

Flavor Physics Workshop Neckarzimmern

Moving from HEP to Industry with FPGAs

Christian Färber

Neckarzimmern - Germany

15. March 2023

Copyright © 2023 Intel Corporation.

This document is intended for personal use only.

Unauthorized distribution, modification, public performance,
public display, or copying of this material via any medium is strictly prohibited



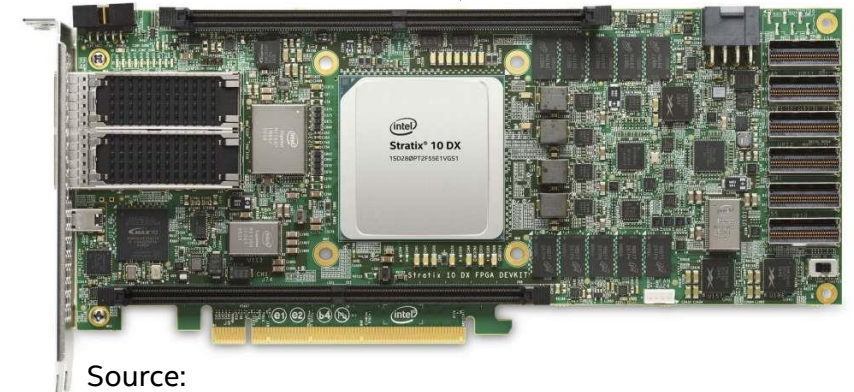
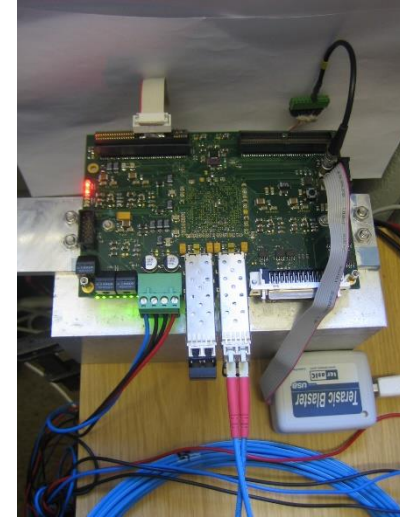
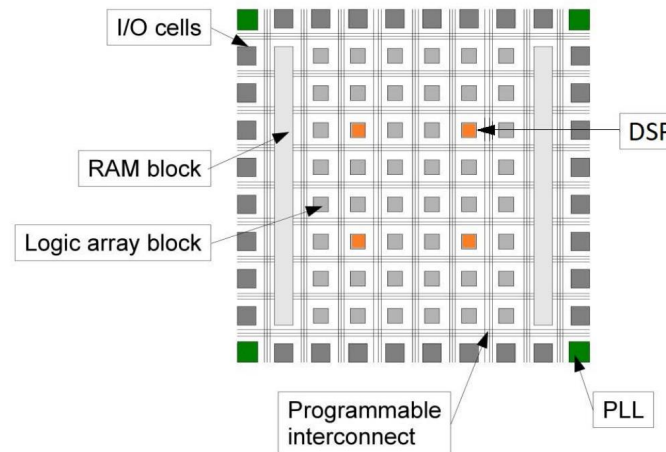
Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

Agenda

Moving from HEP to Industry with FPGAs

- Career Path
 - Heidelberg University
 - Vector Informatik
 - CERN
 - Thales
 - Intel
- What I found useful
 - Interviews
 - General Thoughts
 - Paths of FPGA Colleagues
- Summary
- FPGA Introduction
- FPGA compute acceleration

Field Programmable Gate Array



Source:

<https://www.intel.de/content/www/de/de/products/details/fpga/development-kits/stratix/10-dx.html>

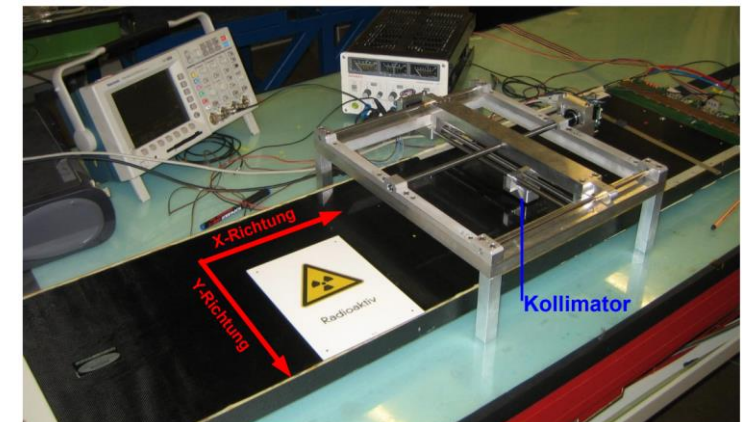
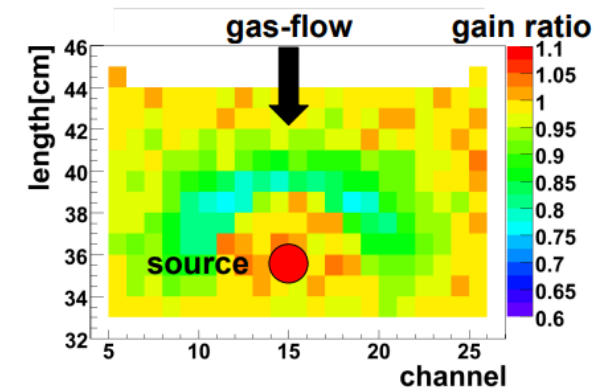
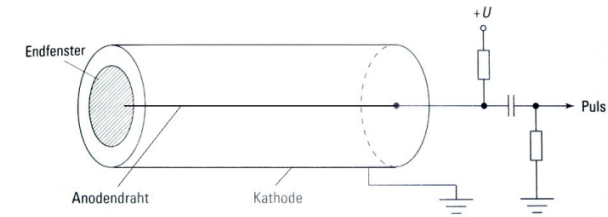


Section: Career Path

Sub-Topics:

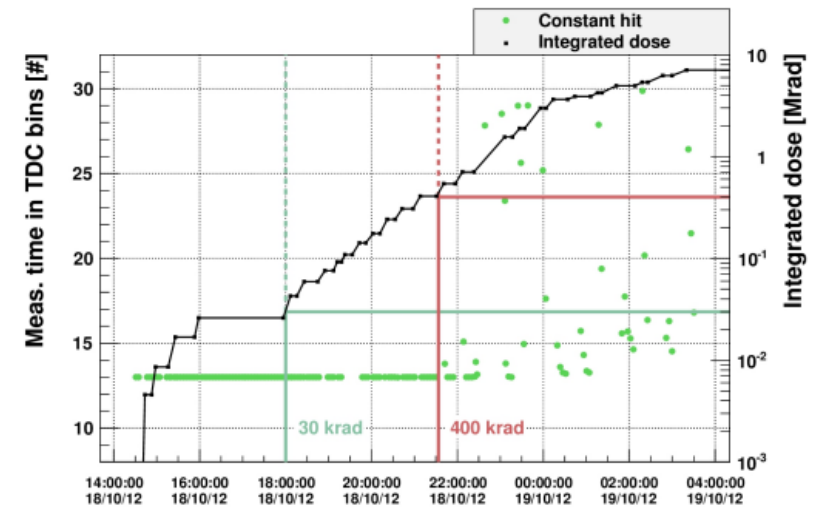
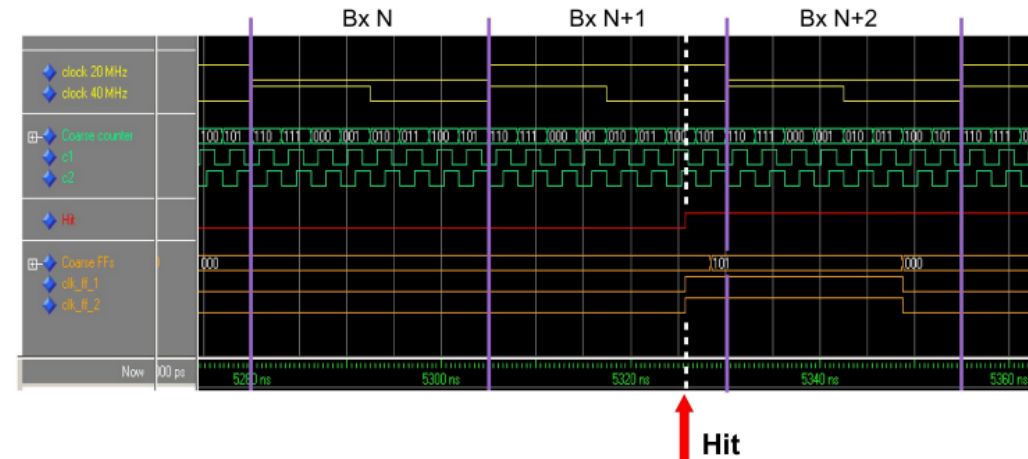
- Heidelberg University
- Vector Informatik
- CERN
- Thales
- Intel

- Investigating Outer Tracker ageing
- Development of automatic ageing scanner
- Adding oxygen to counting gas and investigate ozone creation to reduce ageing
- Skills:
 - Undertaking irradiation tests in the lab
 - Implement a lab readout chain as also lab experiment control
 - Analyzing data with C++ and ROOT



PhD

- Prototyped FPGA-based readout board for potential OT upgrade
 - FPGA-based Time to digital converter ($t_{res} < 1\text{ns}$)
 - FPGA irradiation tests at MPIK and HIT
 - Collaboration with engineers from PI workshop
- Support of Outer Tracker commissioning
- Outer Tracker gas monitoring
- Skills:
 - Electronics design with FPGAs, optical high-speed transceivers, VHDL, C++, analysis of experiments, presenting at conferences, ...
 - Developed muscle to work into complex new fields, having a high frustration tolerance and the mindset to get things really working
 - With HEP great preparation for joining academia and industry!



https://www.phys.uni-heidelberg.de/Publications/Dissertation_Christian_Faerber.pdf



Section: Career Path

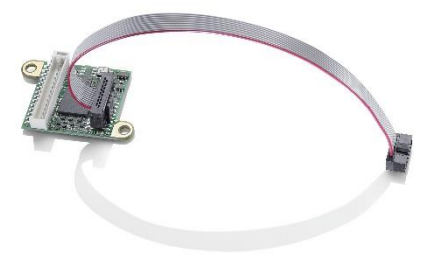
Sub-Topics:

- Heidelberg University
- **Vector Informatik**
- CERN
- Thales
- Intel

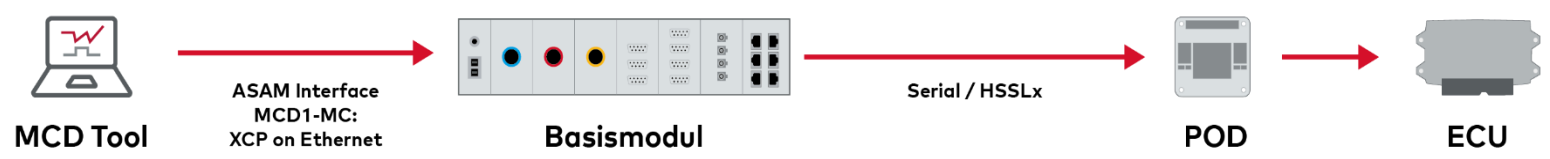
Vector Informatik

Hardware Development Engineer

- Vector: Develops software tools and components for networking of electronic systems based on serial bus systems for **automotive**
- Project: FPGA development of a high-speed interface and a memory synchronization block for the VX1000 hardware, used for vehicle ECU calibration
- Team of 5 FW engineers
 - Next to software team and PCB team, in total 20
- Programming Guide for VHDL & C++
- Industry timelines
- Working time
- Trainings necessary for CAN ...



