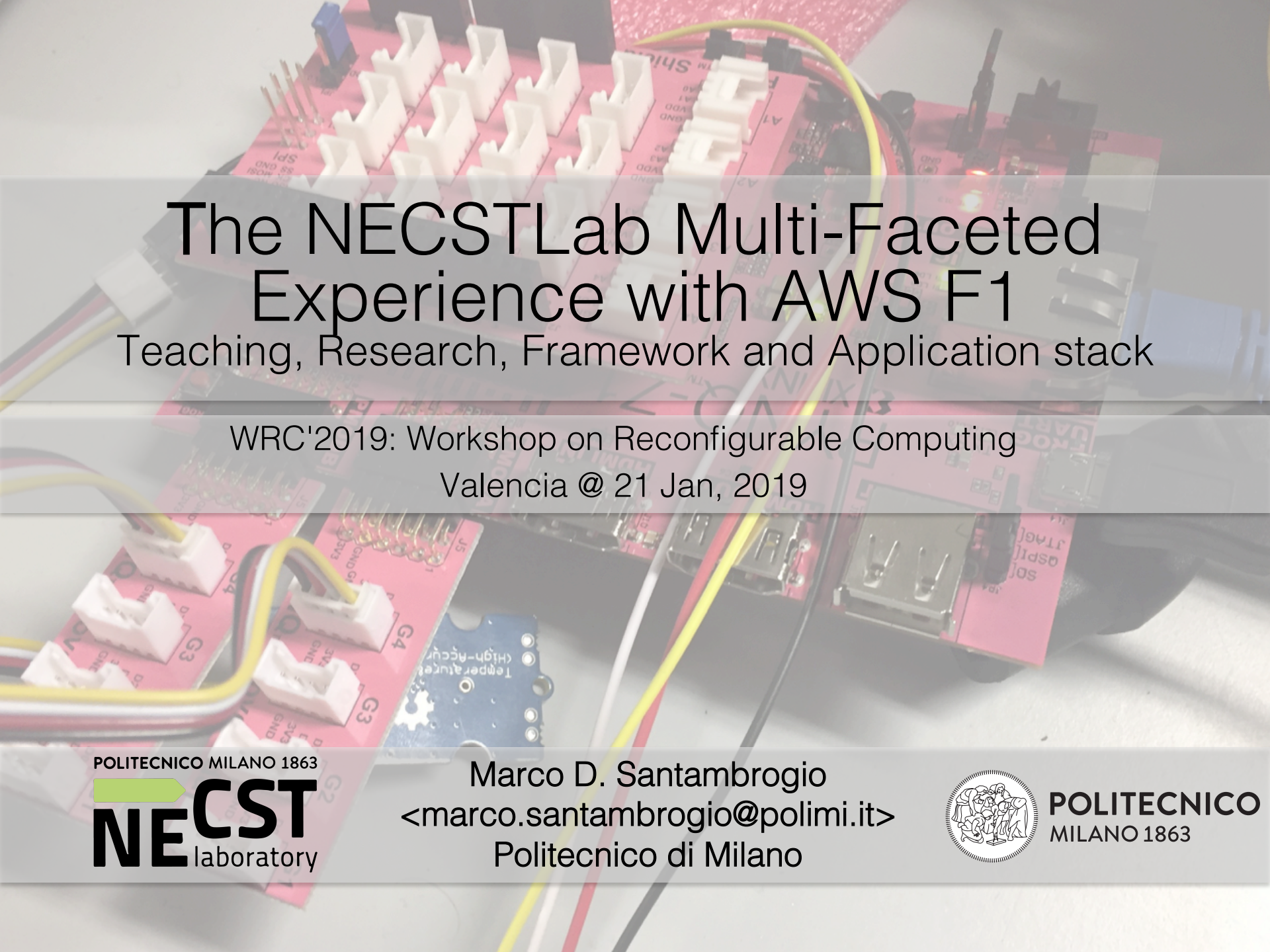


The NECSTLab Multi-Faceted Experience with AWS F1

Teaching, Research, Framework and Application stack

WRC'2019: Workshop on Reconfigurable Computing

Valencia @ 21 Jan, 2019

POLITECNICO MILANO 1863

NECST laboratory

Marco D. Santambrogio
<marco.santambrogio@polimi.it>
Politecnico di Milano

POLITECNICO MILANO 1863

What This Talk Is All About

What This Talk Is All About

What This Talk Is All About

What This Talk Is All About

HUG

CAOS

Computing systems are getting…

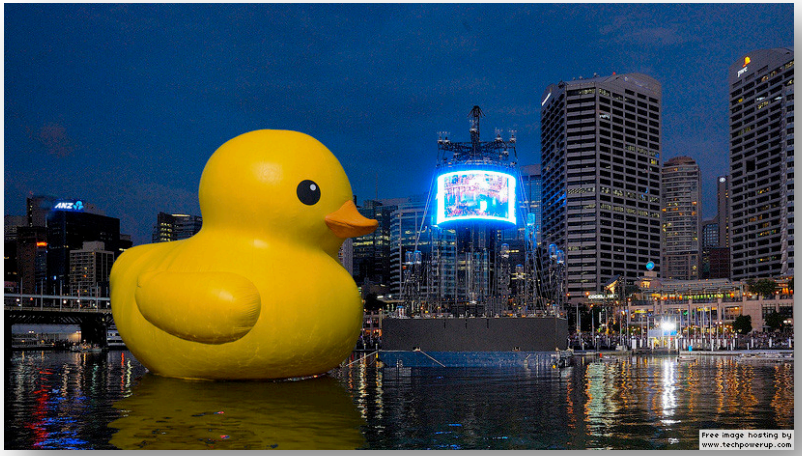
Computing systems are getting…

little…

Computing systems are getting…

little…

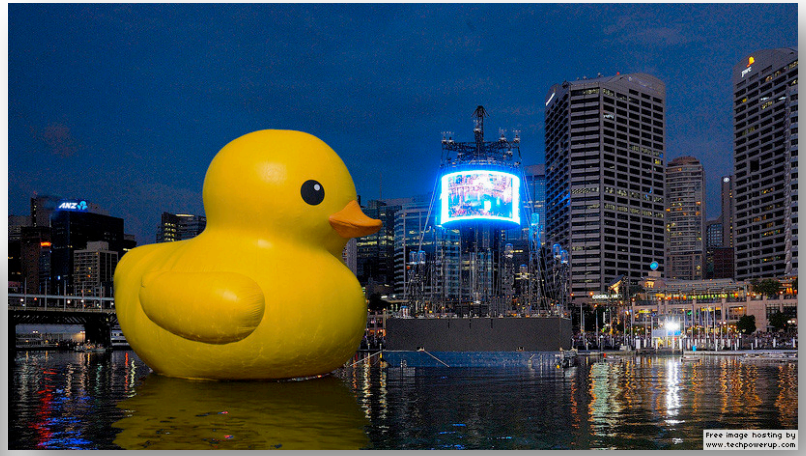
little+Big



Computing systems are getting…

little…

little+Big



