FPGAs and the Cloud – An Endless Tale of Virtualization, Elasticity and Efficiency

13th HiPEAC Workshop on Reconfigurable Computing (WRC) // València // January 21st, 2019

Dr.-Ing. Oliver Knodel // o.knodel@hzdr.de

HZDR

HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Our Background @ HZDR

Main Challenge: Pre-/post-processing and archiving of research data

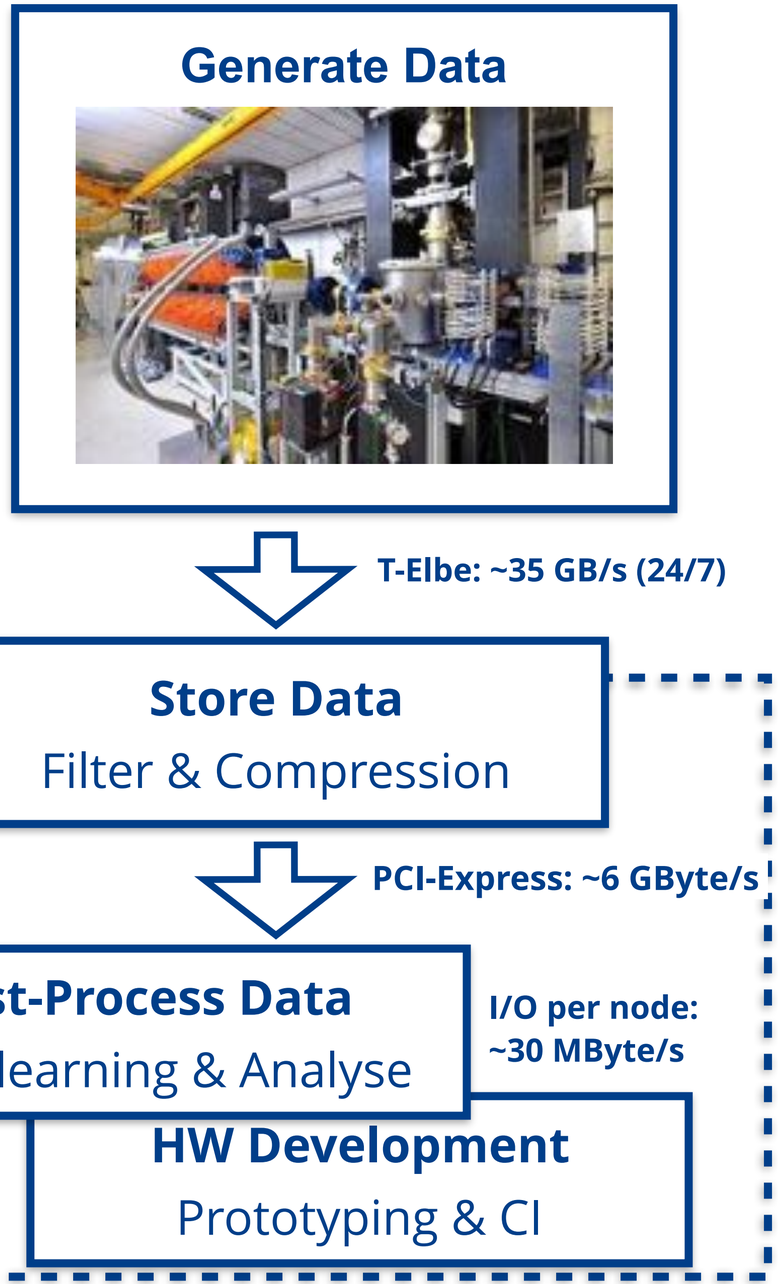
- Filter and compress measured or simulated data at the *edge* of the datacenter (Cloud).
- Accelerate compute-intensive tasks with dedicated low-latency, high-bandwidth hardware.

HPC and hardware acceleration (OpenCL)

- Many research questions require compute intensive deep learning approaches suitable for GPUs and FPGAs.
- The research data is located in the data centre anyway.

Prototyping and Continuous Integration (CI)

- The custom FPGA designs have to be tested and verified with every development cycle to meet high requirements.
- FPGAs in the datacenter can be used for that during idle times.



Structure

1. Introduction and motivation
2. Requirements analysis and system design
   - System architecture of the cloud — RC3E
   - Virtualization of the FPGAs — RC2F
3. Results of the prototypical implementation
   - Required FPGA-Resources
   - Behavior simulation of a FPGA-Cloud
4. Outlook: The future FPGA-Cloud at HZDR


I. Requirements and stakeholder analysis
II. System architecture of the cloud — RC3E
III. Virtualization of the FPGAs — RC2F

Requirements and Stakeholder — Provision, Security, Service Models

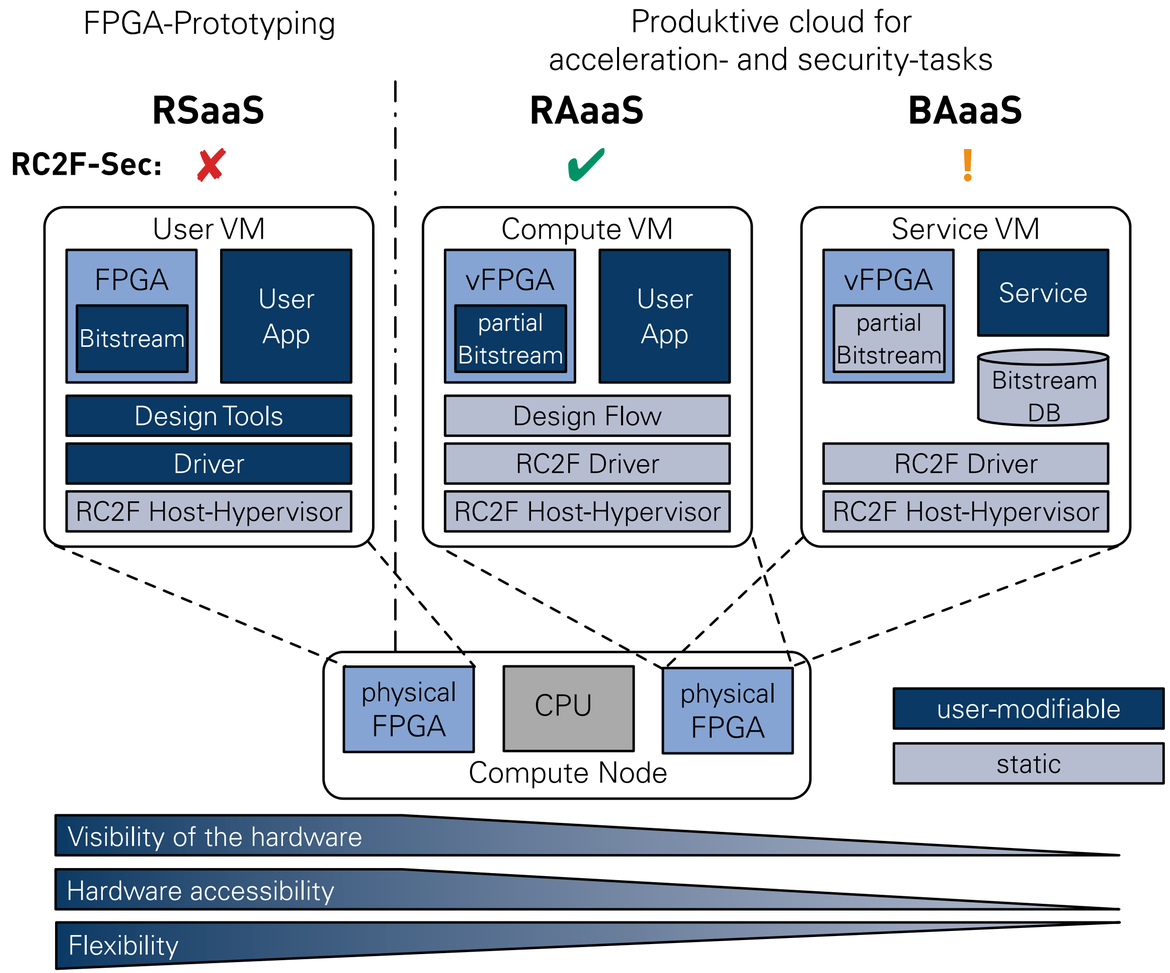
Approach

— The user groups require a different level of visibility and virtualization.

— With division into service models, this abstraction can be integrated into a resource management system (RC3E).

Service-Models

1. Reconfigurable Silicon as a Service (RSaaS).
2. Reconfigurable Accelerator as a Service (RAaaS).
3. Background Acceleration as a Service (BAaaS).

