Open. Together.
Open Domain-Specific Architecture (ODSA) Sub-project Launch

Bapi Vinnakota, Netronome
ODSA: A New Server Subgroup (Incubation)

Extending Moore’s Law

- Domain-Specific Accelerators: Programmable ASICs to accelerate high-intensity workloads (e.g. Tensorflow, Network Flow Processor, Antminer…)
- Chiplets: Build complex ASICs from multiple die, instead of as monolithic devices, to reduce development time/costs and manufacturing costs.

Open Domain-Specific Architecture: An architecture to build accelerators

- Today: All multi-chiplet products are based on proprietary interfaces
- Tomorrow: Select best-of-breed chiplets from multiple vendors
- Incubating a new group, to define a new open interface, build a PoC
Server Productivity on a Steep Decline

Death of Moore’s Law means general purpose CPUs cannot keep up with demands of new workloads. OCP exploring accelerators.

Source: Netronome based on internal benchmarks and industry reports
Domain-Specific Architectures

Tailor architecture to a domain

- Server-attached devices — programmable, not hardwired
- Integrated application and deployment-aware development of devices, firmware, systems, software
- 5-10X power performance improvement

A New Golden Age for Computer Architecture
John L. Hennessy, David A. Patterson

Google TPU vs. CPU and GPU
Source: "An in-depth look at Google’s first Tensor Processing Unit (TPU)," Google Cloud, May 2017

Netronome NFP vs. CPU and FPGA
Source: Netronome, based on internal benchmarks and industry reports related to Xeon CPUs and Arm FPGAs
Exponential Costs of Silicon Development

- Designs are too costly at advanced nodes
- Impossible to justify for smaller markets
- Only the largest companies can afford
- Stymies innovation
- Limits choice

https://semiengineering.com/big-trouble-at-3nm/
Solution: Move From Monolithic to Chiplets

Shrink: Monolithic process shrink
Integration: Multi-chip on same process

Integration provides nearly all the benefits of a shrink at a fraction of the cost, because of efficient inter-chiplet interconnect

https://www.netronome.com/media/documents/WP_ODSA_Open_Accelerator_Architecture.pdf
Chiplet Use Cases

High-Bandwidth Memory
• 3D stacked memory with interposer and wide parallel standard interface
• Open inter-chiplet interface

Xilinx Versal
• 3D stacked FPGA, SerDes, Application Processor

Intel Foveros
• 3D stacked CPU, GPU, Application Processor

3D INTEGRATION
All the benefits of 2D integration plus a new level of density thanks to Foveros, allowing for a radical re-architecture of systems-on-chips
Growing Interest in Chiplets

I want to...

- Provide chiplets to...
- Provide Design Services to...
- Provide Technology (...
- Provide an open standard...
- Consumer of chiplets to...

The biggest problem is...

- Business model (including I...
- Packaging and test...
- Architecture/st standardization...
- Die-to-die PHY layer
- POC prototype

Great interest across the ecosystem
Great confusion about how this all works

With an Open System:

- More Choice
- Best-of-breed
- Leverage economies of scale
- Cheaper
Multiple chiplets need to function as though they are on one die.
Multi-Chiplet Reference Architecture for DSA

<table>
<thead>
<tr>
<th>Design Function</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP Qualification</td>
<td>Verified IP for inter-chiplet communication</td>
</tr>
<tr>
<td>Architecture</td>
<td>Leverage reference architecture.</td>
</tr>
<tr>
<td>Verification</td>
<td>Focus investment on domain-specific logic.</td>
</tr>
<tr>
<td>Physical</td>
<td>Reuse chiplets instead of IP for 40% of the functions in a monolithic design</td>
</tr>
<tr>
<td>Software</td>
<td>Open source firmware and software for host-attached operation</td>
</tr>
<tr>
<td>Prototype</td>
<td>Aim for reference package design with area, power budgets and pinouts for components</td>
</tr>
<tr>
<td>Test and Validation</td>
<td>Develop workflow for chiplets</td>
</tr>
</tbody>
</table>

Open. Together.
ODSA in OCP Server Project

- Multiple OCP projects use accelerators
- Open architectural interface to support accelerator designs across multiple carrier cards
- Power, management, reliability requirements vary across sockets
- Enable a collection of ODSA-compliant chiplets, packages, sockets, in the OCP marketplace
Define an open cross-chiplet fabric interface

Build an open multi-company chiplet PoC
How to Participate: Join a Workstream

Join Interface/Standards: (Mark Kuemerle/Aaron Sullivan)

Join the PoC, Build fast: (Quinn Jacobson/Jawad Nasrullah)

Develop Packaging + Socket, Dev Board

Provide FPGA IP

Define Architectural Interface

Provide PHY technology

Develop software

Provide ODSA chiplets

Join Business, IP and workflow: (Sam Fuller/Jeff McGuire)

Define test and assembly workflow

Provide Chiplet IP

Workstream contact information at the ODSA wiki
Call to Action

Timeline
- ODSA announced, 7 companies: 10/1/18
- White paper, 10 companies: 12/5/18
- Workshop, 35 companies: 1/28/19
- OCP Incubation: 3/15/19

How to Participate
- Subscribe to the Mailing list: https://ocp-all.groups.io/g/OCP-ODSA
- Attend the next workshop: 3/28/19 @ Samsung, - deep dive and where we need help register at wiki, or search “ODSA” at www.eventbrite.com
- Join a workstream: PoC, Interface/Standards, Business IP/Workflow

Thanks to:
Achronix: Quinn Jacobson, Manoj Roge; Aquantia: Ramin Farjad; Avera Semi: Dan Greenberg, Mark Kuemerle, Wolfgang Sauter; Ayar Labs: Shahab Ardalan; ESNet: Yatish Kumar; Kandou: Brian Holden, Jeff McGuire; Netronome: Sujal Das, Jim Finnegan, Jennifer Mendola, Brian Sparks, Niel Viljoen; NXP: Sam Fuller; OCP: Bill Carter, Archna Haylock, Dharmesh Jani, Steve Roberts, Seth Sethapong, John Stuewe, Aaron Sullivan, Siamak Tavallaei; Samtec: Marc Verdiell; Sarcina: Larry Zu; zGlue: Jawad Nasrullah.
Open. Together.

OCP Global Summit | March 14–15, 2019