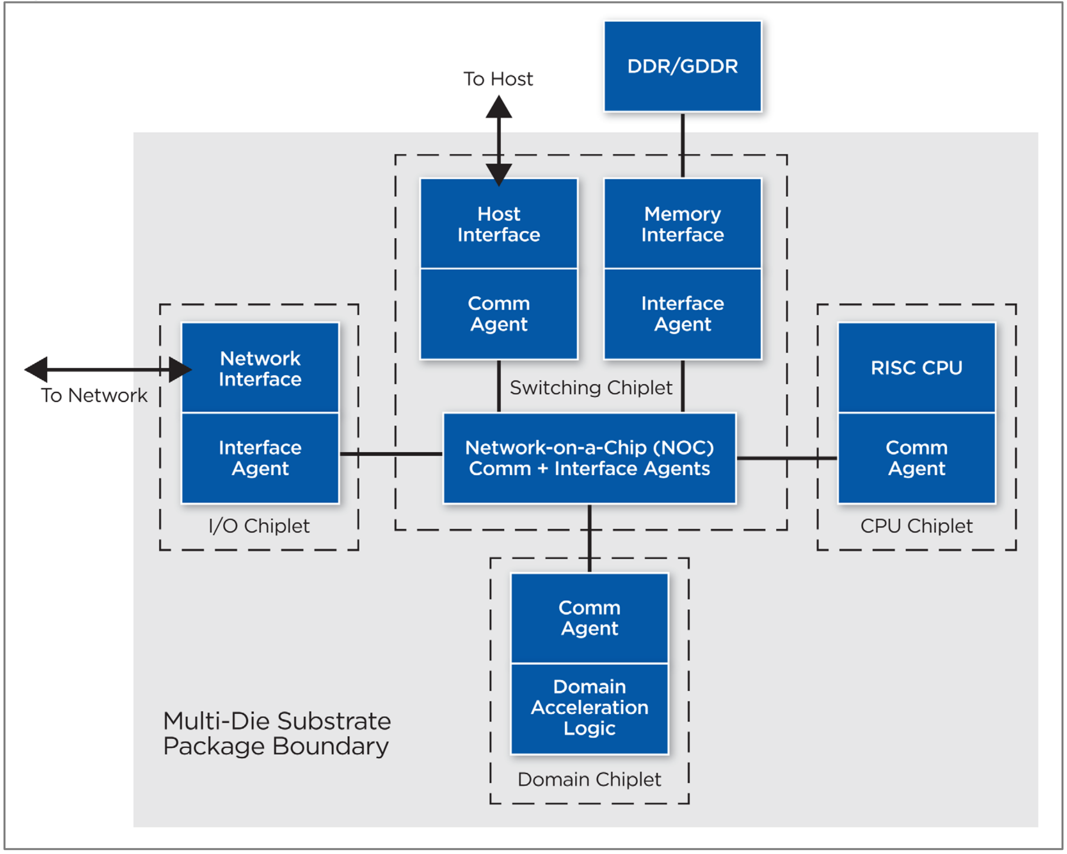
KEY FEATURES OF OPEN DOMAIN-SPECIFIC ACCELERATORS

Architecture

The open domain-specific accelerator architecture being developed in the ODSA Workgroup enables chiplet-based silicon design to be composed using best-of-breed components such as processors, hardware engines, and memory and I/O peripherals using optimal process nodes. The open architecture will provide a complete stack of components (known good die, packaging, interconnect network, software integration stack) that lowers the hardware and software costs of developing and deploying domain-specific accelerator solutions. By implementing open specifications, any vendor's silicon die can become a building block that can be utilized in a chiplet-based SoC design.

Domain-specific accelerator silicon design needs to comprehend both the developer programming model for firmware and how the accelerator is integrated into the system-level application data flow and control management. The ODSA architecture shown in **Figure 4** defines a reference multi-die architecture that is derived from elements common to a wide range of commercial and academic accelerators.