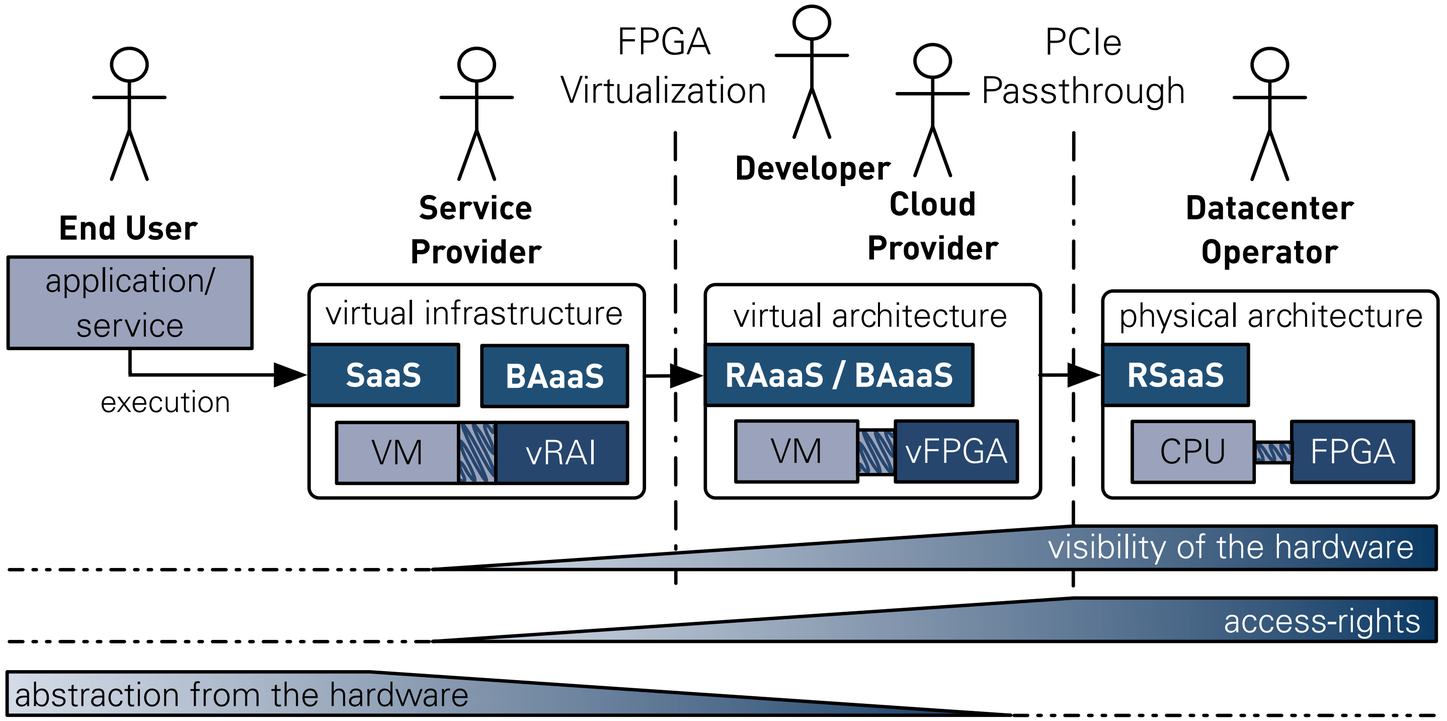


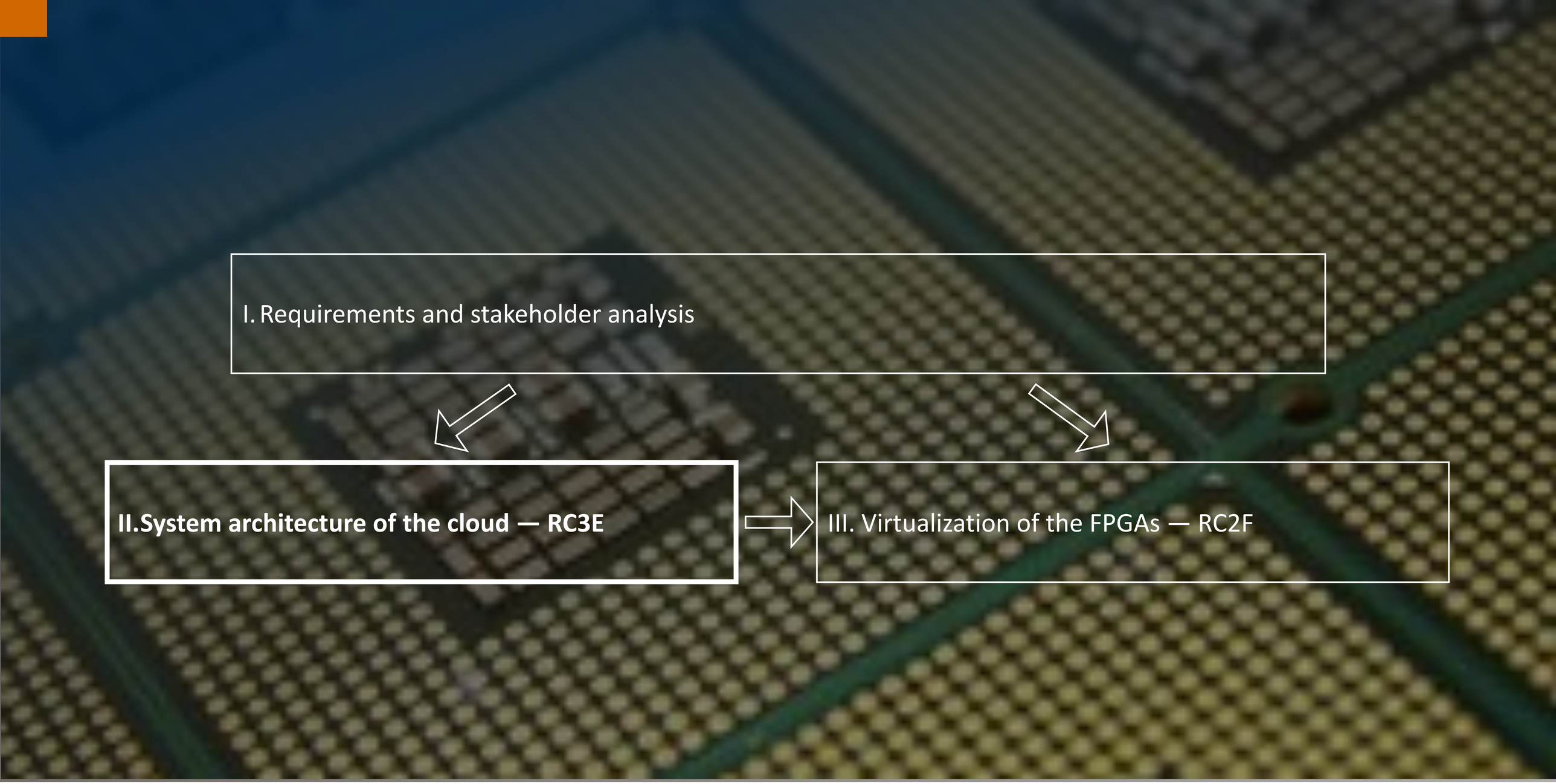
Requirements and stakeholder — The user in the context of Cloud-FPGAs

Objectives

— Flexible deployment for different groups of people with different needs.

— The FPGA is not visible to the enduser and only virtual resources are used by the service providers.



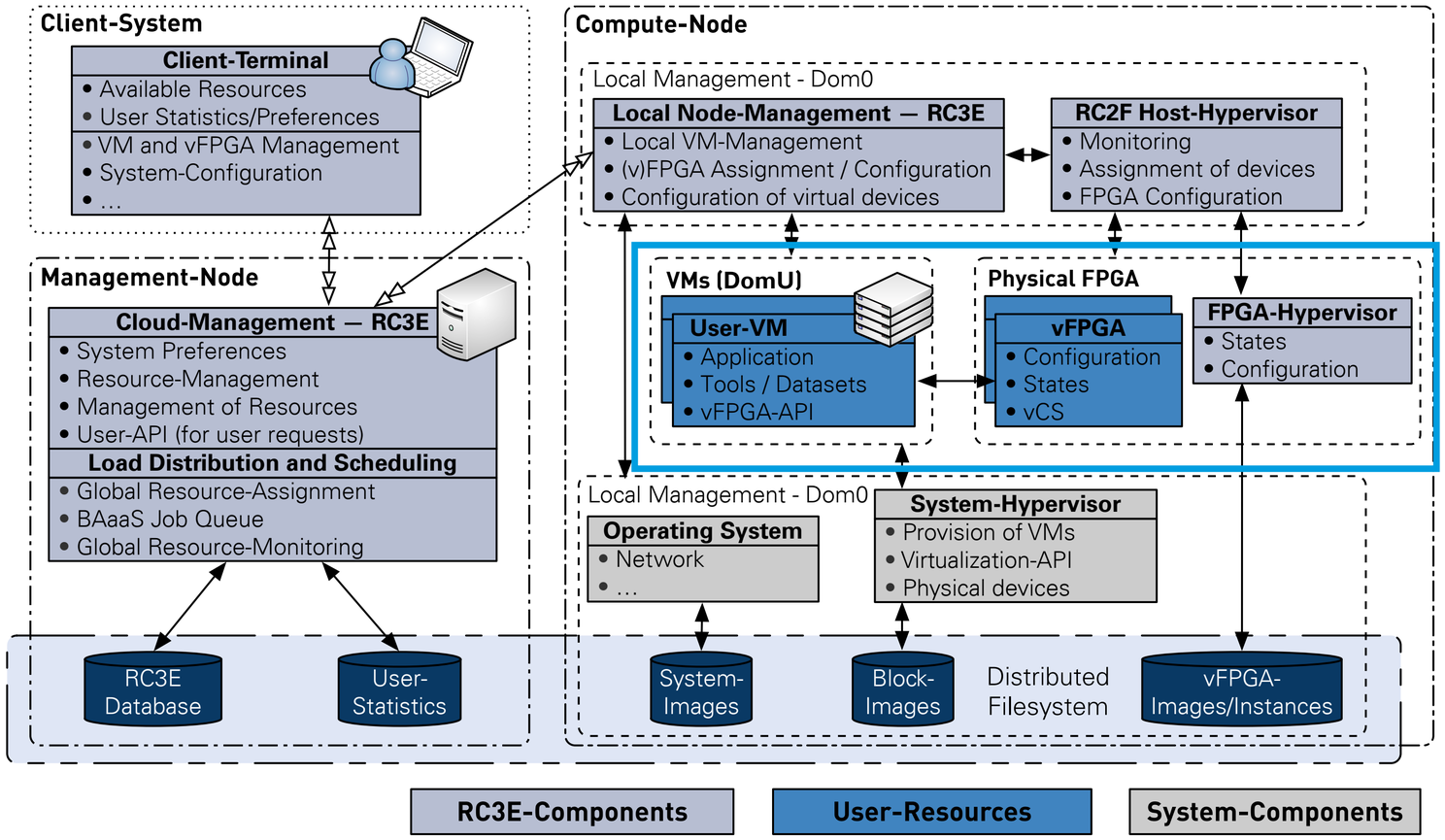

I. Requirements and stakeholder analysis
II. System architecture of the cloud — RC3E
III. Virtualization of the FPGAs — RC2F

II. RC3E System architecture — Architecture of the RC3E prototype

— View of the hardware architecture:
  - The cloud consists of nodes with one CPU each and (two) PCIe-coupled FPGAs.
  - Each physical FPGA can contain multiple virtual FPGAs.

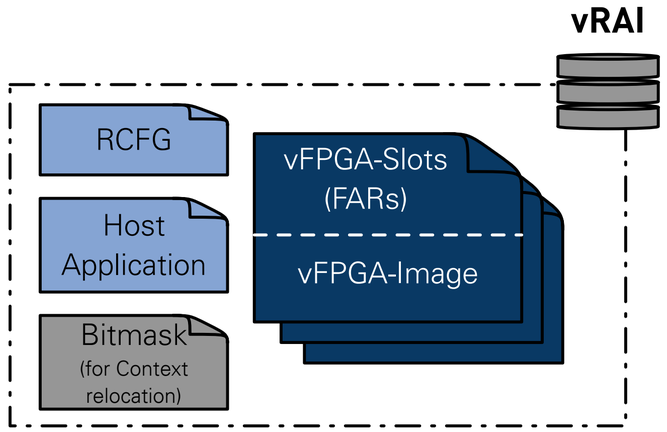
— Manage resources (vFPGA and VMs) across all cloud nodes according to service models.