little+Big and heterogeneous

Heterogeneous Complex Systems

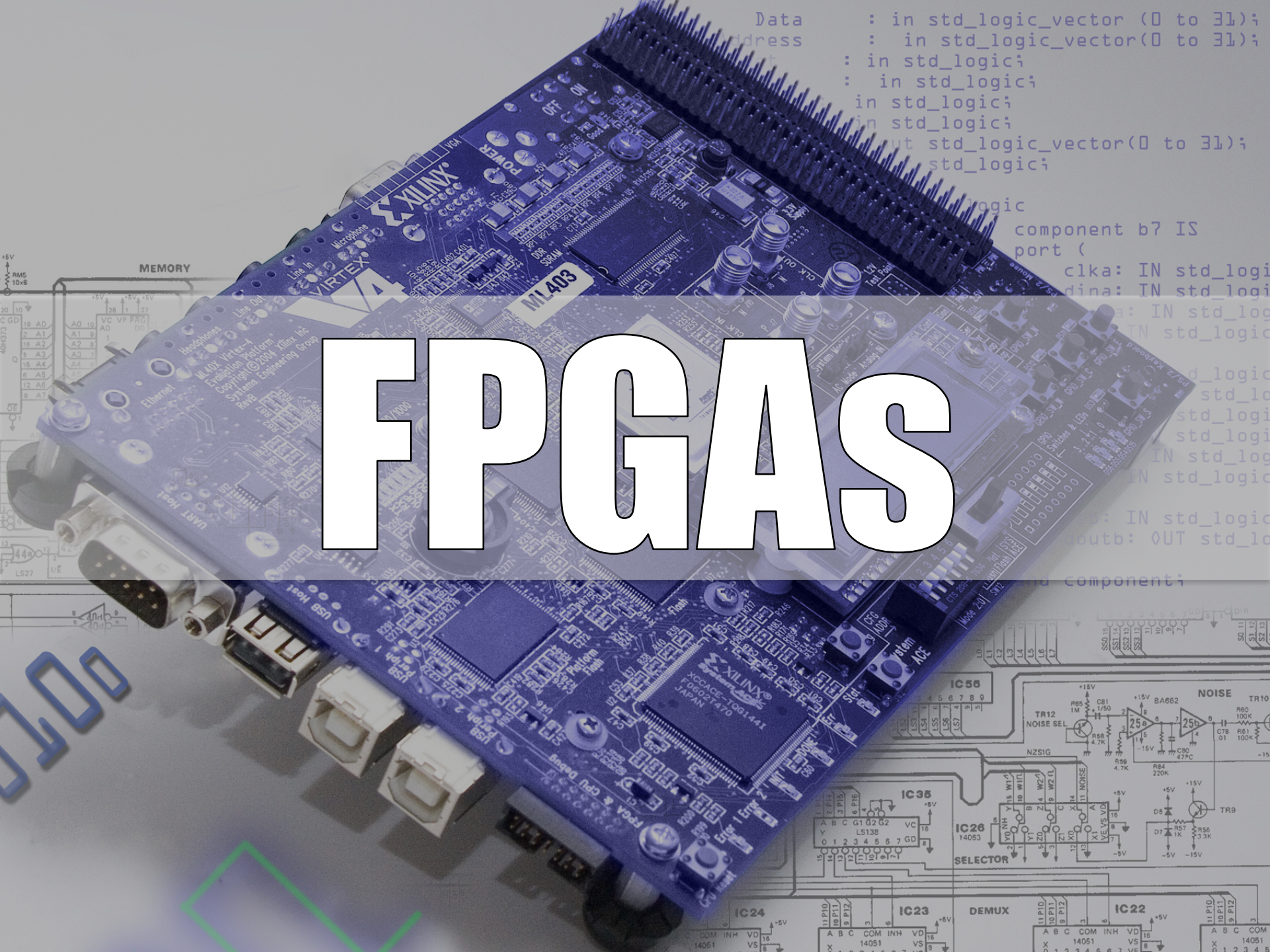
- Ryft ONE
  - Big Data infrastructure due to an FPGA-accellerated architecture
  - http://www.ryft.com/
- IBM Power8
  - Introducing the Coherent Accelerator Processor Interface (CAPI) port that is layered on top of PCI Express 3.0
  - http://www-304.ibm.com/webapp/set2/sas/f/capi/home.html
- Microsoft Catapult
  - Stratix V (Arria 10 FPGA)
  - http://research.microsoft.com/en-us/projects/catapult/
- Amazon EC2 F1 Instances
  - Xilinx UltraScale Plus FPGA
  - https://aws.amazon.com/about-aws/whats-new/2017/04/amazon-ec2-f1-instances-customizable-fpgas-for-hardware-acceleration-are-now-generally-available/
- OpenPower Foundation
  - http://openpowerfoundation.org/