Figure 4: ODSA Reference Multi-Die Architecture



Source: ODSA

The reference design consists of the following chiplets:

- A network I/O chiplet
- A RISC CPU chiplet
- A domain-specific acceleration chiplet, which may be implemented as one or more instances of: an FPGA; a many-core RISC processor; or domain-specific logic
- A switching and interface chiplet to which all the other chiplets are connected

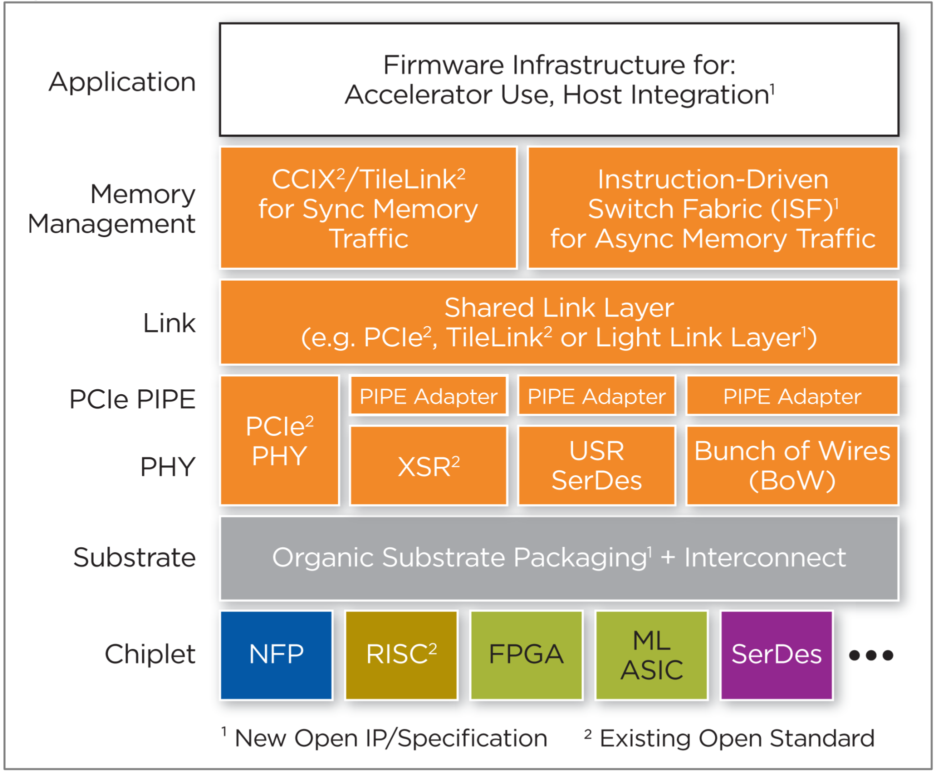
These chiplets are packaged together on a common substrate to form an SoC solution.

Key benefits of using open domain-specific accelerators and chiplets include reduced NRE and time to market, reduced circuit board area and power consumption. By using an open architecture, such as that being developed by the ODSA Workgroup, accelerator developers can assemble systems integrating best-of-breed components from multiple vendors. Development effort and investment can therefore be focused on their own domain expertise.

Protocol Stack

To present a monolithic IC-like development environment, the chiplets in the ODSA implement communication agents to support inter-die networking. The network infrastructure offers a memory model and performance comparable to a monolithic architecture. To compensate for the difference between monolithic and heterogeneous implementations, the ODSA reference architecture will support a complete protocol stack, as shown in **Figure 5**.