Readout chain



Source: <https://www.vector.com/us/en/>

Vector Informatik

How I got this job

- Job portals / social networks
 - Stepstone
 - For large companies check their career portal!
 - LinkedIn / (in the past Xing)
- [Job-Messe University Heidelberg](#)
- Created a professional resume – 2 pages only!
- Applied for several positions and companies
- Shared and discussed experience with my FPGA network
 - Dry run of interview with people from the field





Section: Career Path

Sub-Topics:

- Heidelberg University
- Vector Informatik
- **CERN**
- Thales
- Intel

CERN - Openlab

Senior Applied Fellow

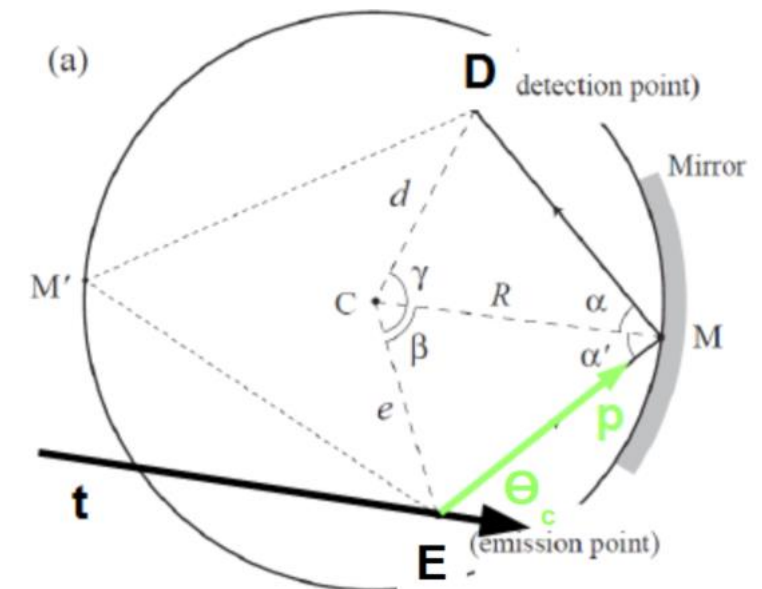
- HTCC - CERN Openlab collaboration with Intel
- Feasibility study to use FPGA-based compute acceleration in the HPC sector of High-Energy physics
 - RICH photon reconstruction
 - Calorimeter raw data encoding
 - CNN inference study with MNIST
- **Highlight:** Prototypes of the Intel[®] Xeon[®]+FPGA hybrid server-processor (HARP - Intel Hardware Accelerator Research Program)
- Got to know Intel development team in Hillsboro very well



CERN - Openlab

RICH PID

- Calculate Cherenkov angle Θ_c for each track t and detection point D , not a typical FPGA algorithm
- RICH PID is not processed for every event, processing time is too long!
- 748 clock cycle long pipeline written in Verilog
 - Additional blocks developed: cube root, complex square root, rot. matrix, cross/scalar product,...
- Lengthy task in Verilog with all test benches

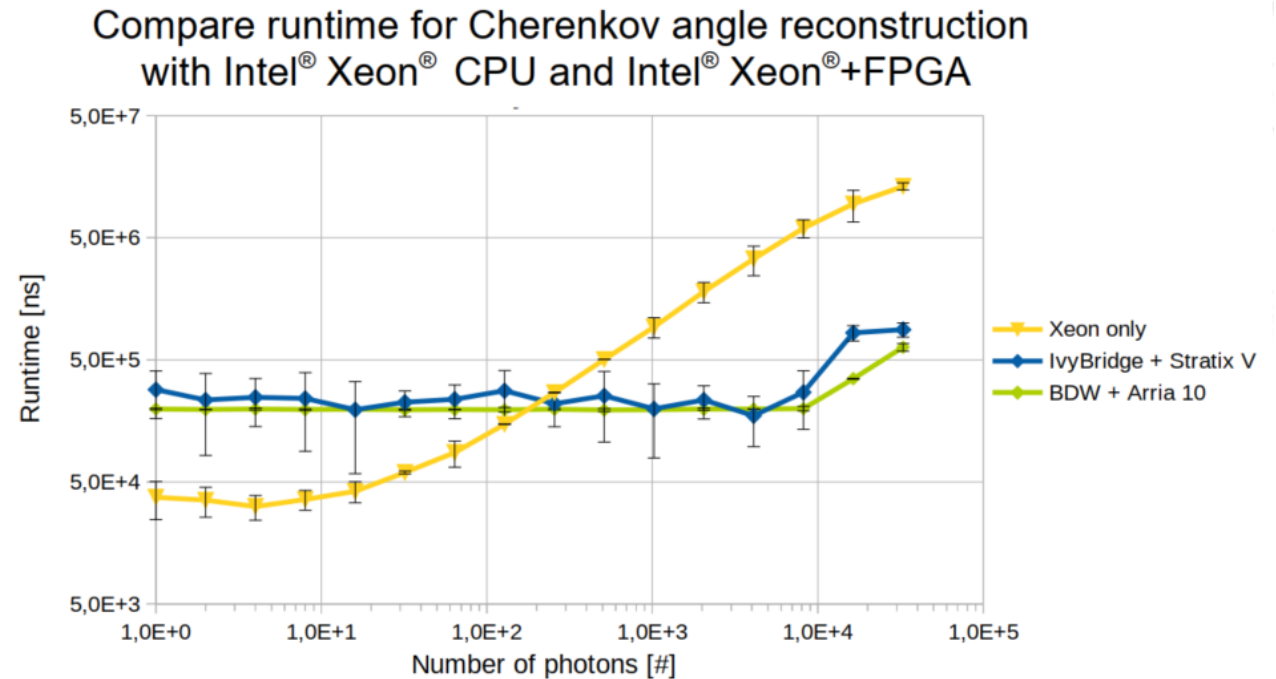


Reference: LHCb Note LHCb-98-040

CERN - Openlab

RICH PID

- Acceleration of up to factor 35 with Intel[®] Xeon[®]+FPGA
 - Theoretical limit of photon pipeline: a factor 64 with respect to single Intel[®] Xeon[™] thread, for Arria[™] 10 a factor ~ 300
 - Bottleneck: Data transfer bandwidth to FPGA, caching can improve this, tests ongoing
- Energy efficiency improvement
- Later implemented with OpenCL and standard PCIe acceleration cards



https://indico.cern.ch/event/669298/attachments/1551772/2433879/CERN_Computing_Seminar_Faerber_20171031.pdf



Section: Career Path

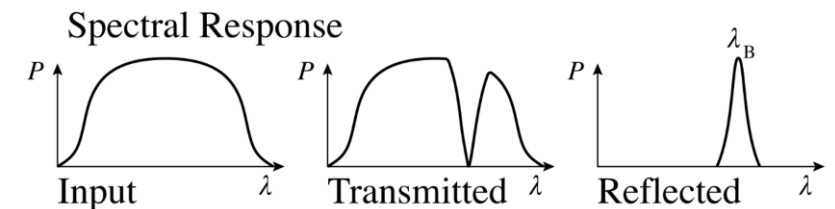
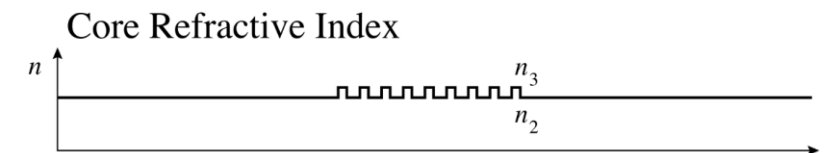
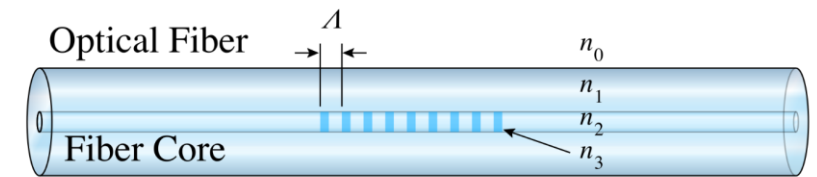
Sub-Topics:

- Heidelberg University
- Vector Informatik
- CERN
- **Thales**
- Intel

Thales - MLS

Senior FPGA Design Engineer

- Thales Group is a French multinational company that designs, develops and manufactures electrical systems for the aerospace, defence, transportation and security sectors.
- Algorithm development and sensor optimization for applications in the **rail industry**
 - Got trainings in **CENELEC EN 50126, ...**
- Fiber Bragg Gating based axle counting system
 - Simulation of sensor geometries
 - Rail field tests
 - Regular presentation of project status to Thales MLS management
 - Collaboration with product group and business unit **supporting early customers**
- Worked in a team of 6 FPGA design engineers, depart. 100.
- Direct manager and project manager



https://en.wikipedia.org/wiki/Fiber_Bragg_grating

Thales - MLS

Axle Counting

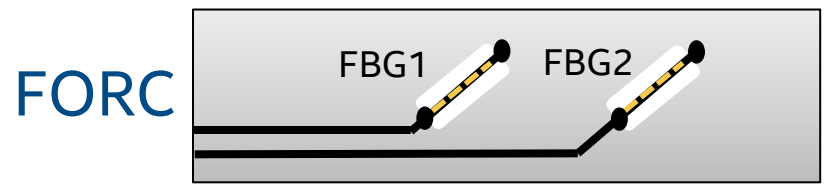
- Why axle counting?
- Old axle counting used magnet field
- New idea to measure bending of rail with FBG
- Replace expensive copper with cheaper fiber
- No power at sensor location needed O(km)



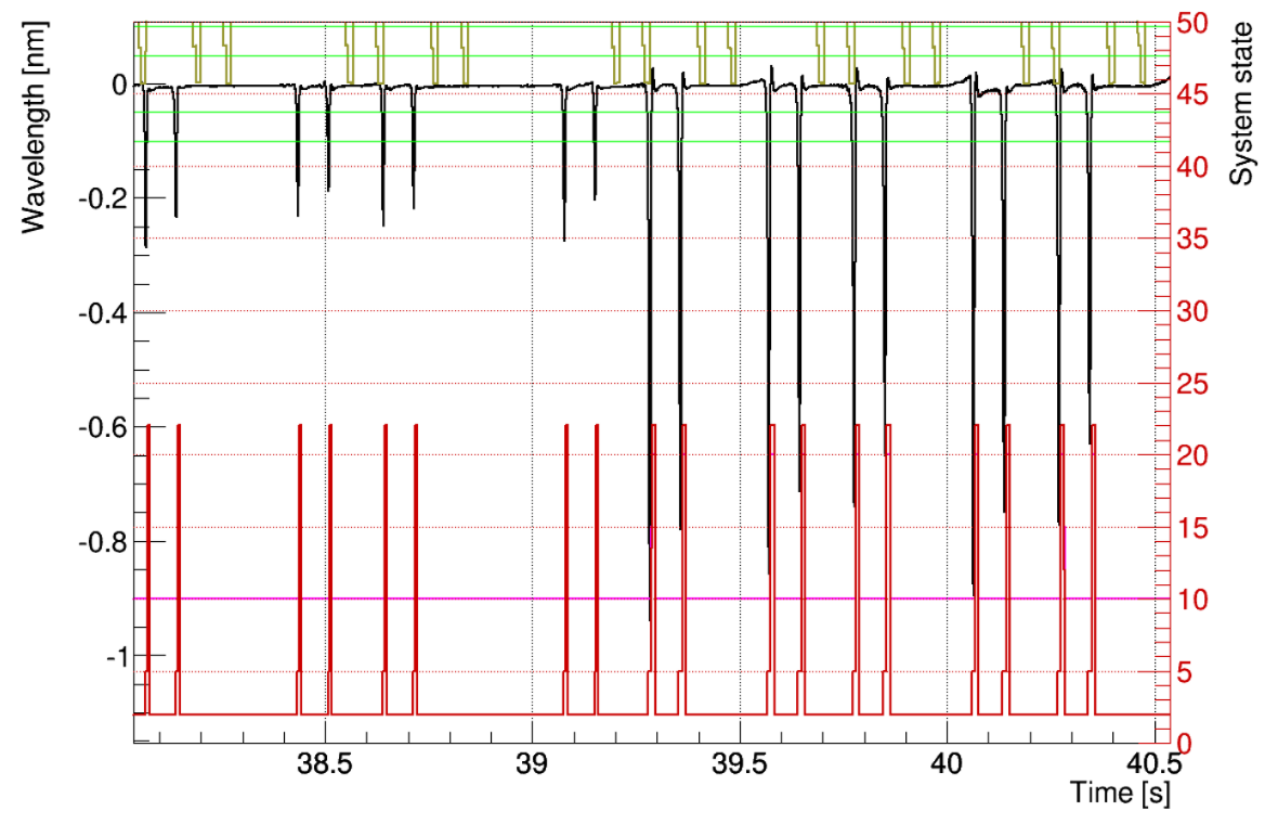
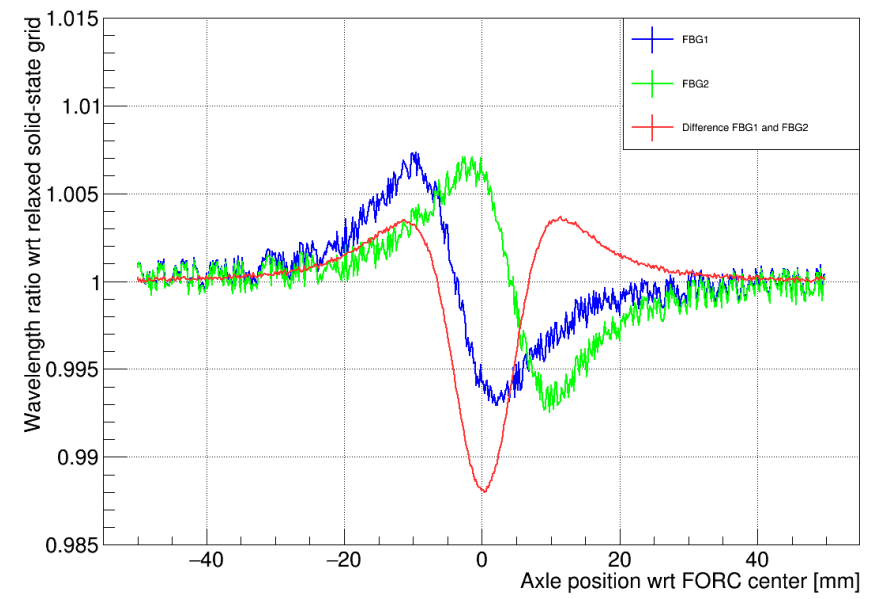
<https://www.thalesgroup.com/en/countries/europe/germany/transportation/rail-field-equipment>

Thales - MLS Axle Counting

Working at Thales is a great possibility to influence modern transportation 😊



FiCoS wavelength shift with axle position: currentGeometry



Great to see your algorithm working on the FPGA in the readout electronics for fields tests with real trains



Section: Career Path

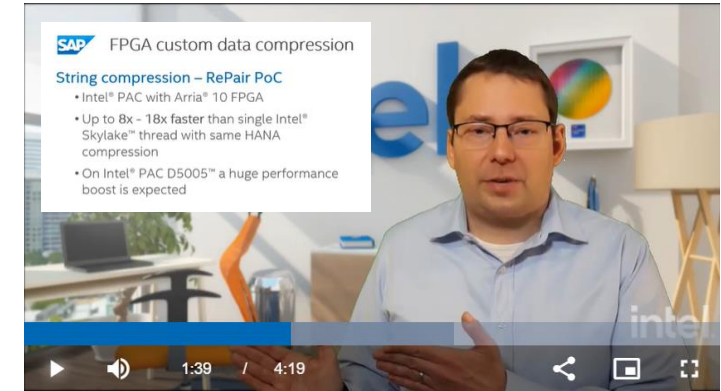
Sub-Topics:

- Heidelberg University
- Vector Informatik
- CERN
- Thales
- Intel

Intel - PSG

Senior Field Application Engineer

- FAE is responsible for technical success of customer projects – customer first mentality
 - Customer trainings onsite in RTL and Intel® oneAPI
 - Realizing Proof of Concepts with customers
 - Support of productization
- New feature development with Intel engineering
- Marketing support
- Find solutions for many different customers from different fields
- Working in a team of ~10 FAEs



<https://salesvideos.intel.com/detail/video/6234855095001/intel%20AE-fpga-acceleration-of-data-compression>



<https://www.intel.com.br/content/dam/www/central-libraries/us/en/documents/2022-09/sap-open-fpga-stack-white-paper.pdf> **intel**



Section: What I found useful

Sub-Topics:

- Interviews
- General Thoughts
- Paths of FPGA Colleagues

Interviews I/II

What to expect

- Headhunter
 - High level, wants normally a CV and 3Qs:
 - Are you in contact with other companies already?
 - [Salary](#) (portals and talk with others)
 - When can you start?
- HR
 - Be confident
 - Rare: Some will challenge you like, why should we hire you and not the others?
 - Often: Some want to understand, what they need to do, that you join
- First time with hiring manager, job and team description, check your skills and expectations



Interviews II/II

What to expect

- Coding interview
 - Live programming, prepare: e.g. [coderbyte](#)
 - Short project, e.g. 10 days to finish
- Technical interview, often with future team members, in the past F2F
 - Prepare to talk about successful projects and small technical questions, no test
 - Be proud what you did
- Last clarification with hiring manager
- HR with last negotiation

- Ask questions!