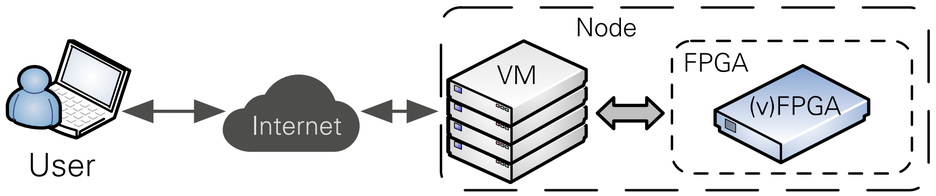
— Required components:
  - Local node-management,
  - RC2F host-hypervisor and
  - FPGA-hypervisor (located on the FPGA device)



II. RC3E System architecture — FPGA Images (vRAIs)

- All necessary files for a background accelerator are encapsulated in **v**irtual **R**econfigurable **A**cceleration **I**mages (vRAIs).
- The vFPGA itself is described in the form of a **R**econfigurable (Device) **C**on**fig**uration (RCFG).
- A simple VM with hardware accelerator can be easily described:

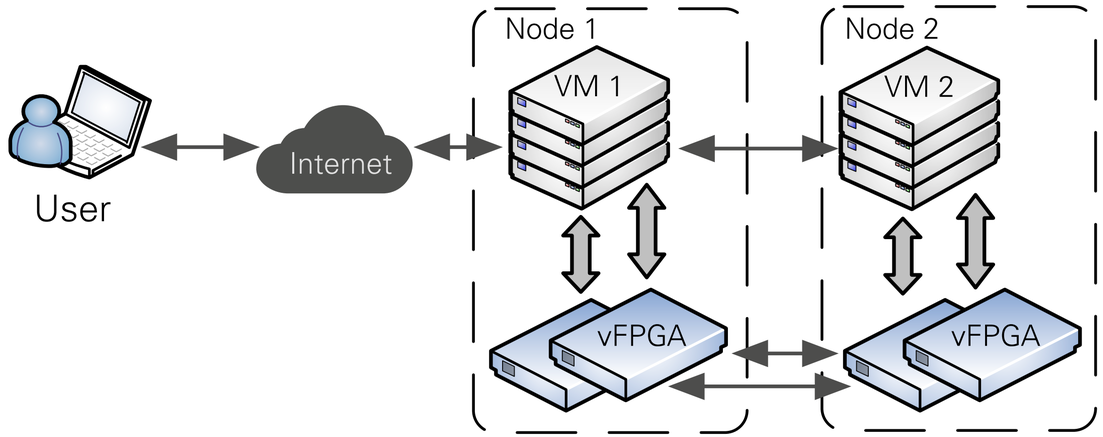


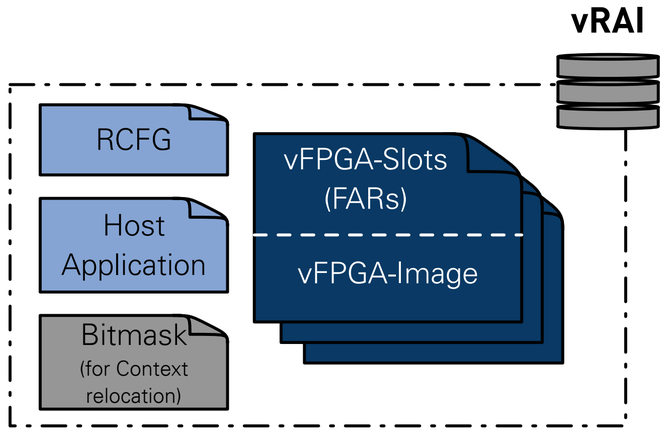



```
service = 'ba'                          #Service Model BAaaS
name = ['vfpga-kmeans']                 #vFPGA/User Design Name
vm = ['vm1-pvm']                        #VM-Instance Name

vfpga = [1]                             #Number of vFPGAs
size = [4]                              #vFPGA-Slots
memory = [4000]                         #DDR-Memory Size

boot= ['idle']                          #Initial vFPGA-State
design = ['kmeans-quad.vrai']           #Initial Design
```

Example of a RCFG for a vFPGA in model BAaaS.

II. RC3E System architecture — Description of a more complex vRAI

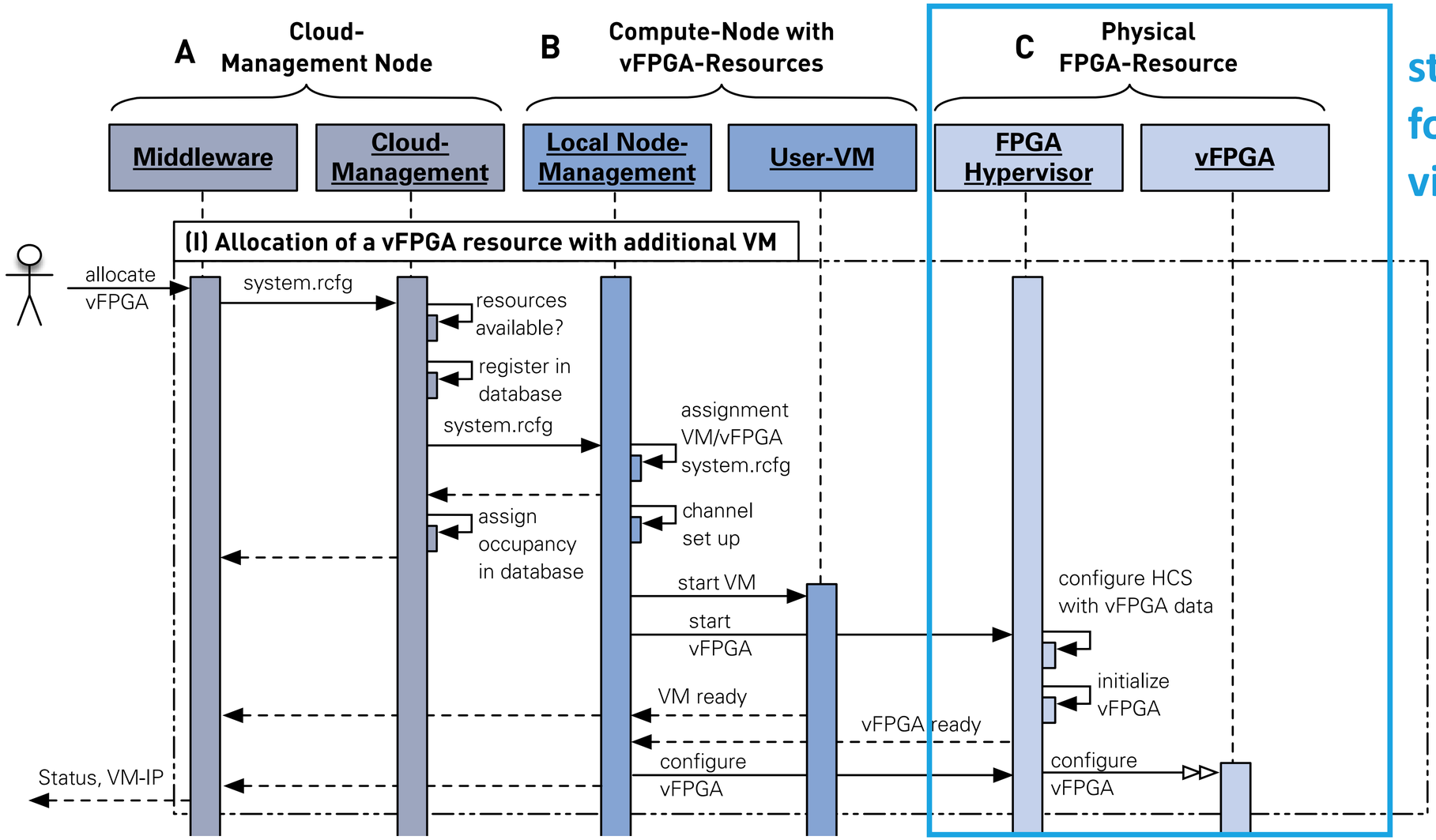




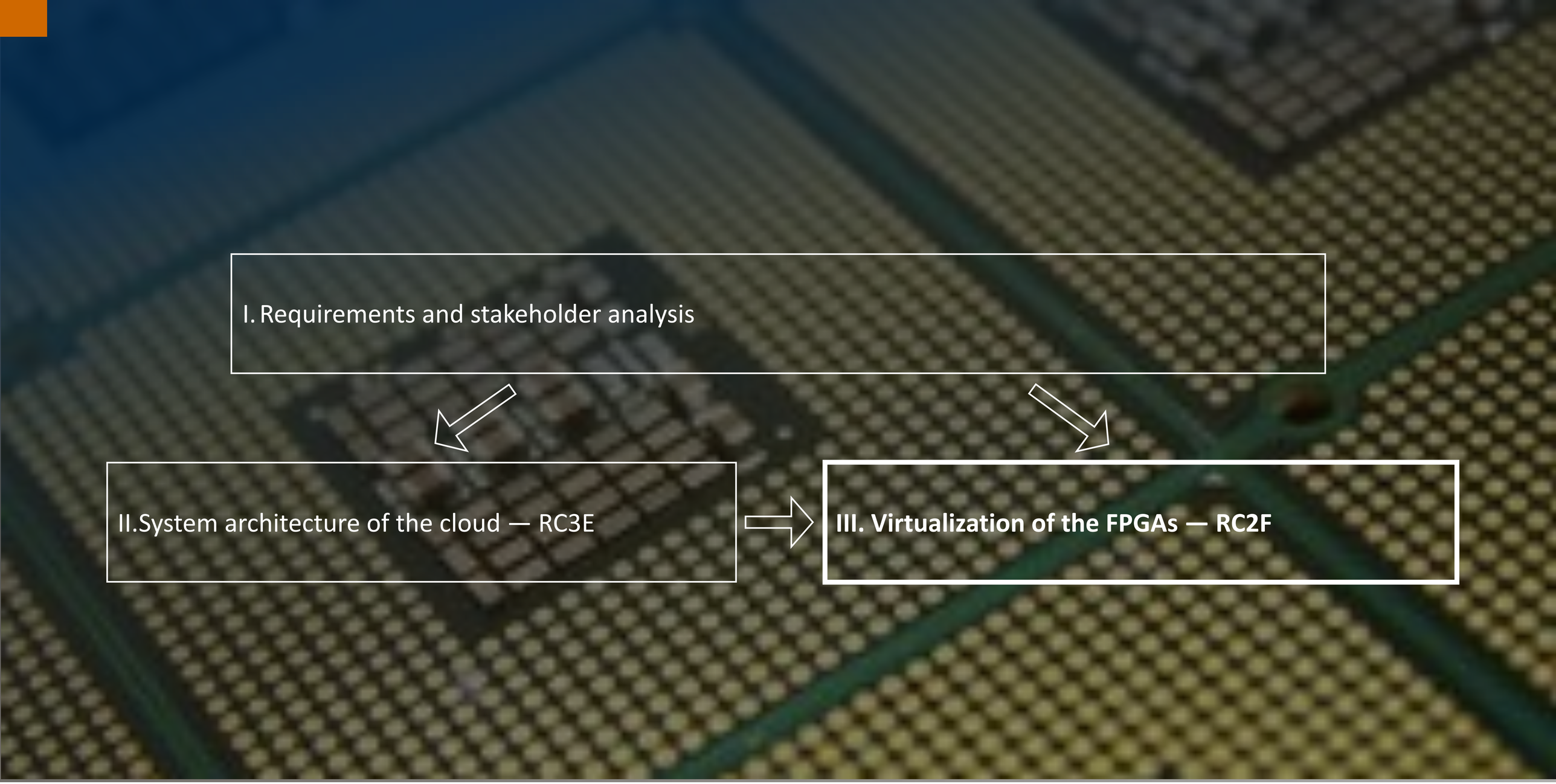

```
service = 'rs'                                  #Service Model RSaaS
name = ['cluster']                              #vFPGA/User Design Name
vm = [‘vm1’, 'vm2']                             #VM-Instance Name

vfpga = [2, 2]                                  #Number of vFPGAs
size = [6, 6]                                   #vFPGA-Slots
memory = [2000]                                 #DDR-Memory Size
debug = 'csp'                                   #Debug Interface
vif = [‘ip=10.0.0.42’; ‘ip=10.0.0.45’]          #IP-Adress-Range

boot= [‘idle']                                  #Initial vFPGA-State
```