Figure 5: ODSA Protocol Stack


Application
Firmware Infrastructure for: Accelerator Use, Host Integration[1]
Memory Management
CCIX[2]/TileLink[2] for Sync Memory Traffic
Instruction-Driven Switch Fabric (ISF)[1] for Async Memory Traffic
Link
Shared Link Layer (e.g. PCIe[2], TileLink[2] or Light Link Layer[1])
PCIe PIPE
PHY
PCIe[2] PHY
PIPE Adapter
PIPE Adapter
PIPE Adapter
XSR[2]
USR SerDes
Bunch of Wires (BoW)
Substrate
Organic Substrate Packaging[1] + Interconnect
Chiplet
NFP
RISC[2]
FPGA
ML ASIC
SerDes
•••
[1] New Open IP/Specification
[2] Existing Open Standard

Source: ODSA

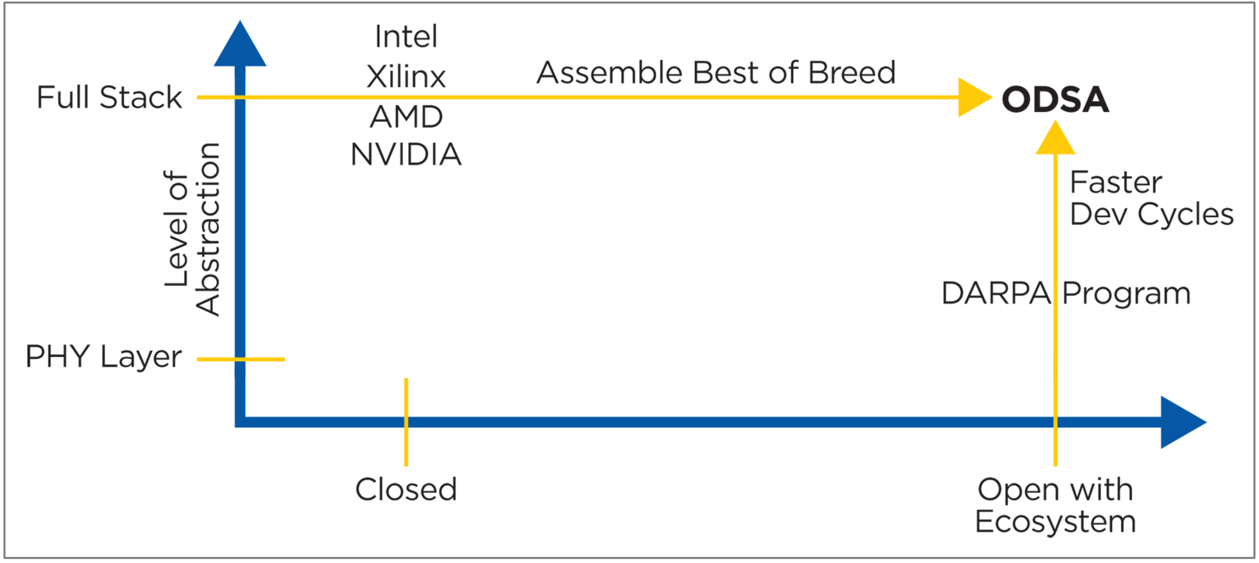
This stack addresses the following issues:

- A network layer to move data around with three sublayers:
  - PHY layer: Protocols for physical inter-die communication
  - Link layer: A protocol for logical communication between physically connected die
  - Routing layer: A protocol to route transactions between devices not directly connected
- A memory layer that is the infrastructure for all processing activity in the ODSA
  - Data Transaction layer: A common transaction model to transport data
  - Data Transport layer: Protocols for memory management across processing components. The architecture will support coherent memory access and a novel approach to instruction-driven data movement.
- The application layer: Firmware and software to integrate the accelerator with a host.

With this approach, higher layers can be preserved even as the design evolves. For example, the PHY protocol can change from PCIe in a prototype to an Ultra Short SerDes in a production part, without impacting higher-layer software. Performance is achieved with a PHY connectivity technology that enables the architecture to be implemented with cost-efficient organic substrate packaging technology.

Large semiconductor companies, including Intel, Xilinx, AMD and Nvidia, have recently announced chiplet-based products that use proprietary protocols at the PHY layer and higher layers of the on-package network. With the open ODSA architecture, customers can choose best-of-breed products from all the vendors that support the architecture (see **Figure 6**).