Section: What I found useful

Sub-Topics:

- Interviews
- **General Thoughts**
- Paths of FPGA Colleagues

General Thoughts

During my career path so far

- Get a mentor and mentor yourself someone
- Create a large network in and around your field
- Create an elevator pitch
- Be pro-active, take responsibility and speak up, you own your own career!
- Changing job and company is normal
- Remote work is in IT even more remote as you may think



<https://www.td.org/talent-development-glossary-terms/what-is-mentoring>



Section: What I found useful

Sub-Topics:

- Interviews
- General Thoughts
- **Paths of FPGA Colleagues**

Paths of FPGA Colleagues

Many interesting paths

- Colleagues who did their PhD with me and worked also with FPGAs are now:
 - Group leader in engineering
 - Aerospace developing FPGA-based control systems
 - Freelancer
 - Transportation
 - ASIC designer
 - Commercial sensor chips
- Similar experience with job market and applications

The screenshot displays a job search interface with the following elements:

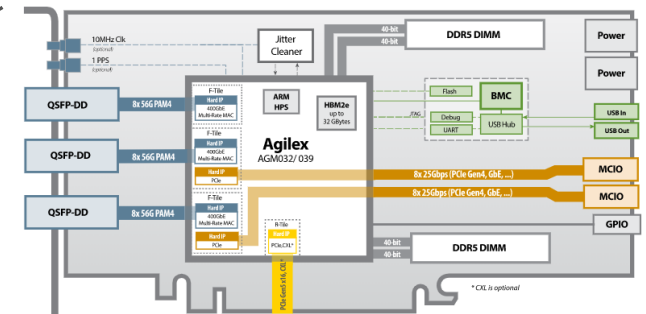
- Search Bar:** Contains 'Fpga' and '(Ort oder 5-stellige PLZ)' with a 'Jobs finden' button.
- Filters:** Includes 'Erscheinungsdatum' (Neuer als 24h, Neuer als 7 Tage), 'Home-Office-Optionen' (Home-Office möglich), 'Gehalt' (Legen Sie fest, wie viel Sie mindestens verdienen möchten), 'Pendelzeit', 'Von wo starten Sie?' (Adresse, PLZ oder Stadt), 'Maximale Dauer' (slider), 'Wie kommen Sie zur Arbeit?' (icons for walking, bicycle, car, bus), 'Bewerbungsart' (Auf Unternehmenswebsite, Schnelle Bewerbung), 'Sprache' (Deutsch, Englisch), and 'Kompetenzen' (Kompetenz Suche, FPGA).
- Job Listings:**
 - FPGA Design Engineer (m/f/d)** at AES Aerospace Embedded Solutions GmbH, München. Posted vor 1 Tag.
 - FPGA Design Engineer (m/w/d)** at Baumer Optronik GmbH, Radeberg. Posted vor 2 Tagen.
 - Applikationsingenieur FPGA (w/m/div.)** at Bosch Gruppe, Reutlingen. Posted vor 8 Tagen.
 - Entwicklungsingenieur „Elektronik / FPGA“ (m/w/d) im Bereich Hardware...** at SCHWIND eye-tech-solutions GmbH, Kleinostheim. Posted vor 1 Woche.
 - FPGA Entwickler/in (m/w/d) für embedded Wärmebildsysteme** at VECTED GmbH, Fürth. Posted vor 1 Woche.
 - FPGA / Embedded Systems Developer (m/f/d) in the field of Machine Vision** at Chromasens GmbH, Konstanz. Posted vor 1 Woche.
 - FPGA Engineer (f/m/d)** at Rockwell Collins Deutschland GmbH, a part of Collins Aerospace.

Summary

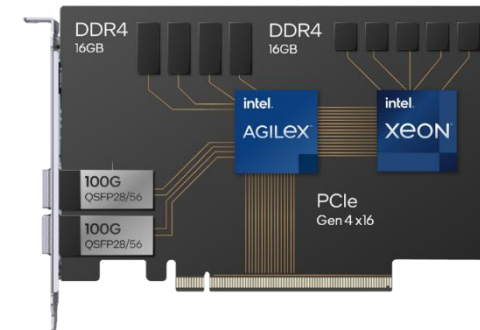
Moving from HEP to Industry with FPGAs

- Working in HEP prepared a large set of skills to be successful in industry 😊
- With hardware & FPGA knowhow you can select the job you want, great future perspective !
- FPGAs are used in a large variety of electronics from edge to cloud
 - FPGA field is growing fast
- Intel® oneAPI enables software engineers to use FPGAs without using RTL 😊
- FPGAs get used also as compute accelerators and infrastructure processing units in data centers and cloud

1
oneAPI



Source: <https://www.bittware.com/intel>



Further questions? 😊

Contact:

christian.ferber@intel.com

Or

<https://www.linkedin.com/in/dr-christian-ferber-b7211590/>

intel®



Section: FPGA Introduction

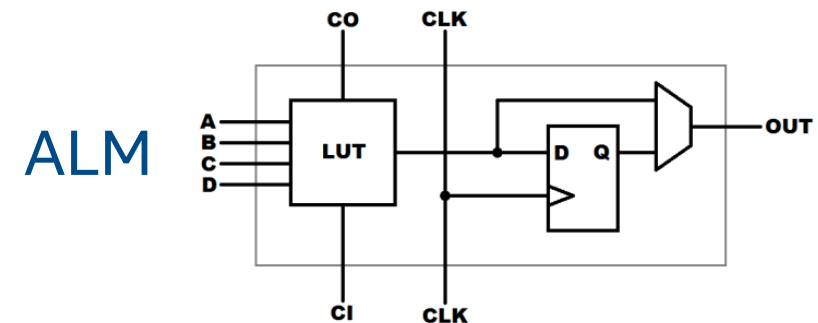
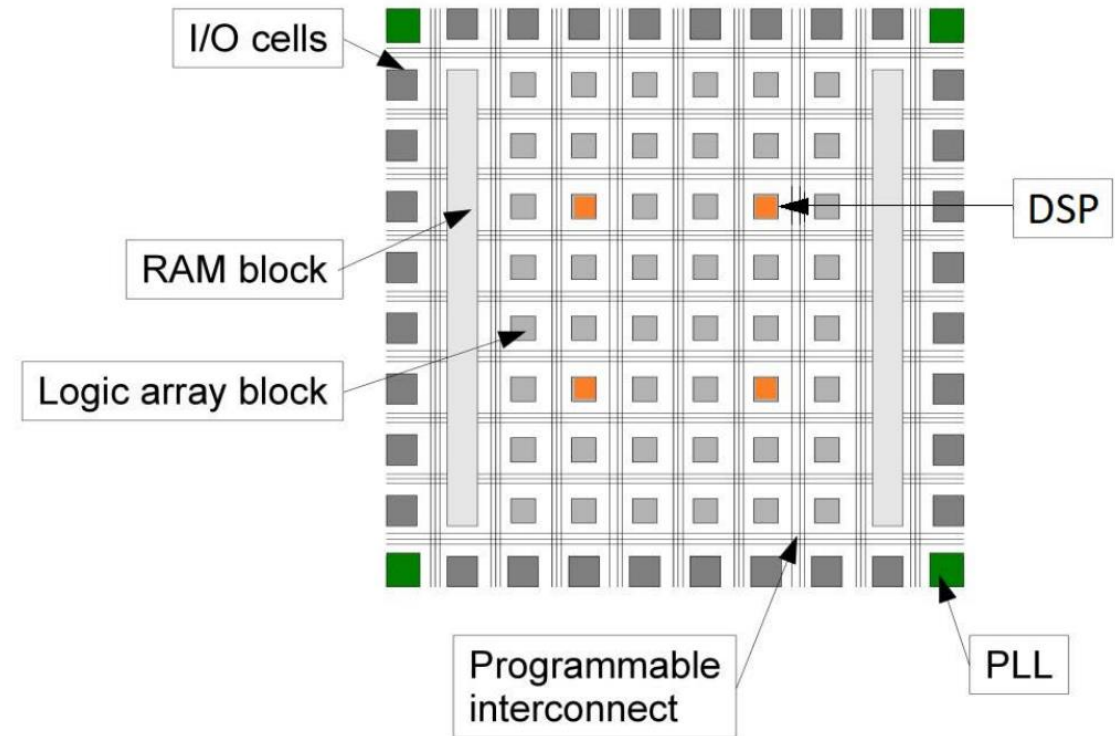
Sub-Topics:

- FPGA Architecture
- Use Cases

FPGA Architecture

Base Components

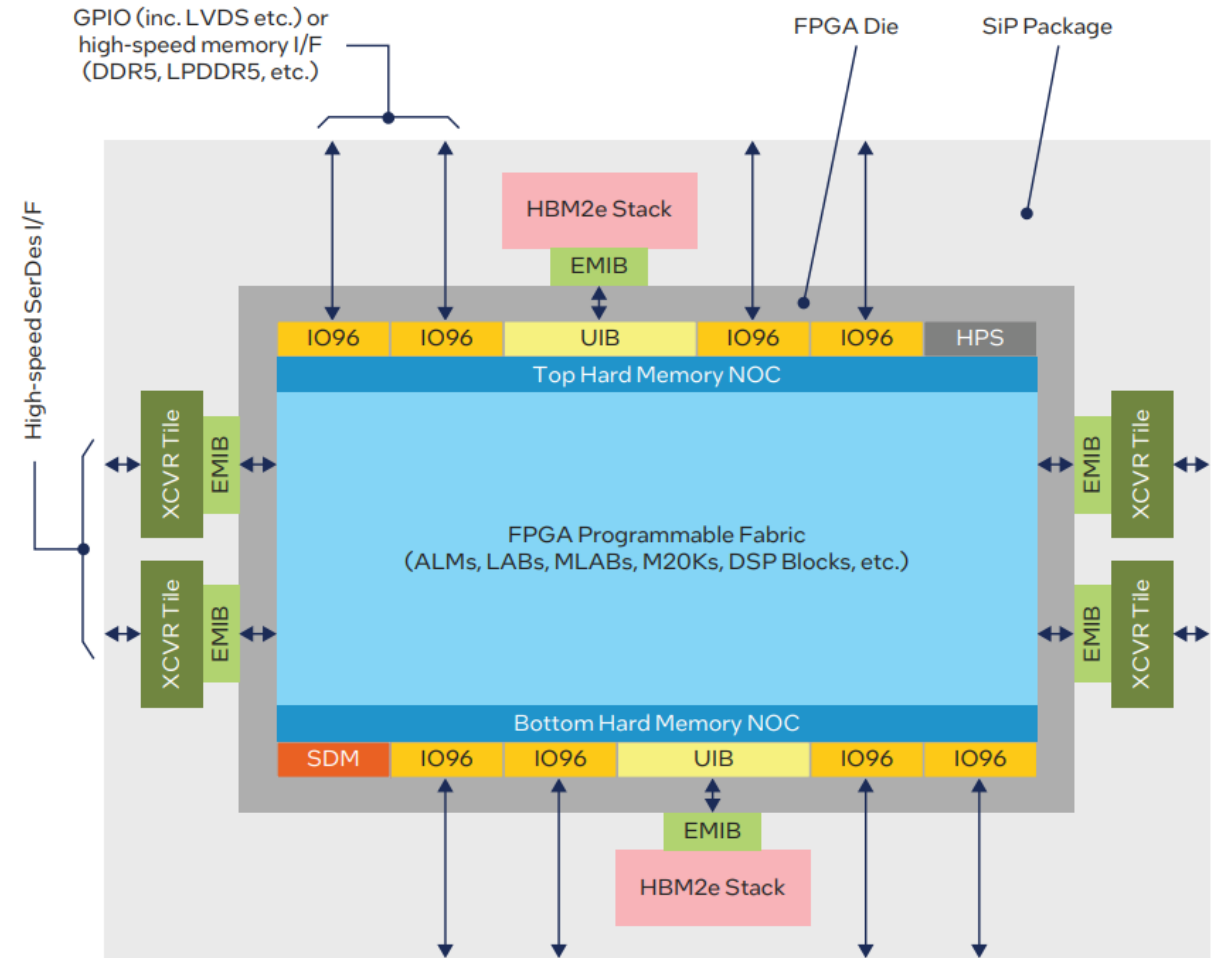
- ALMs (# up to 4M)
 - configurable building block that combines a look-up table and a flip-flop to implement logic functions and register-based storage.
- DSPs (# up to 12k)
 - specialized hardware block with high-speed arithmetic and signal processing capabilities used for implementing complex digital signal processing algorithms.
- RAM blocks (30MB), I/O, PLLs
- Programming: HDL (Hardware Description Language)



FPGA Architecture

Advanced Components

- HPS (Hard Processor System)
- Floating Point and ML DSPs
- Hyperflex registers
- Hardened IPs (e.g. DDR4, PCIe)
- High speed Transceivers
- High Bandwidth Memory
- NoC (Network on Chip)
- Chiplet tiles (PCIe5, CXL, Network, ...)