Example of a RCFG describing a vFPGA cluster with two host VMs.

II. RC3E System architecture — Interactions between components



starting point for RC2F virtualization

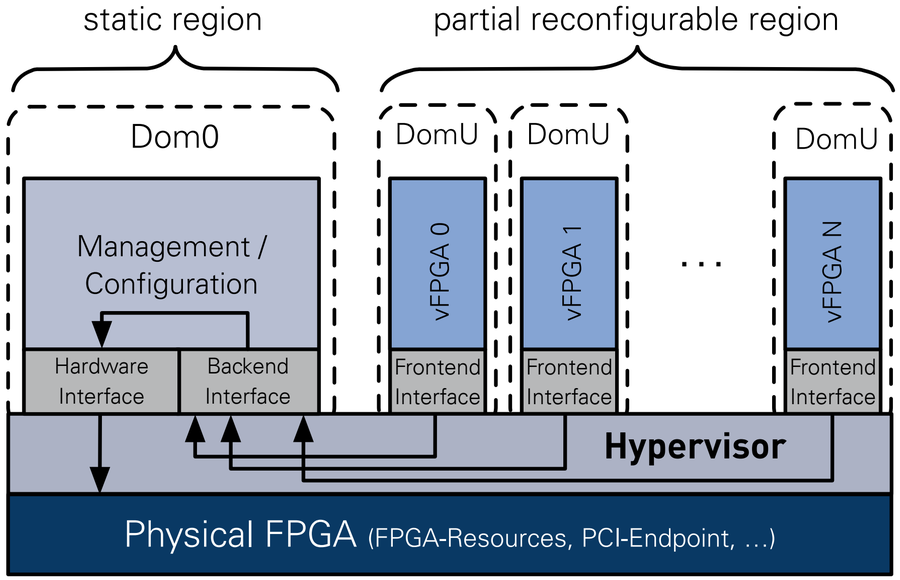

I. Requirements and stakeholder analysis
II.System architecture of the cloud — RC3E
III. Virtualization of the FPGAs — RC2F

III. RC2F Virtualization — V*irtualization* of the FPGA

Concept for the virtualization of FPGAs (approach based on classical System-VMs):

- The user core is located directly (bare-metal) on the FPGA-resources within a dynamically reconfigurable region, the virtual FPGA (vFPGA).
- The static part of the FPGA contains management structures and the physical interfaces.
- The concept is equivalent to traditional virtual machines:

Native Typ 1 System-/Bare-Metal-Virtualization with same ISA



Interface-Virtualization:

- External devices are accessed through **paravirtualization** within the FPGA hypervisor (VMM).
- The interfaces of the vFPGAs are the **frontends**, which are connected to the **backend** in the static region.

III. RC2F Virtualization — Definitions

— A **vFPGA** is perceived by the user as a stand-alone resource with a dynamic number of hardware resources (slices, LUTs, registers, etc.).

— A vFPGA is mapped to **vFPGA-Slots** — smallest possible physical regions with a fixed number of hardware resources.

— A **vFPGA-Design** is the hardware design within a vFPGA.

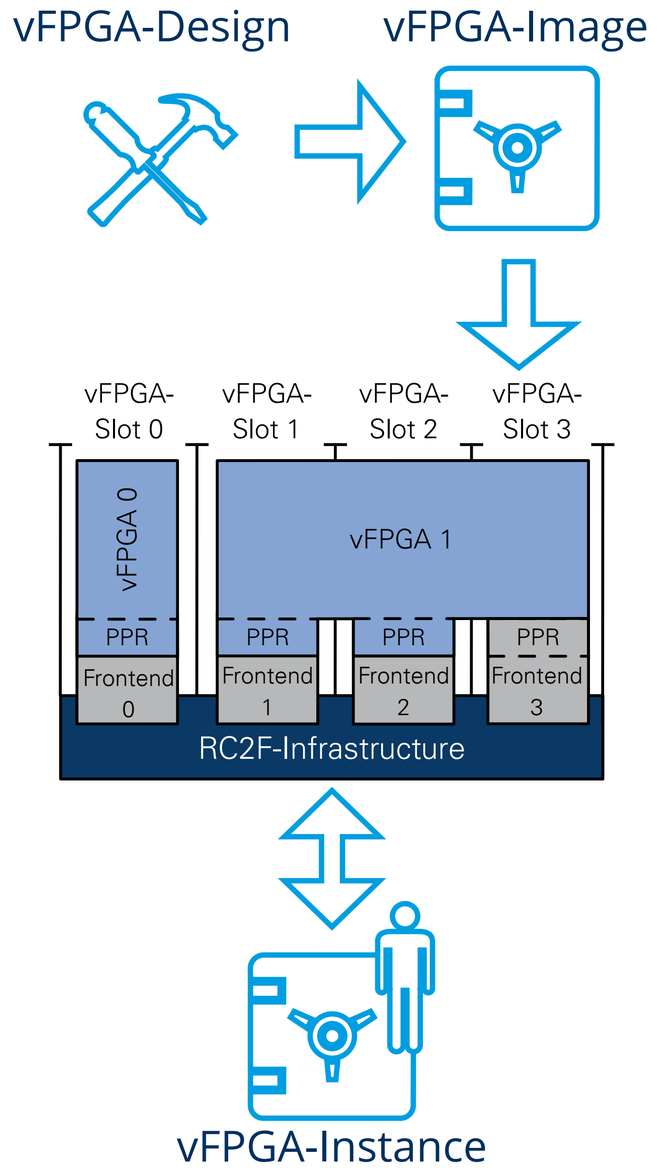
— A partial Bitstream is called **vFPGA-Image** and

— if this is assigned to a user this becomes a **vFPGA-Instance.**

Distinctions in the term “Hypervisor”:

— The management structure for the vFPGAs on their host system is called the **RC2F Host-Hypervisor**.

— The **FPGA-Hypervisor** is the management structure on the FPGA (Dom0).



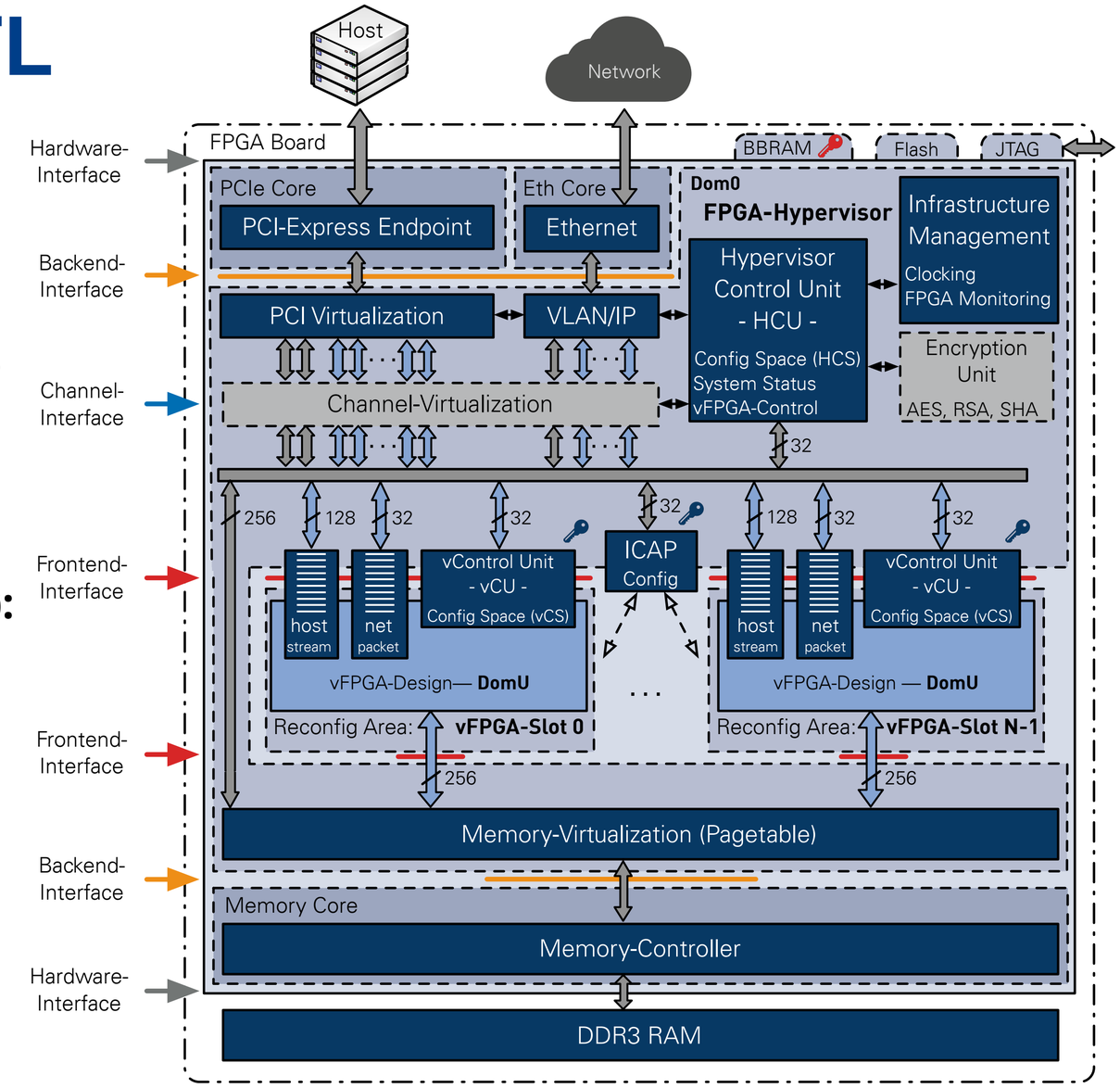
III. RC2F Virtualization — RTL

Provision of vFPGAs on the physical FPGA-device:

- Multiple vFPGAs within vFPGA slots.
- The ICAP is used for dynamic partial reconfiguration.
- Direct management of the states of the vFPGAs on the FPGA within the vControl Unit (**vCU**).