Figure 6: Open & Closed Chiplet Ecosystems



Source: ODSA

The U.S. government agency DARPA recently announced a chiplet initiative. The agency's effort focuses on opening the PHY layer protocol. The ODSA specifies an entire protocol stack above the PHY layer, greatly simplifying and accelerating product development. The

ODSA also supports SerDes technologies that reduce wire count, enabling products to be built with simple organic substrates.

DELIVERING OPEN DOMAIN-SPECIFIC ACCELERATORS

The ODSA Workgroup proposes to make a platform prototype, built from currently existing chiplets, available with the following components:

- A networking chiplet with multiple USR SerDes ports to implement message-based communication
- A long-range SerDes chiplet with a USR SerDes port
- A RISC CPU chiplet with a USR SerDes port
- A multi-chip package with room for an additional accelerator die
- Host integration software for networking acceleration

There are a number of chiplet-based solutions already available including:

- Achronix offers FPGA chiplets today alongside their FPGA chips and embedded FPGA cores. Chiplets allow companies to create very high-performance and power-efficient solutions while keeping die size smaller to achieve good yield.
- Netronome can support the development of ODSA prototypes with the NFP-4000 and 6000 devices. These devices support up to four independent PCIe Gen 3 interfaces. Netronome intends to develop a prototype implementation of the ODSA with the PCIe as the inter-chiplet interface.
- Sarcina is a semiconductor packaging and testing company that supports C4 bump-on-pad, Cu pillar bump on-pad, Cu pillar bump-on-trace, BGA based package, wirebond die, QFN package, TQFP package, die stacking, SMD cap, 0201M silicon cap, discrete devices and other options.
- Averasemi/GLOBALFOUNDRIES is in production with multi-chip solutions today, including multi-die with chip scale package, HBM and other memory integration using organic laminates and silicon interposers.

The vision of chiplets is a broad ecosystem of thousands of interoperable chiplets built in various foundries that deliver a wide variety of functionality for lower costs, faster time-to-market and more cost-effective innovation. Business models will need to support this vision.

In order for this method to succeed, novel business models need to be established. Integrated ASIC providers have already put in place effective models for integration of HBM modules, memory devices and known good die systems. This model can be extended to provide much more complex integration with components from multiple sources.

Technical Challenges

The key technical challenge for open domain-specific accelerators are ensuring the connectivity between chiplets is open and standardized and that applications can be developed and executed on the reference multi-die architecture in a similar way to a monolithic architecture.

The transaction layer used between chiplets is the key enabler for this integration. In order for an integrated CPU or accelerator unit to realize maximum efficiency, access to external data interfaces (PCI Express, Ethernet, Flash memory) must appear as if they are "local" to the processing unit and be totally transparent to the software model. By leveraging the ODSA model, developers are free to choose the optimal solution for each chiplet based on performance needs, IP availability and cost.

As an example, a solution with relatively limited external bandwidth requirements but a need for premium performance in an AI accelerator may choose to implement the accelerator and CPU chiplets in an advanced FinFET node (16/14 nm or beyond) while choosing to use a lower-cost 28 nm technology to implement a small handful of PCI Express and 10G Ethernet links. With the ODSA architecture, the CPU and accelerator chiplets will see these links as local resources that appear to the system as if they were attached to the same internal bus.

Beyond this, multi-chiplet packaging will have to deal with the well-known problem of selecting known good die to integrate into a package.

Business Challenges

Although offering a similar functionality in a system, chiplet technology will require a different business model than that used for silicon IP. The reason for this is that chiplets, unlike silicon IP, will need to be processed, manufactured and quality-guaranteed for years, if not decades.

Silicon IP requires a very intensive engagement during the design process, but relatively little need for continuing support after production. A chiplet business will need to provide manufacturing guarantees regarding how long the supplier will guarantee that a chiplet will be in production, alternatively a chiplet provider may offer the transfer of manufacturing rights to the MCM developer in exchange for royalty payments. This approach would simplify the selection of chiplets by a vendor with regard to concerns about long-term availability.