<https://www.intel.de/content/www/de/de/products/docs/programmable/agilex-7-fpga-m-series-memory-bandwidth-wp.html>



Section: FPGA Introduction

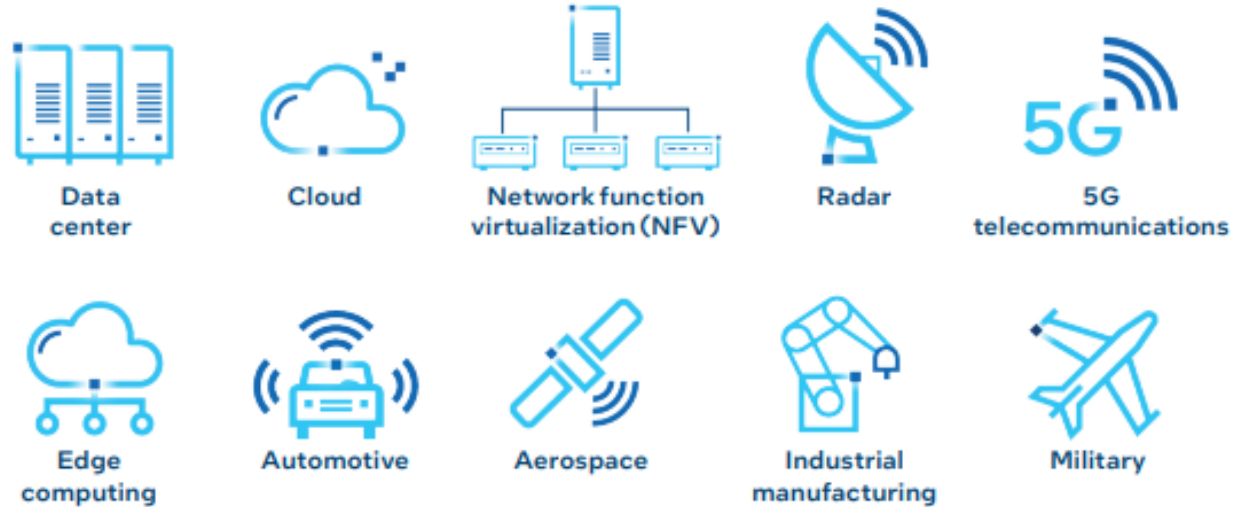
Sub-Topics:

- Architecture
- Use Cases

Use Cases

FPGA base use cases

- Digital Signal Processing, Embedded Systems and Prototyping
- Aerospace, Medical, and Automotive
- Networking and Communications
- Industrial Automation and Control
- High-Performance Computing



New Demands Driving Customization Needs

**Embedded /
Edge**

**Real-Time Actionable
Intelligence**

**Customized Connectivity
Low Latency Compute**

**Communications
Infrastructure**

**High-Bandwidth
Aggregation & Processing**

**Maximum Data
Throughput with Network
Acceleration**

**Cloud /
Enterprise**

**Managing, Organizing &
Processing the Explosion
of Data**

**Customized Acceleration
of Diverse Workloads**



Section: FPGA compute acceleration

Sub-Topics:

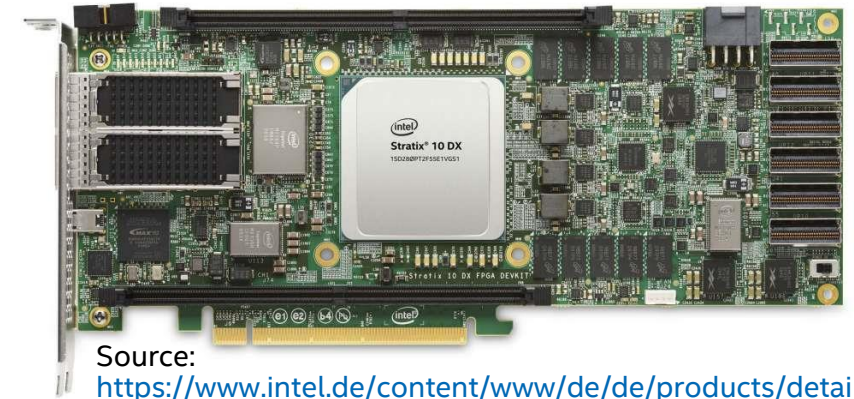
- Hardware
- Intel® oneAPI
- Intel® DevCloud
- FPGAs in HPC
- FPGAaaS

FPGA supported Technologies

FPGA Advantages

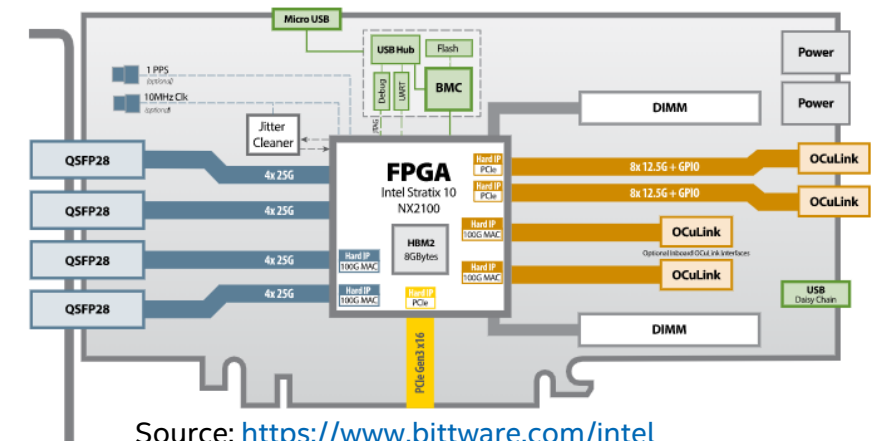
- UPI and now CXL
 - Cache-coherent, high bandwidth and low latency interface
 - Memory extension with DDR4/5
 - High Bandwidth Memory
 - In package 32GB HBM2e, up to 1TB/s
- IKL – FPGA to FPGA connection
- Network interfaces
- Special AI DSP blocks – S10NX
 - <https://www.intel.com/content/www/us/en/products/programmable/stratix-10-nx-fpga-vs-gpu-ai-conference-paper.html>

Intel® S10DX



Source: <https://www.intel.de/content/www/de/de/products/details/fpga/development-kits/stratix/10-dx.html>

BittWare: 520NX

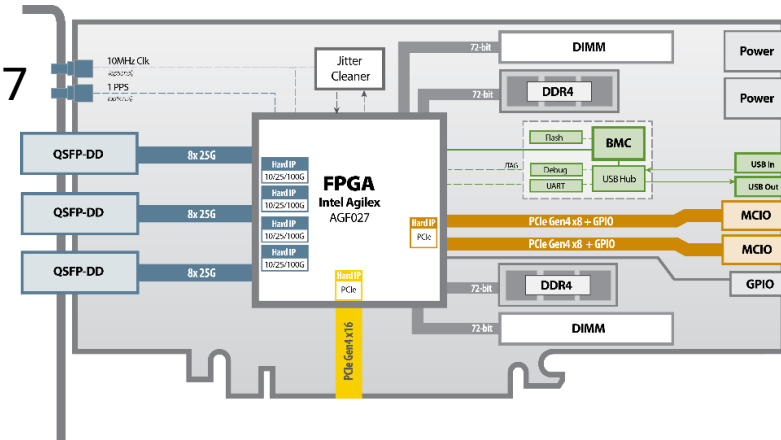


Source: <https://www.bittware.com/intel>

Intel® Agilex™ FPGA Cards with Intel® oneAPI

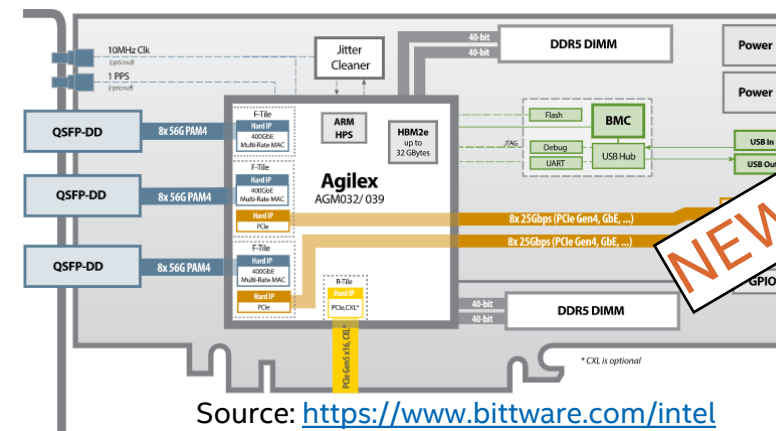
- PCIe Full Size, Dual Width Intel® Agilex™ FPGA : AGF027
- PCIe Gen4 x16
- 128GB DDR4 (4x banks)
- 3x QSFPDD (200G); 2x MCIO x8 expansion
- BittWare BMC + SDK

BittWare: IA-840f



- PCIe Dual Width Intel® Agilex™ FPGA : AGM039
- PCIe Gen5 x16 + CXL
- 32GB HBM2e
- 32GB DDR5 (2x banks)
- 3x QSFP-DD (400G)
- BittWare BMC + SDK

BittWare: IA-860m



Source: <https://www.bittware.com/intel>



Section: FPGA compute acceleration

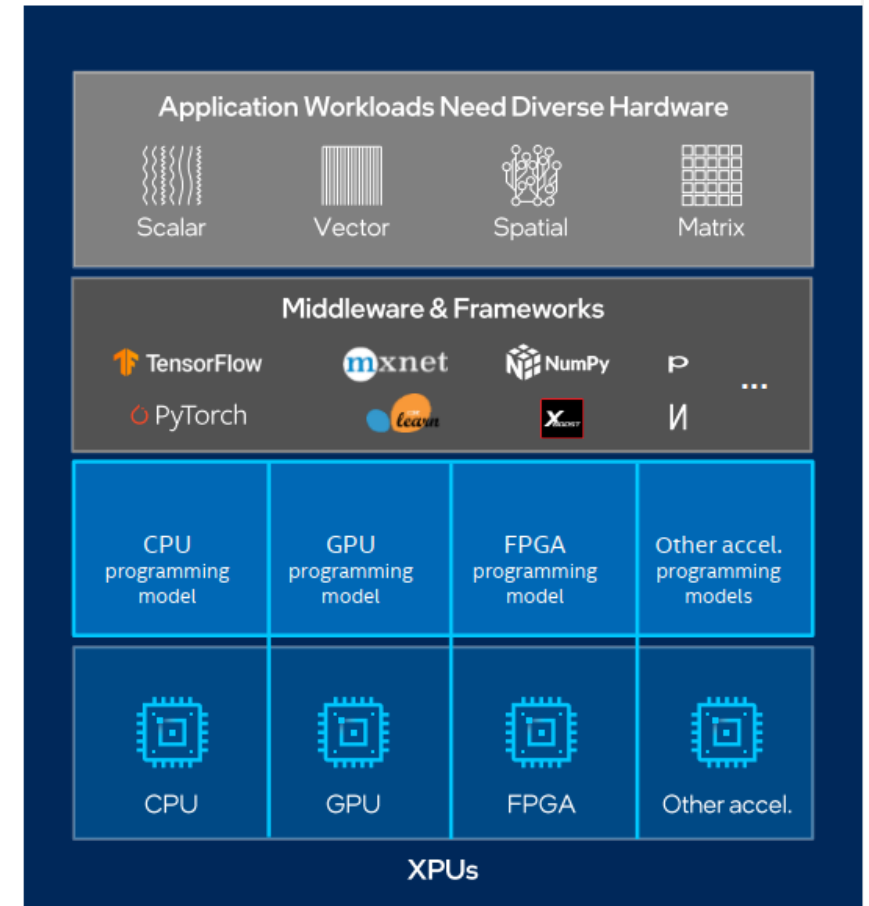
Sub-Topics:

- Hardware
- **Intel® oneAPI**
- Intel® DevCloud
- FPGAs in HPC
- FPGAaaS

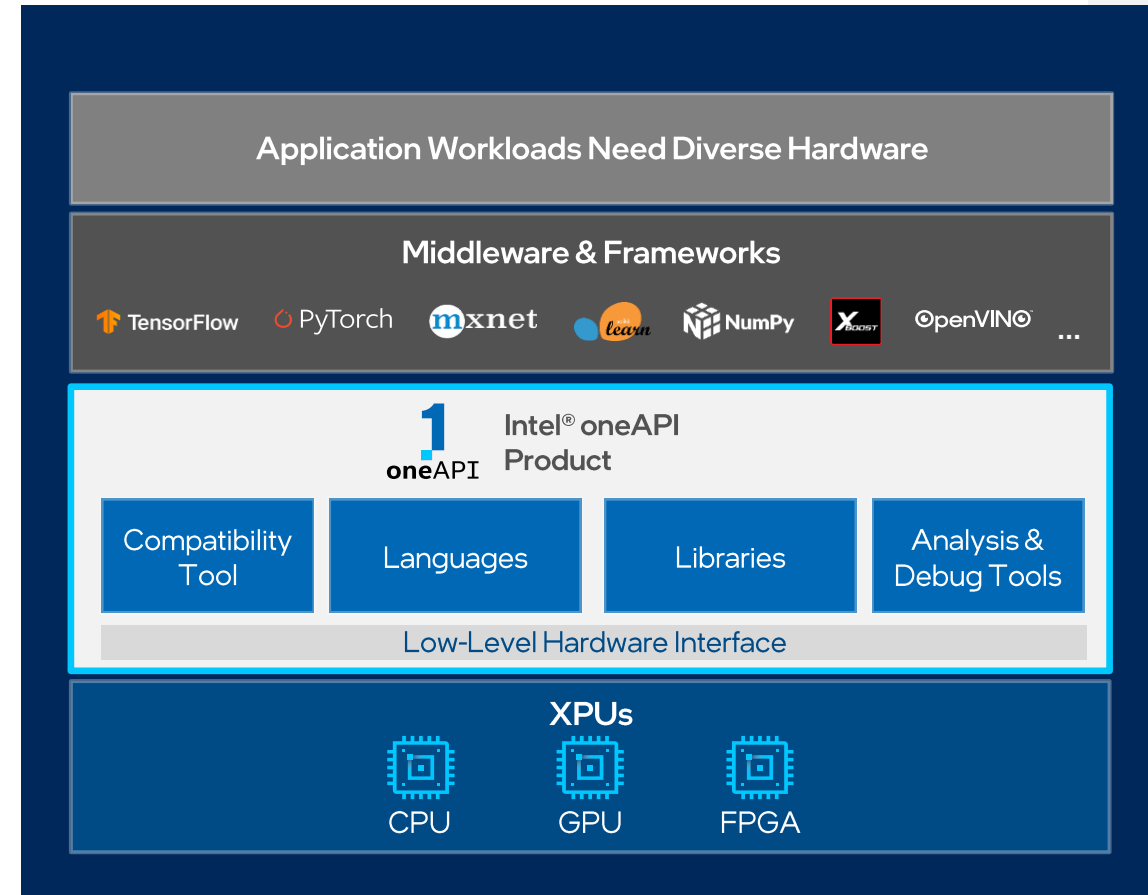
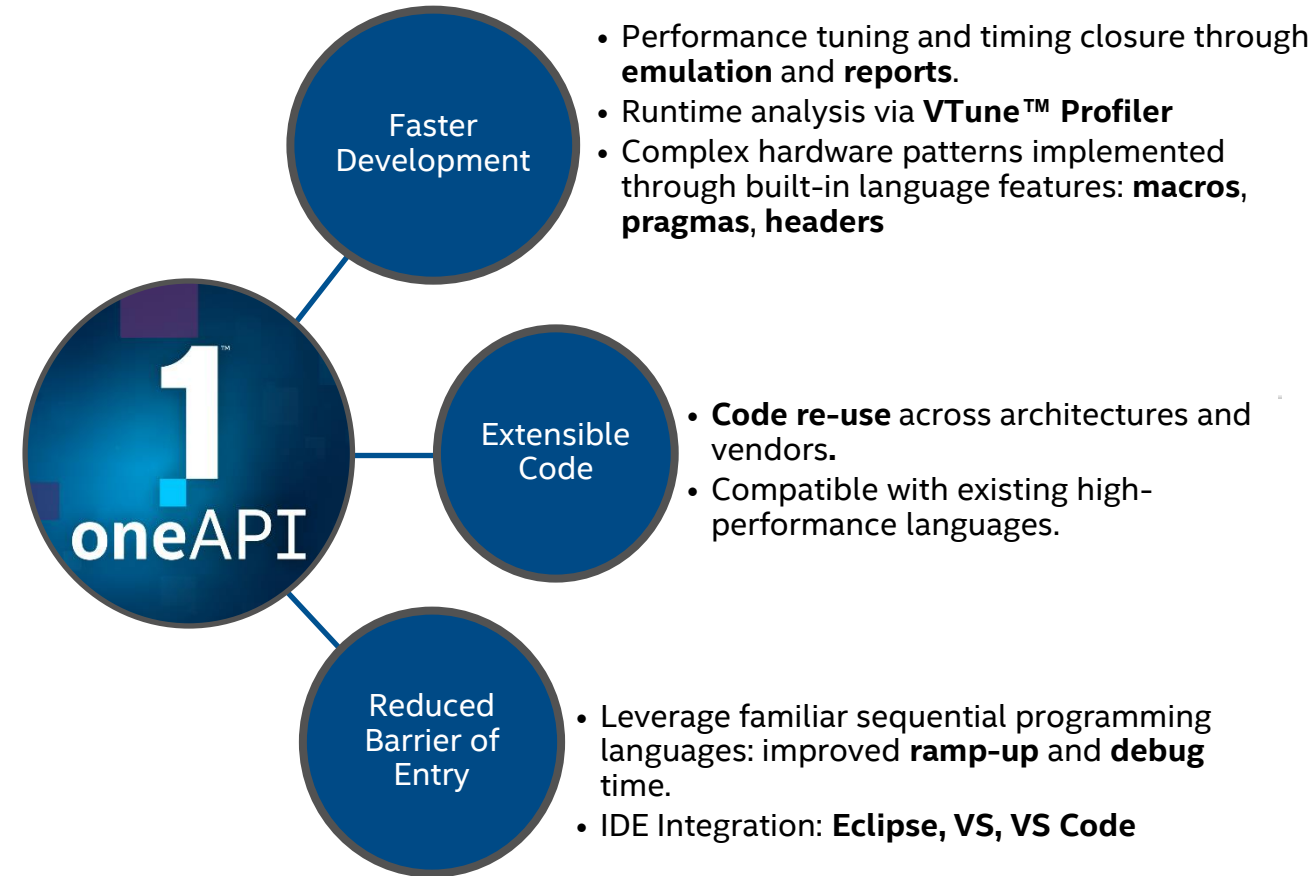
Programming Challenges

Multiple Architectures

- **Separate** programming models and toolchains for each architecture.
 - Required **training** and **licensing** – compiler, IDE, debugger, analytics/monitoring tool, deployment tool, et al. – per architecture.
 - Challenging experience in **debug**, **monitoring**, and **maintenance** of a cross-architectural source code.
 - Difficult integration across proprietary IPs and architectures and no code re-use.
- Software development complexity **limits** freedom of architectural choice.
 - Isolated investments required for technical expertise to overcome the barrier-to-entry



Intel® oneAPI Product



[Available Now](#)

DPC++: Three Scopes

- DPC++ Programs consist of 3 scopes:
 - **Application scope** - Normal host code
 - **Command group scope** - Submitting data and commands that are for the accelerator
 - **Kernel scope** – Code executed on the accelerator
- The full capabilities of C++ are available at application and command group scope
- At kernel scope there are limitations in accepted C++
 - Most important is no recursive code
 - See SYCL specification for complete list

```
void dpcpp_code(int* a, int* b, int* c) {  
    //Set up an FPGA device selector  
    INTEL::fpga_selector selector;  
    // Set up a DPC++ device queue  
    queue q(selector);  
    // Setup buffers for input and output vectors  
    buffer buf_a(a, range<1>(N));  
    buffer buf_b(b, range<1>(N));  
    buffer buf_c(c, range<1>(N));  
  
    //Submit Command group function object to the queue  
    q.submit([&](handler &h){  
        //Create device accessors to buffers  
        accessor a(buf_a, h, read_only);  
        accessor b(buf_b, h, read_only);  
        accessor c(buf_c, h, write_only);  
        //Dispatch the kernel  
        h.single_task<VectorAdd>([=]() {  
            for (int i = 0; i < kSize; i++) {  
                c[i] = a[i] + b[i];  
            }  
        });  
    });  
}
```

Application Scope

Command Group Scope

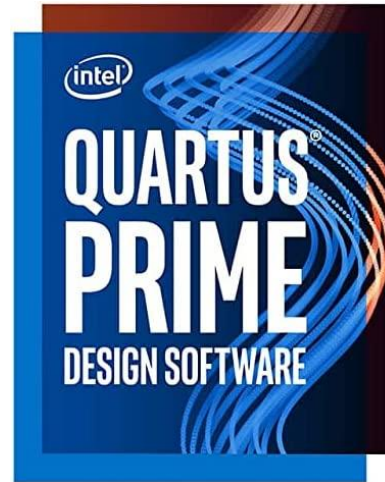
Kernel Scope

Getting Started with oneAPI on an FPGA



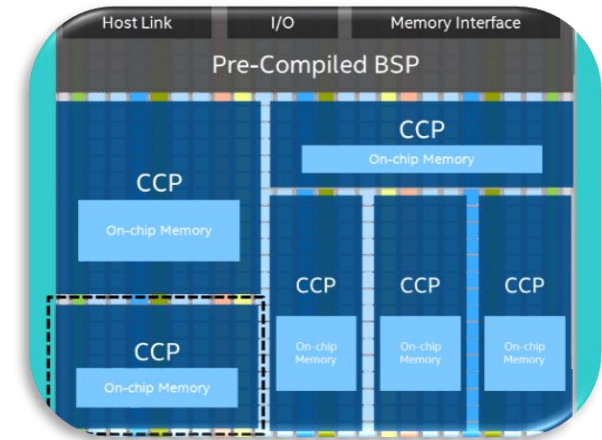
Intel® oneAPI Base Toolkit

+



Intel Quartus® Prime Design Software

+



Board Support Package (BSP)

Note: Developers using custom platforms should write with IOFS their own BSP or obtain a BSP from their 3rd part platform vendor.



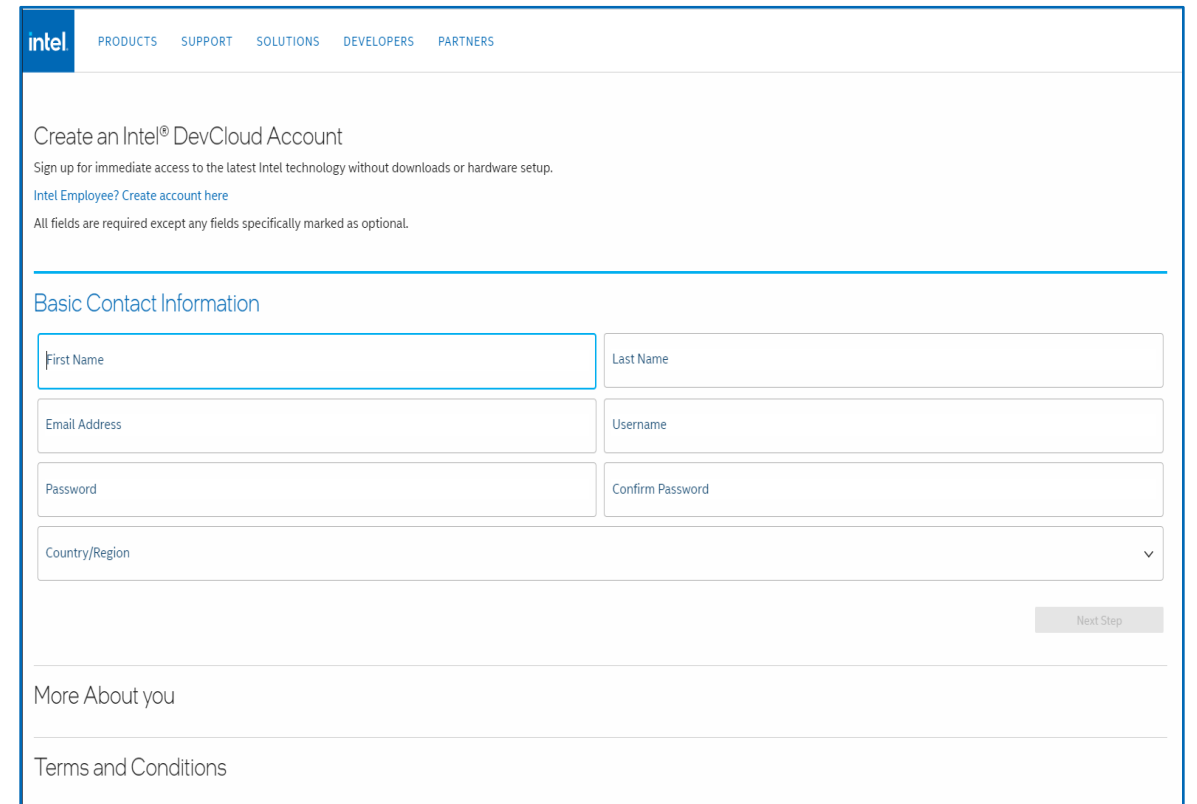
Section: FPGA compute acceleration

Sub-Topics:

- Hardware
- Intel[®] oneAPI
- Intel[®] DevCloud
- FPGAs in HPC
- FPGAaaS

Intel® DevCloud

- Sign up here:
 - <https://software.intel.com/devcloud>
 - Account for 120 days
 - Nodes with cards installed in the group fpga_runtime
 - Nodes with extra memory for full FPGA compiles in the group fpga_compile
 - Intel® oneAPI environment already ready