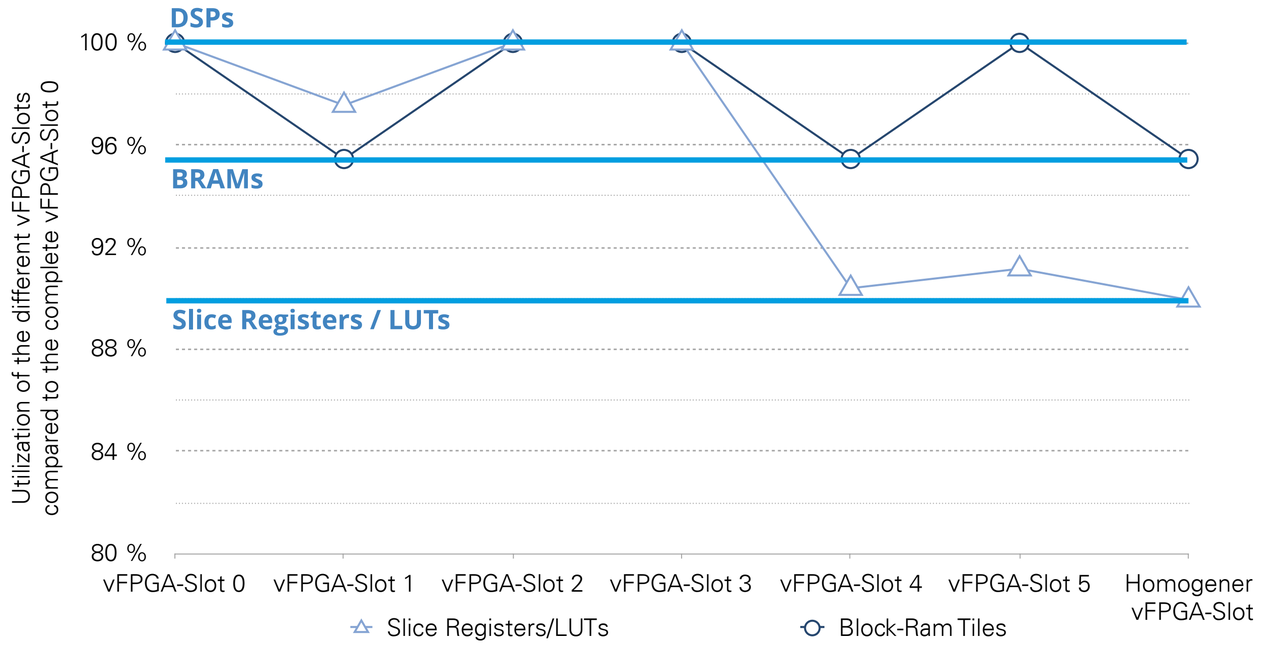
Components of the infrastructure (FPGA-Hypervisor):

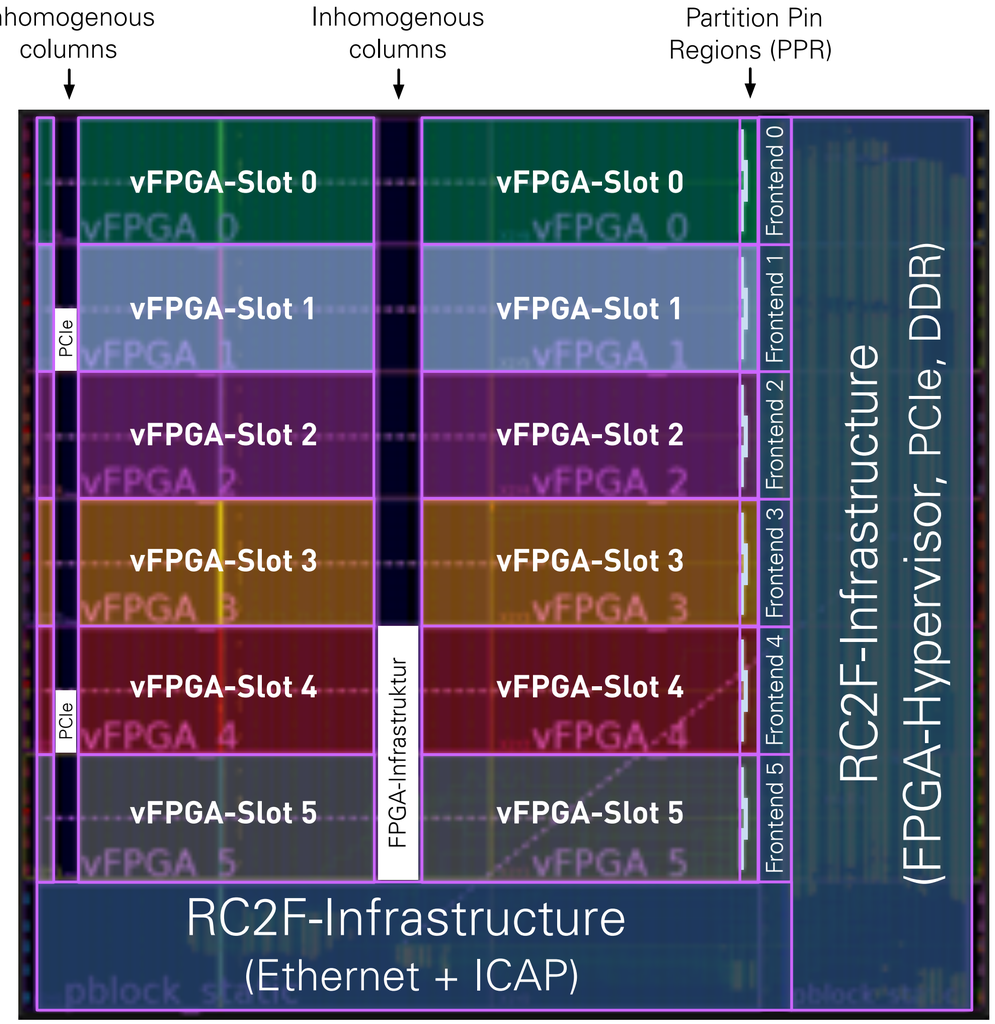
- Internal system bus used for the Paravirtualization.
- Hypervisor Control Unit (**HCU**) to monitor the physical FPGA infrastructure.
- **PCIe** and **Ethernet**-Interfaces.
- Access to **DDR3 memory** on the FPGA board via page tables.



III. RC2F Virtualization — Homogeneous vFPGA-Slots

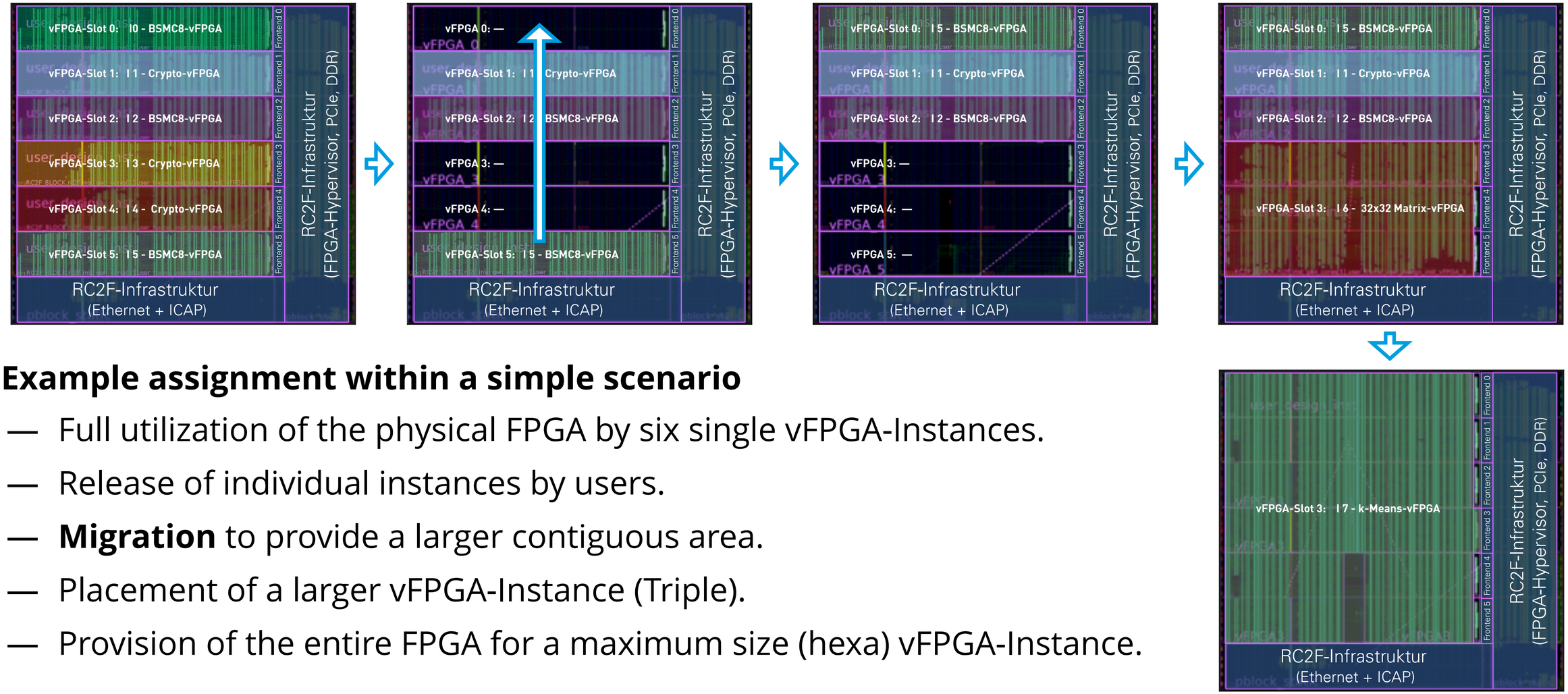
- Establishment of vFPGA-Slots within the clock regions.
- Realization of **homogeneous** vFPGA-Slots on the physical FPGA to enable **migration** of vFPGA-Instances.
- The distribution is dictated by the inhomogeneous FPGA architecture.