ODSA WORKGROUP

The Importance of the ODSA Workgroup

Through the ODSA, the silicon industry has the opportunity to expand the use of chiplets from single vendors to a complete ecosystem, enabling system developers and service providers to develop and deploy advanced SoC solutions.

The following quotes from leading industry figures provide some insight into current thinking:

- "We are extremely excited to collaborate with industry leaders and contribute significant intellectual property and related open specifications derived from the proven NFP products and apply that effectively to the open and composable chiplet-based architecture being developed in the ODSA Workgroup." said Niel Viljoen, founder and CEO at Netronome.

- "We are delighted to join and bring our embedded FPGA technology to the ODSA Workgroup to enable customers to bring open, cost-efficient accelerator products to market." said Steve Mensor, vice president of marketing at Achronix.
- "Our collaboration efforts with the ODSA Workgroup ensure an additional option to enable data center SoC accelerator technology supporting applications from deep learning for artificial intelligence to next-generation 5G networks." said Kevin O'Buckley, general manager of Avera Semiconductor.
- "With unprecedented bandwidth and ultra-low power, Glasswing enables companies to quickly and efficiently build flexible yet optimized solutions for workload-specific applications," said Dr. Amin Shokrollahi, founder and CEO at Kandou. "Kandou supports the ODSA Workgroup and delivering Glasswing as a critical component."
- "NXP is pleased to join the ODSA Workgroup and provide its Multicore Arm SoC solutions to enable low-power, low-latency, open accelerator solutions that deliver greater cost and performance efficiencies." said Sam Fuller, director of marketing at NXP.
- "We are pleased to be a member of the ODSA Workgroup and look forward to SiFive's leading RISC-V Core IP being available in chiplet form, potentially via our silicon capabilities, to enable customers to create open, heterogeneous, best-in-class accelerators at low cost." said Dr. Naveed Sherwani, president and CEO at SiFive.

Planned Developments

Specifications and contributions within the ODSA Workgroup relate to open and standards-based connectivity between chiplets. They are in the following areas:

- Application layer: Open source firmware infrastructure and host software for accelerator use and host integration; new open IP/specification
- Memory management layer:
    - Protocols for synchronous memory traffic based on an open protocol leveraging CCIX and TileLink; existing open standard
    - Protocols for asynchronous memory traffic such as Instruction-driven Switch Fabric (ISF); new open IP/specification based on Netronome's on-chip communication protocol
- Link layer: A light link layer protocol based on and leveraging PCIe and TileLink; existing open standard
- Multiple PHY layer interfaces:
    - Based on and leveraging PCIe XSR; existing open standard
    - Based on USR SerDes, and Bunch of Wires (BoW) abstracted through an open, common interface such as PCIe PIPE
- Substrate layer: Based on organic substrate packaging and interconnect; new open IP/specification

CONCLUSIONS

CPU performance cannot keep up with the processing required to deliver cloud and network services and many other applications. Domain-specific accelerators can dramatically increase the performance of servers for specific applications, reducing the number of servers required and thus saving cost and power consumption. To gain the best benefits from domain-specific accelerators, these must be optimized for specific applications and integrate best-of-breed components.

Chiplets offer an attractive alternative to monolithic designs for domain-specific accelerators and other applications. The development and manufacturing costs of multiple die and MCM assembly can be significantly smaller than the costs of developing and manufacturing complete SoC solutions on 7 nm or 10 nm silicon technology. Domain-specific accelerators are optimized for specific applications and therefore companies may want to use multiple accelerators with different combinations of chiplets.

Open standards are key to a strong chiplet ecosystem. As more chiplets become available with compatible interfaces, companies will be able to create solutions more quickly and cost-effectively. The ODSA architecture and related specifications will help the development of the chiplets with open interfaces and a common protocol stack that are required to build this strong ecosystem.

If you wish to help build this ecosystem or partner with the ODSA and its members, please contact info@odsachiplets.org.