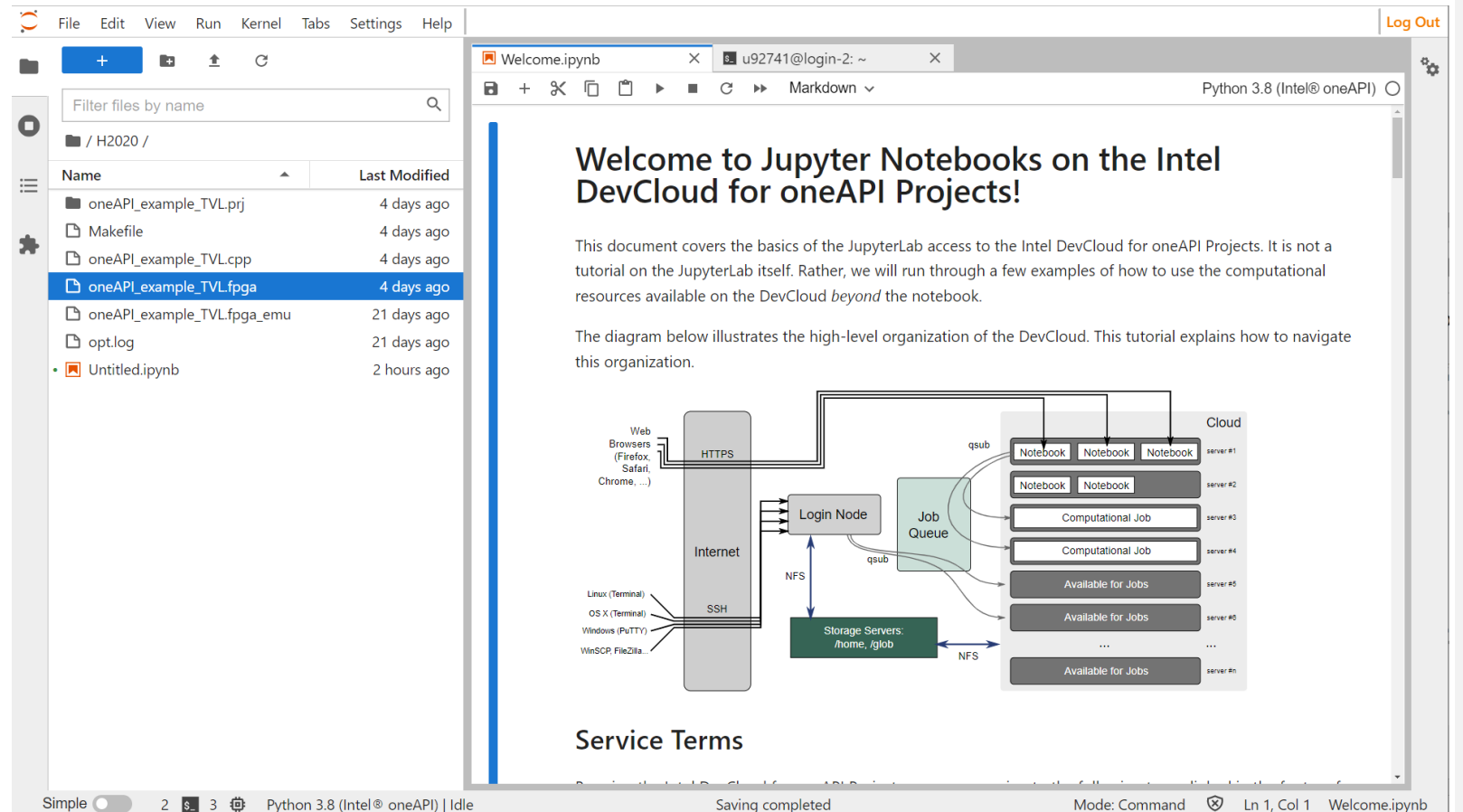
The screenshot shows the Intel DevCloud account creation page. At the top, there is a navigation bar with the Intel logo and links for PRODUCTS, SUPPORT, SOLUTIONS, DEVELOPERS, and PARTNERS. The main heading is "Create an Intel® DevCloud Account", followed by a sub-heading "Sign up for immediate access to the latest Intel technology without downloads or hardware setup." and a link "Intel Employee? Create account here". A note states "All fields are required except any fields specifically marked as optional." Below this is a section titled "Basic Contact Information" containing several input fields: First Name, Last Name, Email Address, Username, Password, and Confirm Password. There is also a dropdown menu for Country/Region. A "Next Step" button is located at the bottom right of the form. Below the form, there are links for "More About you" and "Terms and Conditions".

Use the Intel[®] DevCloud

- SSH to gateway
- Or Jupyter Notebooks via browser
- Job queue for compile and run
 - Job output into log file
- Access to single node also possible
- Session time limit:
 - default : 6hrs
 - max : 24hrs
- Many samples and tutorials in Git repro

<https://github.com/oneapi-src/oneAPI-samples/tree/master/DirectProgramming/DPC%2B%2BFPGA>

Jupyter Notebooks



The screenshot shows the Jupyter Notebook interface. On the left is a file browser for the directory /H2020/. The file list is as follows:

Name	Last Modified
oneAPI_example_TVL.prj	4 days ago
Makefile	4 days ago
oneAPI_example_TVL.cpp	4 days ago
oneAPI_example_TVL.fpga	4 days ago
oneAPI_example_TVL.fpga_emu	21 days ago
opt.log	21 days ago
Untitled.ipynb	2 hours ago

The main notebook area displays a "Welcome to Jupyter Notebooks on the Intel DevCloud for oneAPI Projects!" message. Below the message is a diagram illustrating the high-level organization of the DevCloud. The diagram shows the following components and their interactions:

- Internet:** Acts as the central hub for external access.
- Web Browsers (Firefox, Safari, Chrome, ...):** Access the system via **HTTPS**.
- Linux (Terminal), OS X (Terminal), Windows (PuTTY), WinSCP, FileZilla, ...:** Access the system via **SSH**.
- Login Node:** Receives connections from the Internet and interacts with **Storage Servers** via **NFS**.
- Storage Servers (/home, /glob):** Provide storage for the system.
- Job Queue:** Manages the execution of jobs, receiving input from the Login Node and interacting with the Cloud via **qsub**.
- Cloud:** Consists of multiple servers (server #1 to server #n). Each server can host **Notebook** instances or be in a state **Available for Jobs**. The Job Queue interacts with the Cloud via **qsub** to execute **Computational Jobs**.

Below the diagram, the text "Service Terms" is visible.

Intel® DevCloud – Available FPGA Hardware

- What are you trying to use the DevCloud for?

- 1) Arria™ 10 PAC - RTL AFU, OpenCL
- 2) Arria™ 10 - OneAPI, OpenVINO
- 3) Stratix™ 10 - RTL AFU, OpenCL
- 4) Stratix™ 10 – OneAPI
- 5) Emulation
- 6) Compilation (bitstream creation)

Soon: Intel® Agilex™ cards with Intel® oneAPI





Section: FPGA compute acceleration

Sub-Topics:

- Hardware
- Intel® oneAPI
- Intel® DevCloud
- **FPGAs in HPC**
- FPGAaaS

FPGAs in HPC

Examples

Reasons: Acceleration, Perf/W, Perf/V

Use cases:

- Scientific calculations
- Genomics
- HEP trigger
- Data compression
- Image processing
- Data Base acceleration
- File parsing
- Financial
- ...

Particle Identification on an FPGA Accelerated Compute Platform for the LHCb Upgrade

Christian Färber, Rainer Schwemmer, Jonathan Machen, and Niko Neufeld

Abstract—The current LHCb detector system is being upgraded to a “triggerless” system. The Large Hadron Collider corresponding bandwidth from dedicated computing farms to trigger, has to be increased currently 500 GHz up to 40 Tbps. This will reduce the time to write the interesting physics data to a challenging task. Technologies are considered at different parts of the system. A farm at the event filter field programmable gate array platform is considered and accelerators are used more as for Microsoft Bing search we use hosts a general fast FPGA based via the high-up An acceleration is implemented these platforms, which are built computing, are also very fast. First, the performance of the existing LHCb RICH and is ported to the experiment. We have compared the performance of the existing LHCb RICH and is ported to the experiment. We have compared the performance of the existing LHCb RICH and is ported to the experiment. We have compared the performance of the existing LHCb RICH and is ported to the experiment.

Accelerating Re-Pair Compression using FPGAs
Robert Lauch, Suleyman S. Demirosoy, Norman May, robert.lauch@intel.com, suleyman.demirosoy@intel.com, norman.may@intel.com, Intel Corporation (UK) Limited
Veeharaghavan Ramamurthy, Christian Färber, veeharaghavan.ramamurthy@intel.com, christian.farber@intel.com, Intel Corporation

FPGA-Accelerated Compression of Integer Vectors
Mahmoud Mohsen, Norman May, mahmoud.mohsen@sap.com, norman.may@sap.com, SAP SE
Christian Färber, christian.farber@intel.com, Intel Corporation
David Broneska, david.broneska@ovgu.de, University of Magdeburg

PipeJSON: Parsing JSON at Line Speed on FPGAs
Jonas Dann, Royden Wagner, Daniel Ritter, jfdann@posteo.de, rdwagner@posteo.de, dritter@posteo.de, SAP SE, Waldorf, Germany
Christian Färber, christian.farber@intel.com, Intel Corporation
Holger Fröning, holger.froening@hamburg.de, Heideberg University Heidelberg, Germany

Resource-Efficient Database Query Processing on FPGAs
Mehdi Moghaddamfar, Mehdi.moghaddamfar@sap.com, TU Dresden & SAP SE
Christian Färber, christian.farber@intel.com, Intel Corporation
Wolfgang Lehner, wolfgang.lehner@tu-dresden.de, TU Dresden
Norman May, norman.may@sap.com, SAP SE
Akash Kumar, akash.kumar@tu-dresden.de, TU Dresden

Accelerating Re-Pair Compression using FPGAs
Abstract: Re-Pair is a compression algorithm which requires random accesses to compressed data. This is not supported by hardware accelerators. We propose a hardware accelerator for Re-Pair compression on FPGAs. The accelerator is implemented on a Xilinx Zynq-7010. It achieves a compression ratio of 2.0 and a throughput of 1.2 Gbps. The accelerator is implemented on a Xilinx Zynq-7010. It achieves a compression ratio of 2.0 and a throughput of 1.2 Gbps. The accelerator is implemented on a Xilinx Zynq-7010. It achieves a compression ratio of 2.0 and a throughput of 1.2 Gbps.

FPGA-Accelerated Compression of Integer Vectors
Abstract: An efficient compression algorithm which requires random accesses to compressed data. This is not supported by hardware accelerators. We propose a hardware accelerator for integer vector compression on FPGAs. The accelerator is implemented on a Xilinx Zynq-7010. It achieves a compression ratio of 2.0 and a throughput of 1.2 Gbps. The accelerator is implemented on a Xilinx Zynq-7010. It achieves a compression ratio of 2.0 and a throughput of 1.2 Gbps.

PipeJSON: Parsing JSON at Line Speed on FPGAs
Abstract: JavaScript Object Notation (JSON) is a data exchange and storage format. While modern CPUs show an improved JSON parsing performance, the limited pipelining of CPUs prevent from reaching the practical limit of line speed. We present PipeJSON, the first standard parser to process tens of gigabytes of JSON data on FPGAs. PipeJSON is implemented on a Xilinx Zynq-7010. It achieves a parsing rate of 1.2 Gbps. PipeJSON is implemented on a Xilinx Zynq-7010. It achieves a parsing rate of 1.2 Gbps.

Resource-Efficient Database Query Processing on FPGAs
Abstract: FPGA technology has introduced new ways to accelerate database query processing, that often result in higher performance and energy efficiency. This is due to the unique architecture of FPGAs, which allows for a high degree of parallelism and a high degree of customization. We present a resource-efficient database query processing architecture on FPGAs. The architecture is implemented on a Xilinx Zynq-7010. It achieves a query processing rate of 1.2 Gbps. The architecture is implemented on a Xilinx Zynq-7010. It achieves a query processing rate of 1.2 Gbps.



Section: FPGA compute acceleration

Sub-Topics:

- Hardware
- Intel[®] oneAPI
- Intel[®] DevCloud
- FPGAs in HPC
- **FPGAaaS**

FPGAaaS

Datacenter and cloud

- FPGA accelerator functionality can be offered as a microservice enabling application developers to easily leverage many microservice characteristics:
 - Auto-deployment,
 - Scalability, dynamic configuration
- Data compression as CPU competitive example
- Sharing FPGA with multiple instances important for business case and enables even more use cases later
- IOFS enables virtualization and Docker container available for deployment and easy scaling, OS with DFL driver necessary e.g. GardenLinux
- Availability of Intel® FPGAs

CIDR 2023


DASH: Asynchronous Hardware Data Processing Service

Norman May
Daniel Ritter
Andreas Dostinger
SAP SE
Germany
norman.may@sap.com

Christian Farber
Intel Corporation
United Kingdom
christian.farber@intel.com

Suleyman S. Demireli
Intel Corporation (UK) Ltd.
Germany
suleyman.demireli@intel.com

Abstract
An accelerator used to solve a good balance of high performance at low price. Hardware accelerators promise to accelerate or offload compute-intensive operations from general purpose CPUs. As hardware accelerators like FPGAs become more widely available, e.g. offered by cloud providers or dedicated FPGAs-equipped compute nodes [1, 2], researchers investigate how to apply FPGAs in the context of databases [3]. In this research we demonstrate that FPGAs can help to accelerate database operations to be more energy efficient or to general more cost-efficient. However, prior work assumes a setup where the FPGA is deployed with the database processing server where the FPGA communicates with the host CPU via PCIe or CXL interconnects in a full high bandwidth and data flow between an accelerator and the host (marked green). Other options explored put the FPGA in the path - either in the network or in the data path - which is often limited by latency or bandwidth of the interconnects. In this paper, we propose a new hardware architecture from a different perspective. We focus on highly compute-intensive database operations that are highly compressible and where the transfer time of the data and completion time of the tasks related to the results time are low. Examples of such operations include reorganizing or compressing the persistent data, collecting statistics or secondary checks. With the server-side approach under the creative license of a commercial CPU, we show how to use the FPGA to accelerate such a task. We demonstrate that the proposed architecture can be implemented on a hardware accelerator of the Intel® Open FPGA Stack. The hardware has been tested on Intel® Stratix 10 K10, 10K10, and 10K10B.