(Virtex-7 XC7VX485T)

III. RC2F Virtualization — Application scenario

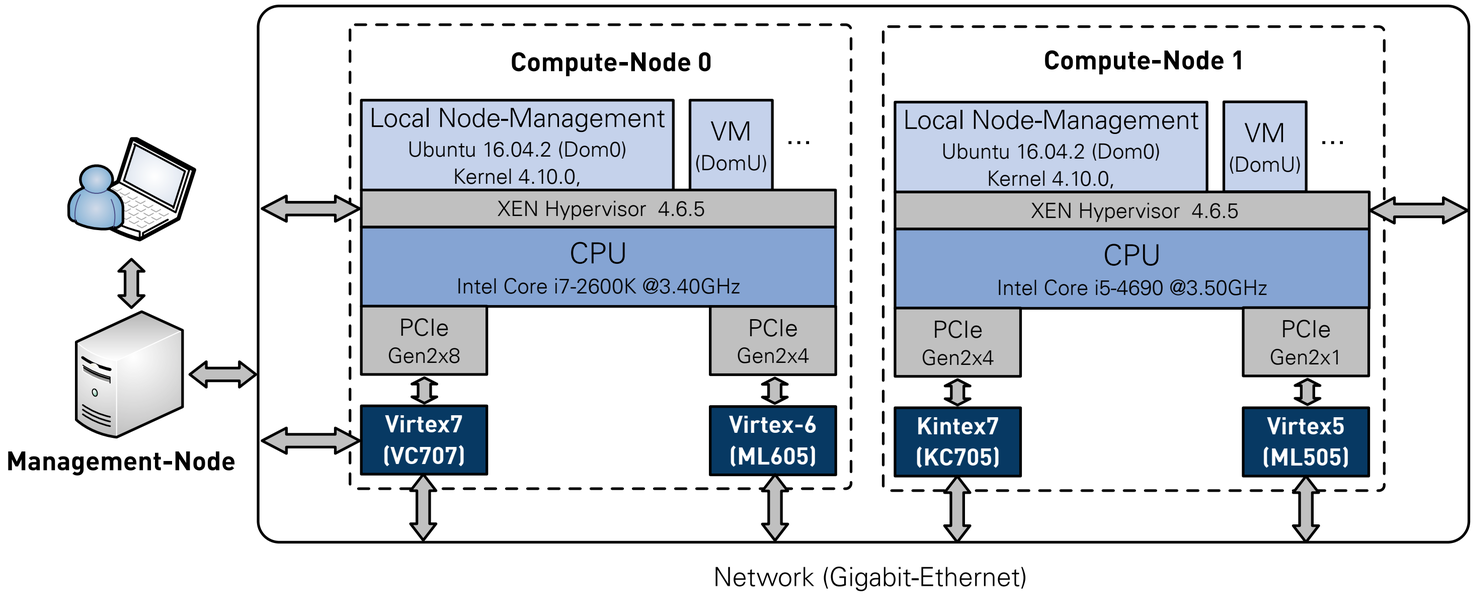


Example assignment within a simple scenario

- Full utilization of the physical FPGA by six single vFPGA-Instances.
- Release of individual instances by users.
- **Migration** to provide a larger contiguous area.
- Placement of a larger vFPGA-Instance (Triple).
- Provision of the entire FPGA for a maximum size (hexa) vFPGA-Instance.

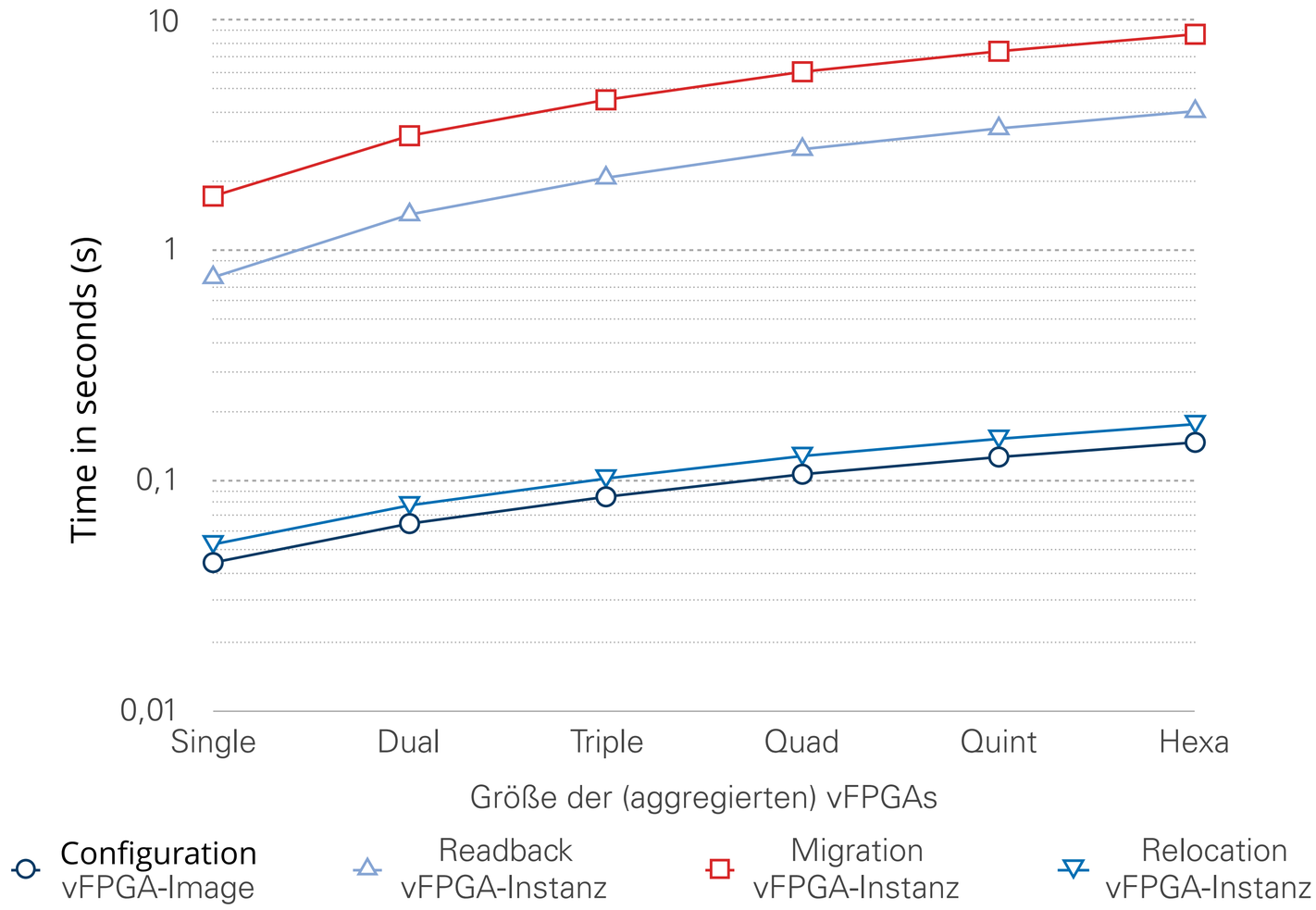
Results of the prototypical implementation of the RC3E & RC2F

Hardware prototype @ TU Dresden



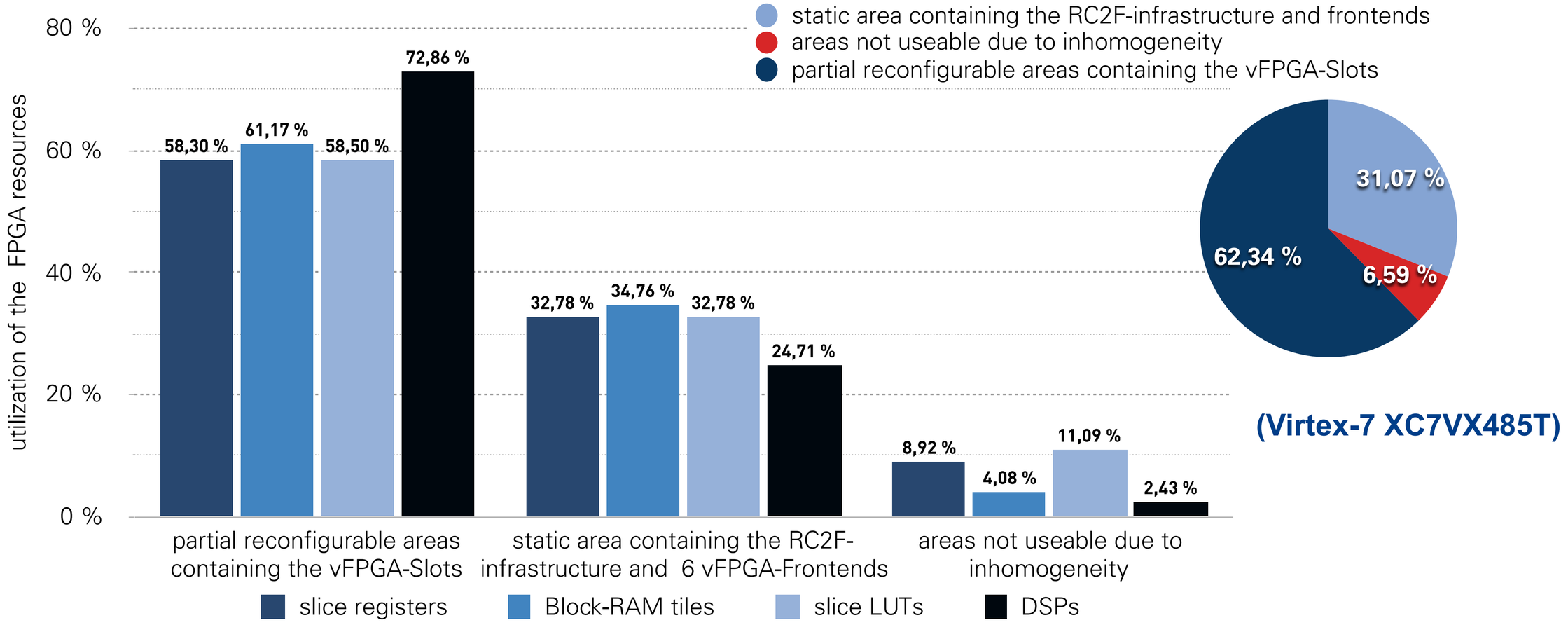
- Structure of the cloud prototype includes different FPGAs.
- Used as a system for development (RSaaS) and teaching (RAaaS).
- For the evaluation of a larger system (BAaaS) a RC3E-Simulation was realized.

Results of the prototypical implementation — Migration



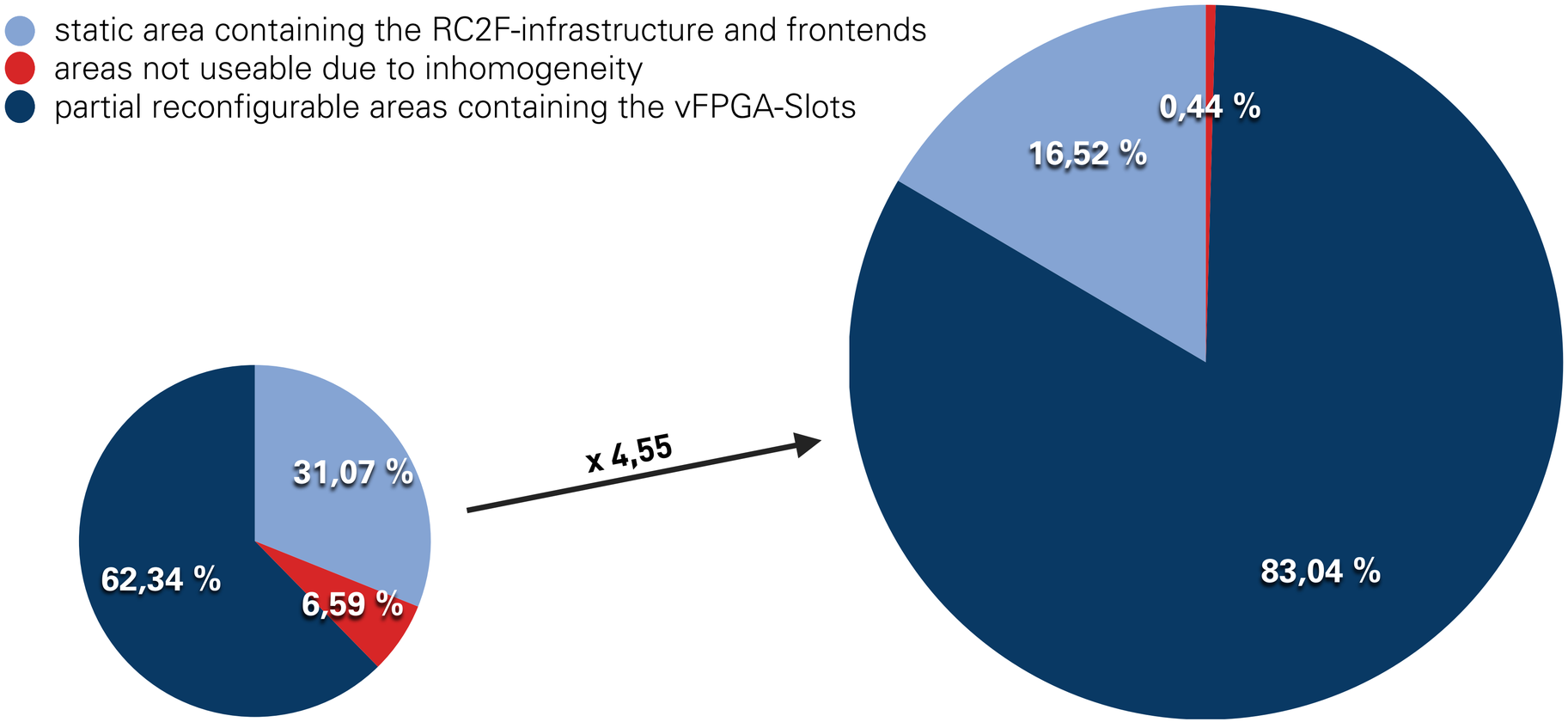
(Virtex-7 XC7VX485T)

Results of the prototypical implementation — Utilization



(Virtex-7 XC7VX485T)

Results of the prototypical implementation — Virtex-7 UltraScale+



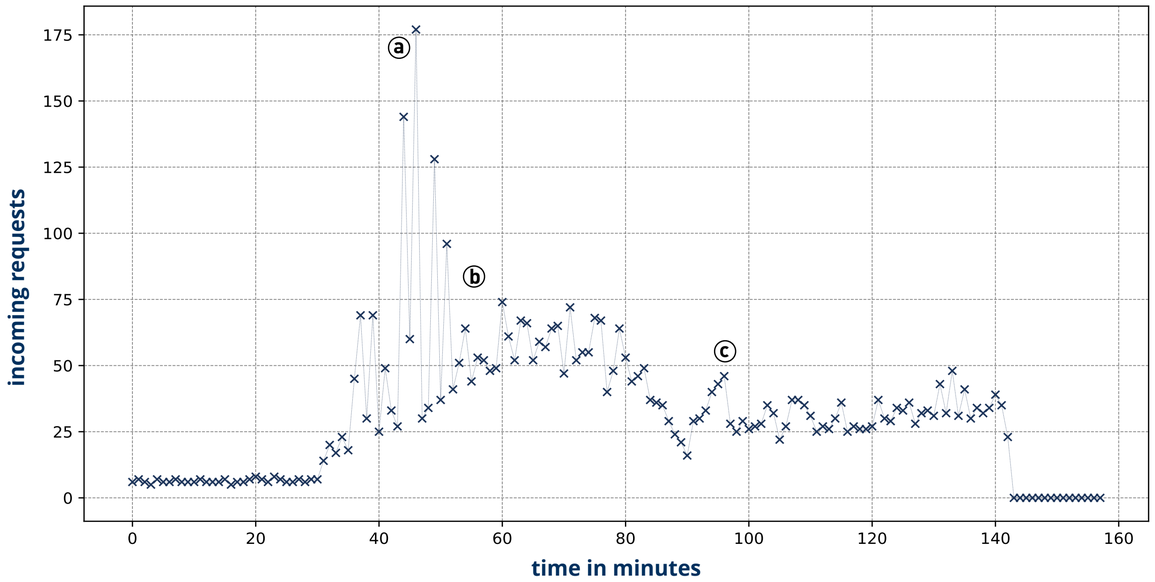
RC2F-Prototype

Virtex-7 XC7VX485T

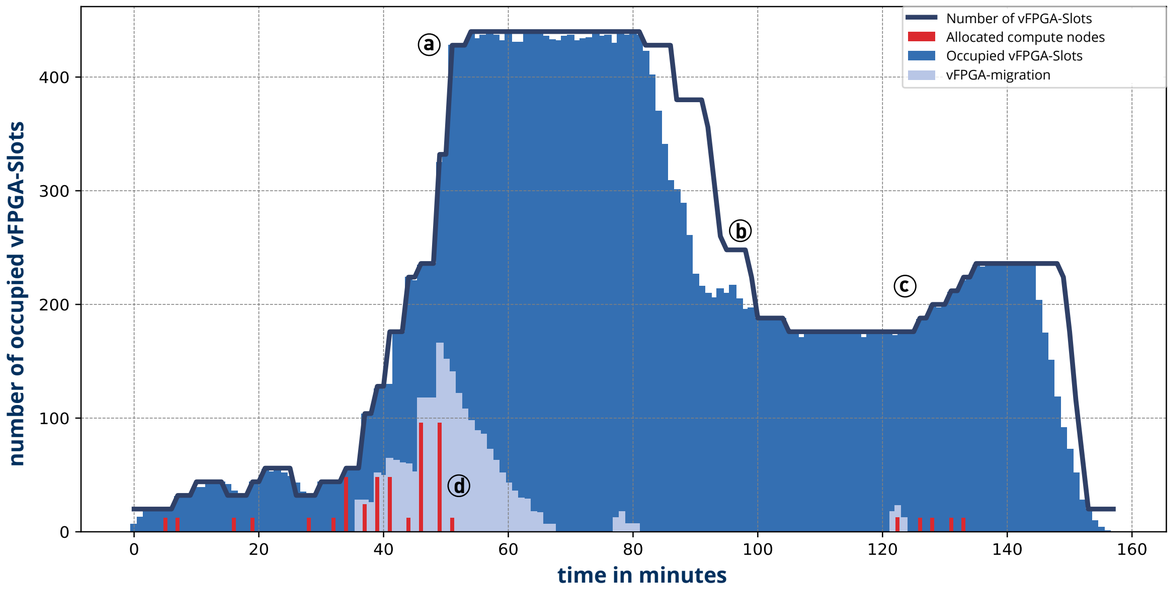
estimation for a productive cloud

Virtex-7 UltraScale+ XCVU9P

Results of the RC3E simulation with a synthetic workload (BAaaS)



Synthetic workload with **4,981 requests** over 150 minutes.



Assignment of vFPGA-Slots in an simulation with additional migration to reduce defragmentation (**+FPGAs +RC2F +migration**).

Results of the RC3E Simulation (BAaaS)

	Scenario I (synthetic) — 4,981 requests			
	Basis (without FPGAs)	+FPGAs	+RC2F	+Migration
Compute Nodes	357	132	26	24
Utilization of the FPGAs	—	27.34 %	78.14 %	85.07 %
Energy demand (kWh)	35.37 kWh	24.53 kWh	8.64 kWh	8.13 kWh
Energy demand (%)	100 %	69.35 %	24.43 %	22.91 %

— By using two Virtex-7 FPGAs per node, the energy demand of the cloud can be reduced by **30.65 %**.

— RC2F virtualization reduces the energy demand of the cloud system to **24.43 %**.

— An additional migration of vFPGAs adds an extra **1.45 %** compared to a simple RC2F virtualization.

Final considerations, summary and outlook

Outlook: Aspects for our Future FPGA Cloud (with virtualization?)

- Provision of reconfigurable Hardware in a Cloud requires an abstraction from the physical hardware and a flexible environment (RC3E & RC2F).
- A fine-grained virtualization with multiple users on the same physical device is possible and can be reasonable for the background acceleration of suitable services in a productive cloud (BAaaS),
- … but inefficient for most of the typical scenarios in our research context at the HZDR.