1 Introduction
Databases need to strike a good balance between high performance at low price. Hardware accelerators promise to accelerate or offload compute-intensive operations from general purpose CPUs. As hardware accelerators like FPGAs become more widely available, e.g. offered by cloud providers or dedicated FPGAs-equipped compute nodes [1, 2], researchers investigate how to apply FPGAs in the context of databases [3]. In this research we demonstrate that FPGAs can help to accelerate database operations to be more energy efficient or to general more cost-efficient. However, prior work assumes a setup where the FPGA is deployed with the database processing server where the FPGA communicates with the host CPU via PCIe or CXL interconnects in a full high bandwidth and data flow between an accelerator and the host (marked green). Other options explored put the FPGA in the path - either in the network or in the data path - which is often limited by latency or bandwidth of the interconnects. In this paper, we propose a new hardware architecture from a different perspective. We focus on highly compute-intensive database operations that are highly compressible and where the transfer time of the data and completion time of the tasks related to the results time are low. Examples of such operations include reorganizing or compressing the persistent data, collecting statistics or secondary checks. With the server-side approach under the creative license of a commercial CPU, we show how to use the FPGA to accelerate such a task. We demonstrate that the proposed architecture can be implemented on a hardware accelerator of the Intel® Open FPGA Stack. The hardware has been tested on Intel® Stratix 10 K10, 10K10, and 10K10B.

Figure 1: Compute topology in the cloud and data centers with FPGAs. From the context of this paper is marked red.

related architecture of the cloud native database services, such tasks could be deployed on accelerators that are deployed on the server as accelerator-based services. As marked red in Figure 1, these accelerator attached nodes are connected to other parts of the database engine via network links. Expanding the FPGA operations as a service enables sharing the FPGA between multiple database instances and using the FPGA based services flexibly based on the demand.

As we discuss in more detail in Section 2, this setup is consistent with the trend to build cloud-native databases that factor database operations into independent services, e.g. compute and storage, optimizer services etc. [4, 5, 26]. Database implementations seek to maximize the hardware costs to support database workloads through the separation of components, and at the same time, it allows them to scale components independently. As a typical consequence of this development, we argue that certain components can be deployed on specialized hardware. While, as typical, such asynchronous operations could also be executed on CPUs, FPGAs - or hardware - for the specific task CPUs or CPUs - may offer better performance and energy efficiency than a CPU making them attractive for the server-side architecture. In this paper, we propose an alternative perspective that can utilize the inherent pipelining and parallelization of the FPGA. With this we aim to reorganize the database operations in a high-degree data flow on an accelerator where prior work had only an unifying compression related to the transaction processing model or supported workload [7, 18, 25].

<https://www.cidrdb.org/cidr2023/papers/p6-may.pdf>

White Paper 

Intel® FPGAs
Intel® Open FPGA Stack

SAP Prototypes a Containerized Compression Workload Leveraging the Intel® Open FPGA Stack (Intel® OFS) Infrastructure

Computationally intensive workloads can benefit from FPGA-based acceleration. To expedite this development, SAP employs the Intel® OFS hardware and software infrastructure.

Interested in Developing a Containerized Application with Intel® OFS?
Read this white paper to learn how Intel partners, SAP, used an Intel® FPGA-based acceleration platform and Intel® Open FPGA Stack to prototype development of a general-purpose (GPG) accelerator in a Docker container.

Executive Summary
Developers at SAP SE are creating a proof-of-concept (POC) of cloud-based Compression as a Service (CaS). They chose to leverage FPGAs to accelerate this computationally intensive workload, and plan to run their workload in Docker containers on their high-performance Analytics Appliance (HANA) cloud using their distribution of the Garden Linux operating system (OS). Using Intel® Open FPGA Stack, or Intel® OFS, they leverage a scalable, secure-accessible hardware and software infrastructure designed to meet custom board and workload development. The flexibility provided by Intel® OFS enables them to develop a POC in their preferred configuration, and the modular architecture expedites their development.

Background
SAP SE is a German multinational software corporation based in Walldorf, Baden-Württemberg, that develops enterprise software to manage business operations and customer relations.

SAP relies on relational database management system developed and marketed by SAP SE. Columnar data in SAP HANA is encoded with dictionaries. String dictionaries may contain vast amounts of textual data that needs to be compressed to minimize memory requirements. There are many different compression algorithms available (LZ77, LZ78, LZSS, LZMA, etc.), but these algorithms are typically not designed to compress large quantities of information into a block. If a dictionary is compressed this way, the entire file would have to be decompressed to access a single entry, which would be extremely inefficient in terms of time, computation, and power consumed. Alternatively, using these algorithms to compress each dictionary entry individually would present its own inefficiencies because they are not optimized to compress small amounts of data.

Authors
Dr. Christian Farber
SAP SE
Tamas Lin
Product Marketing Specialist
Intel Program Manager, Intel Data Center Group
Vincent Riesop
Linux Developer
SAP SE

<https://www.intel.com.br/content/dam/www/central-libraries/us/en/documents/2022-09/sap-open-fpga-stack-white-paper.pdf>



Section: Use cases

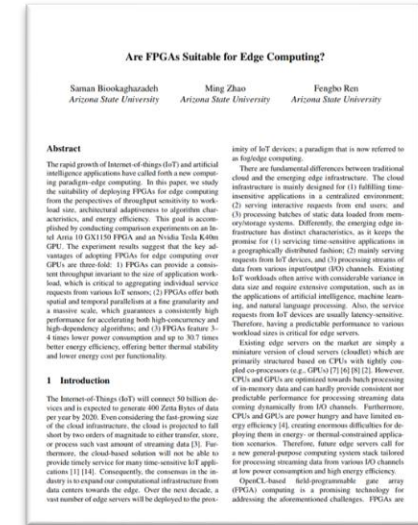
Sub-Topics:

- FPGAs and Edge Computing

FPGAs and Edge Computing

Examples

- Reduce data transfer by compute and decide at the edge
- Benefits of FPGAs:
 - Good fit for customized performant edge systems
 - Easy scalable perf., low latency, low power, thermal stability
- Applications:
 - Internet of Things (IoT)
 - Mobile and 5G
 - Industrial automation
 - Video analytics
 - Autonomous vehicles



<https://www.usenix.org/system/files/conference/hotedge18/hotedge18-papers-bhookhagazadeh.pdf>



<https://www.intel.com/content/www/us/en/products/docs/programmable/cloud-connectivity-solution-brief.html>



Section: FPGA Hardware

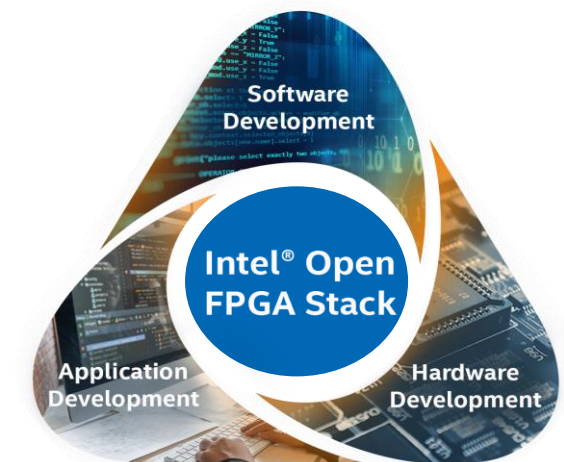
Sub-Topics:

- Intel® Open FPGA Stack
- CXL Interface
- FPGA based IPUs

Intel® OFS – Enabling Scale and Deployment

Intel OFS is a **software and hardware infrastructure** providing an efficient approach to develop a custom FPGA-based platform or workload using an Intel, 3rd party, or custom board.

- Scalable, source-accessible hardware and software framework delivered through **Git repositories**
- **Reduce development time** with **modular** and composable source code used as-is or easily customized
- **Upstreamed Linux kernel drivers** are being adopted by leading OS and orchestration vendors
- **Growing ecosystem** of Intel OFS-enabled boards, workloads, and OS distributions



Intel® OFS Deliverables

Hardware

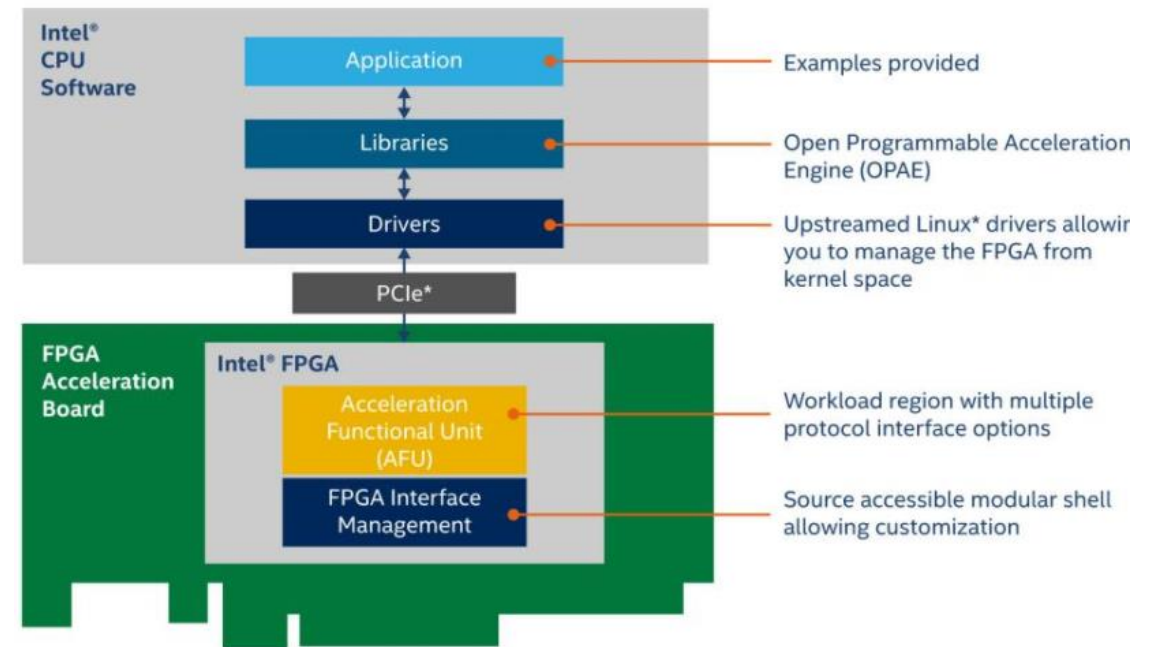
- Acceleration Functional Unit (AFU) Region for Workload Development with Sample AFUs
- FPGA Interface Manager (FIM)
- Board Management Controller (BMC)
- HLD enablement

Software

- Upstreamed, open-source kernel drivers
- OPAE libraries, tools and APIs
- Example Applications

Verification Environment

- UVM Verification environment provided through Git repositories



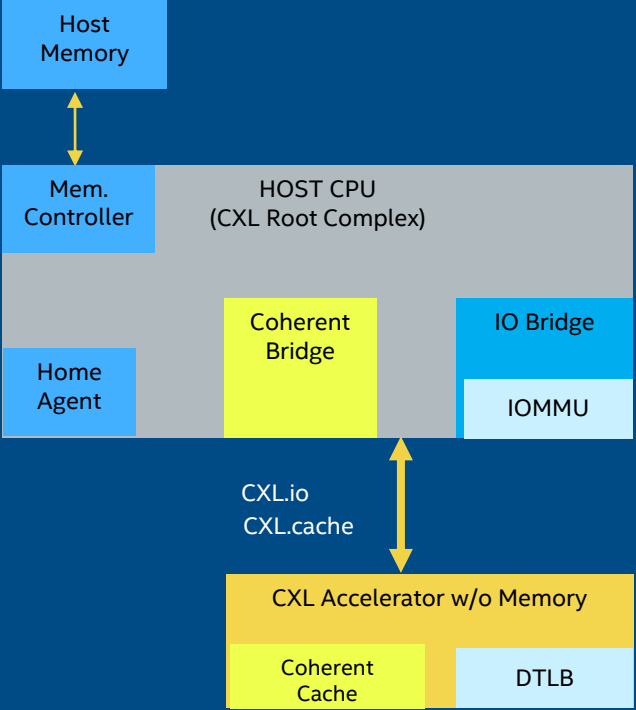


Section: FPGA Hardware

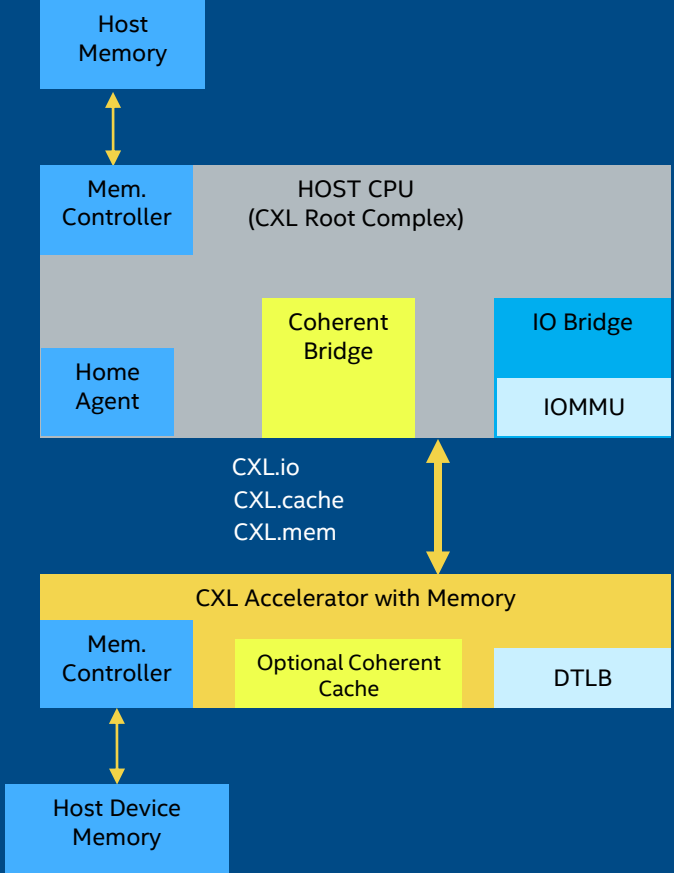
Sub-Topics:

- Intel® Open FPGA Stack
- **CXL Interface**
- FPGA based IPUs

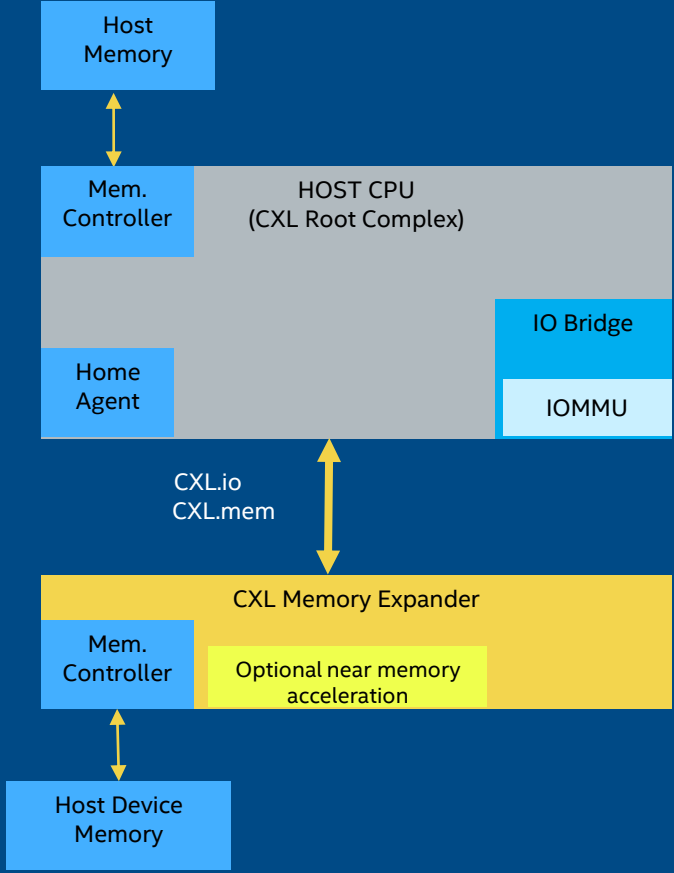
CXL Configuration Types - Type 1, 2, 3



**CXL Type 1
Accelerator Device**

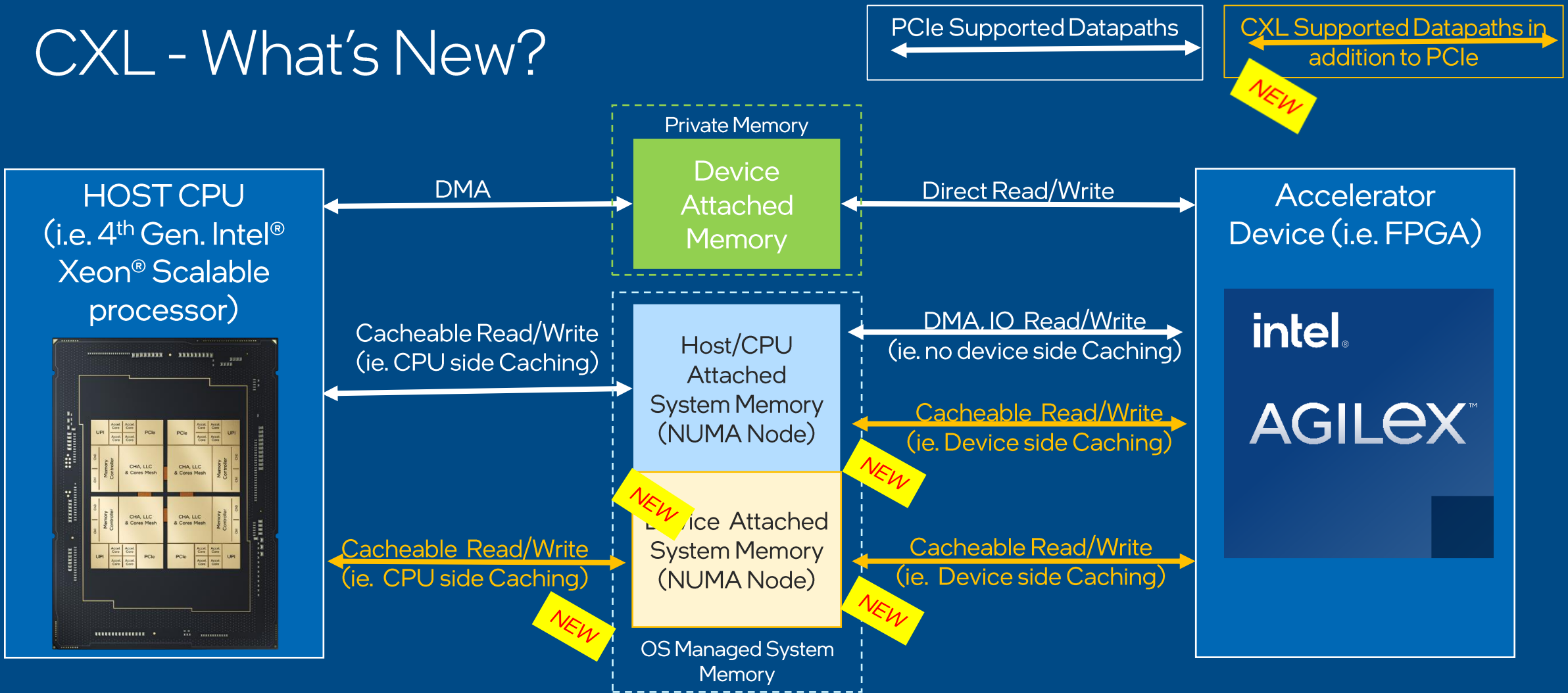


**CXL Type 2
Accelerator Device**



**CXL Type 3
Memory Expansion Device**

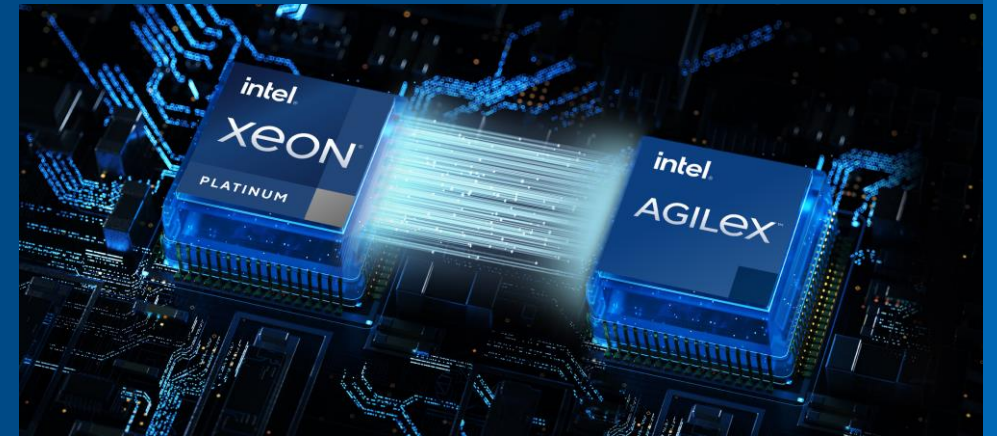
CXL - What's New?



CXL-BASED ACCELERATORS - ELEVATED TO 1ST CLASS PRIORITY

Intel® Agilex™ FPGAs – The 1st FPGA with CXL Hard IP

- FPGA Industry-leading interconnect performance
 - Hard IP for CXL and PCIe 5.0 (x16 lanes)
- Supports CXL across multiple CPU / chipset suppliers
 - CXL v1.1 (now) and v2.0 (future IP release, same silicon)
 - Support for Type 1 and Type 2 Accelerators
 - Support for Type 3 memory expansion
 - Additional algorithm acceleration options may not be available from 3rd party ASSP's / ASIC's



1. Intel estimates based on Intel Agilex FPGA with CXL hard/soft IP bandwidth per port vs. competitors FPGAs using 3rd party soft IP CXL controller.

2. Based on PCI-SIG integrator list results for PCIe 5.0 compliance performance. Intel Agilex R-Tile Gen 5 x16 @ 32 GT/s vs. Xilinx Versal Premium ACAP CPM5 Gen 5 x16 @ 16 GT/s.



Section: FPGA Hardware

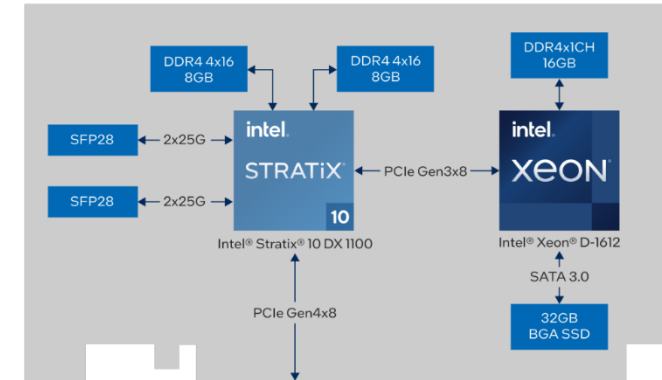
Sub-Topics:

- Intel® Open FPGA Stack
- CXL Interface
- **FPGA based IPU**s

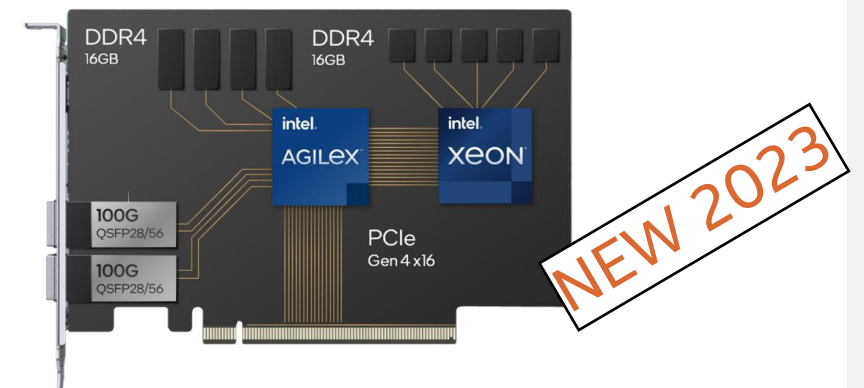
FPGA based IPU

Features	Target Acceleration Workloads
<ul style="list-style-type: none"> • 2 x 25 GbE connectivity • Intel® Stratix™ 10 DX FPGA • Intel® Xeon™ D-1612 Processor • 32GB DRAM • RTL 	<ul style="list-style-type: none"> • Packet processing • Open vSwitch
<ul style="list-style-type: none"> • 2 x 100 GbE connectivity • Intel® Agilex™-F FPGA • Intel® Xeon™ D-1736 Processor • 32GB DRAM • RTL/P4 	<ul style="list-style-type: none"> • Packet processing • Open vSwitch • NVMe-oF/RoCEV2 • RDMA

Intel® IPU Platform C5000X-PL



Intel® IPU Platform F2000X-PL



A coach talks to you, a mentor talks with you, and a sponsor talks about you

(roles may overlap)



COACH

A coach provides guidance for your development, often focused on soft skills (e.g., active listening) rather than technical skills.

Who Drives the Relationship?
You and your coach are responsible for driving the relationship—you can reach out to your coach when you need help, but your coach can also reach out to you.

Actions
Provide development feedback outside the formal performance evaluation process.



MENTOR

A mentor informally or formally helps you navigate your career, providing guidance for career choices and decisions.

Who Drives the Relationship?
You drive the relationship. Your mentor is reactive and responsive to your needs.

Actions
Help you determine possible career paths to meet specific career goals.



SPONSOR

A sponsor is a senior leader or other person who uses strong influence to help you obtain high-visibility assignments, promotions, or jobs.

Who Drives the Relationship?
The sponsor drives the relationship, advocating for you in many settings, including behind closed doors.

Actions
Advocate for your advancement and champion your work and potential with other senior leaders.

Based on catalyst.org guidance

@addyosmani