Store Data
Filter & Compression

Post-Process Data
Deep learning & Analyse

HW Development
Prototyping & CI

CLOUD

Alibaba gets Xilinx FPGA for its F3 cloud

by FUAD ABAZOVIC on 11 MAY 2018

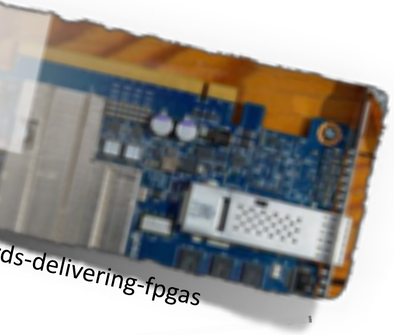
Tweet Like One person likes this. Sign Up to see what your friends like. G+

FONT SIZE PRINT EMAIL

https://fudzilla.com/news/memory-and-storage/46283-alibaba-gets-xilinx-fpga-for-its-f3-cloud

Microsoft takes step towards delivering FPGAs as a service

Launches preview of FPGA-powered Project Brainwave



https://www.computerworld.com.au/article/640970/microsoft-takes-step-towards-delivering-fpgas

WIRED

CADE METZ BUSINESS 09.25.16 07:00 PM

MICROSOFT BETS ITS FUTURE ON A REPROGRAMMABLE COMPUTER CHIP

https://www.wired.com/2016/09/microsoft-bets-future-chip-reprogram-fly/

SEP 27, 2017 @ 08:48 AM 8,423

Amazon And Xilinx Deliver New FPGA Solutions

Forbes

https://www.forbes.com/sites/moorinsights/2017/09/27/amazon-and-xilinx-deliver-new-fpga-solutions/#3f880862370a

Facebook has a new job posting calling for chip designers

Matthew Lynley @mattlynley / Apr 19, 2018

https://techcrunch.com/2018/04/18/facebook-has-a-new-job-posting-calling-for-chip-designers/

FPGAs and the New Era of Cloud-based 'Hardware Microservices'

8 Jun 2017 6:00am, by Mary Branscombe

https://thenewstack.io/developers-fpgas-cloud/

References

[**KG+18**] **Oliver Knodel, Paul Genßler,Fredo Erxleben and Rainer Spallek.** "**An Endless Tale of Virtualization, Elasticity and Efficiency**". In: *International Journal on Advances in Systems and Measurements, vol 11 no 3&4*. IARIA. 2016.

[**KGS17**] **Oliver Knodel, Paul R. Genssler and Rainer G. Spallek**. „Virtualizing Reconfigurable Hardware to Provide Scalability in Cloud Architectures". In: Reconfigurable Architectures, Tools and Applications, RECATA 2017, Rome, Italy, September 10 - 14, ISBN: 978-1-61208-585-2, **CENICS 2017 Best Paper Award.** IARIA. 2017, S. 33–38.

[**KGS16**] **Oliver Knodel, Paul Genßler and Rainer Spallek.** "Migration of long-running Tasks between Reconfigurable Resources using Virtualization". In: *ACM SIGARCH Computer Architecture News Volume 44, HEART 2016, July 25-27, Hongkong*. ACM. 2016.

[**KLS16**] **Oliver Knodel, Patrick Lehmann and Rainer G Spallek.** "RC3E: Reconfigurable Accelerators in Data Centres and their Provision by Adapted Service Models". In: *9th Int'l Conf. on Cloud Computing, Cloud 2016, June 27 - July 2, San Francisco, CA, USA*. IEEE. 2016.

[**Kno14**] **Oliver Knodel.** "Integration von FPGA-Ressourcen als Hardwarebeschleuniger und deren Bereitstellung sowie Verwaltung in einem Mehrbenutzersystem". In: *Dresdner Arbeitsta- gung Schaltungs- und Systementwurf, DASS 2014, Dresden, Germany*. 2014.

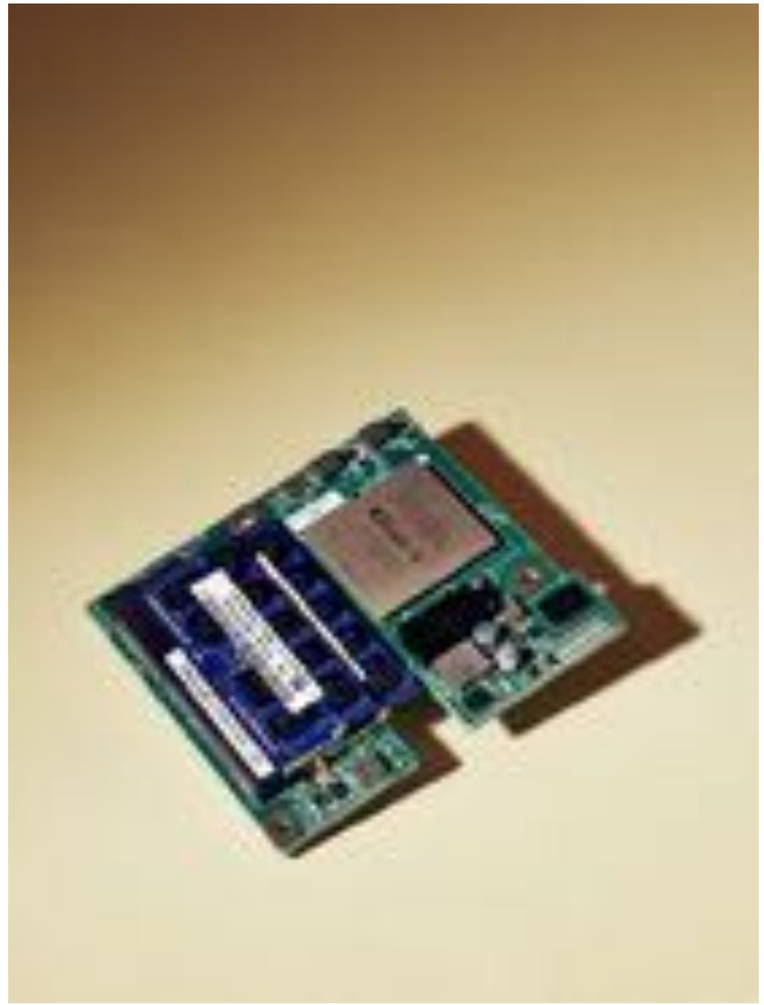
[**KS13**] **Oliver Knodel and Rainer Spallek.** "FPGAs in der Cloud: Integration und Bereitstellung von rekonfigurierbaren Hardware-Ressourcen in einer Cloud-Infrastruktur". In: *5. Workshop für Grid-, Cloud- und Big-Data-Technologien für Systementwurf und -analyse, Grid4Sys 2013, Dresden, Germany*. 2013.

[**KS15**] **Oliver Knodel and Rainer G. Spallek.** "Computing Framework for Dynamic Integration of Reconfigurable Resources in a Cloud". In: *2015 Euromicro Conference on Digital System Design, DSD 2015, Madeira, Portugal, August 26-28*. IEEE. 2015, S. 337–344.

Appendix

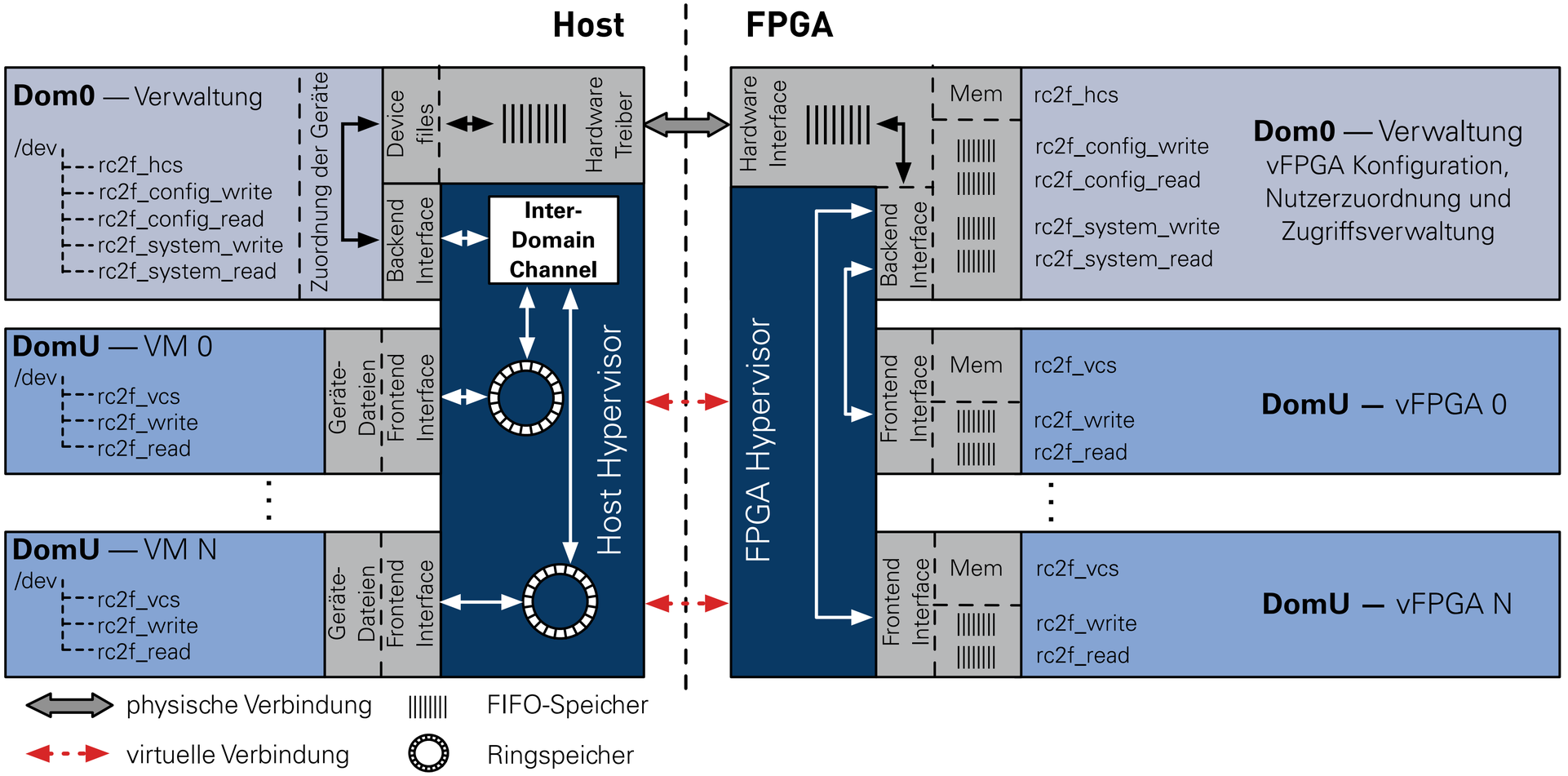
Outlook: Aspects for Future FPGA Architectures

- **SoC-FPGA with a static RC2F-like Infrastructure:**
  Development of a static area for the administration with all necessary interfaces to the outside world.
- **Structural changes within the FPGA architecture:**
  Establishment of homogeneous and areas with own clock regions on the FPGAs.
- **Implementation of a security concept:**
  Security concept based on a trusted authority to provide verifiable RC2F infrastructure.



https://www.wired.com/2016/09/microsoft-bets-future-chip-reprogram-fly/

vFPGAs and the Host



Virtex-7 Ultrascale+