CAPI Technology

intel FPGA

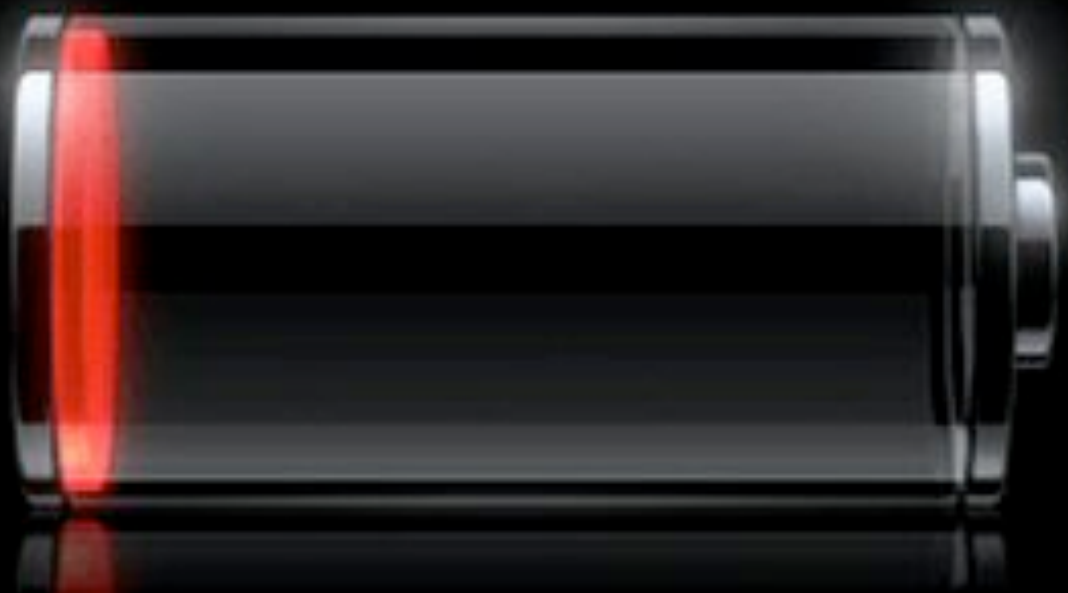
amazon web services

OpenPOWER™


FPGAs


SYSTEM HAS TO BE ADAPTIVE
TO RECOVER
FROM DAMAGES

SYSTEM HAS TO BE ADAPTIVE



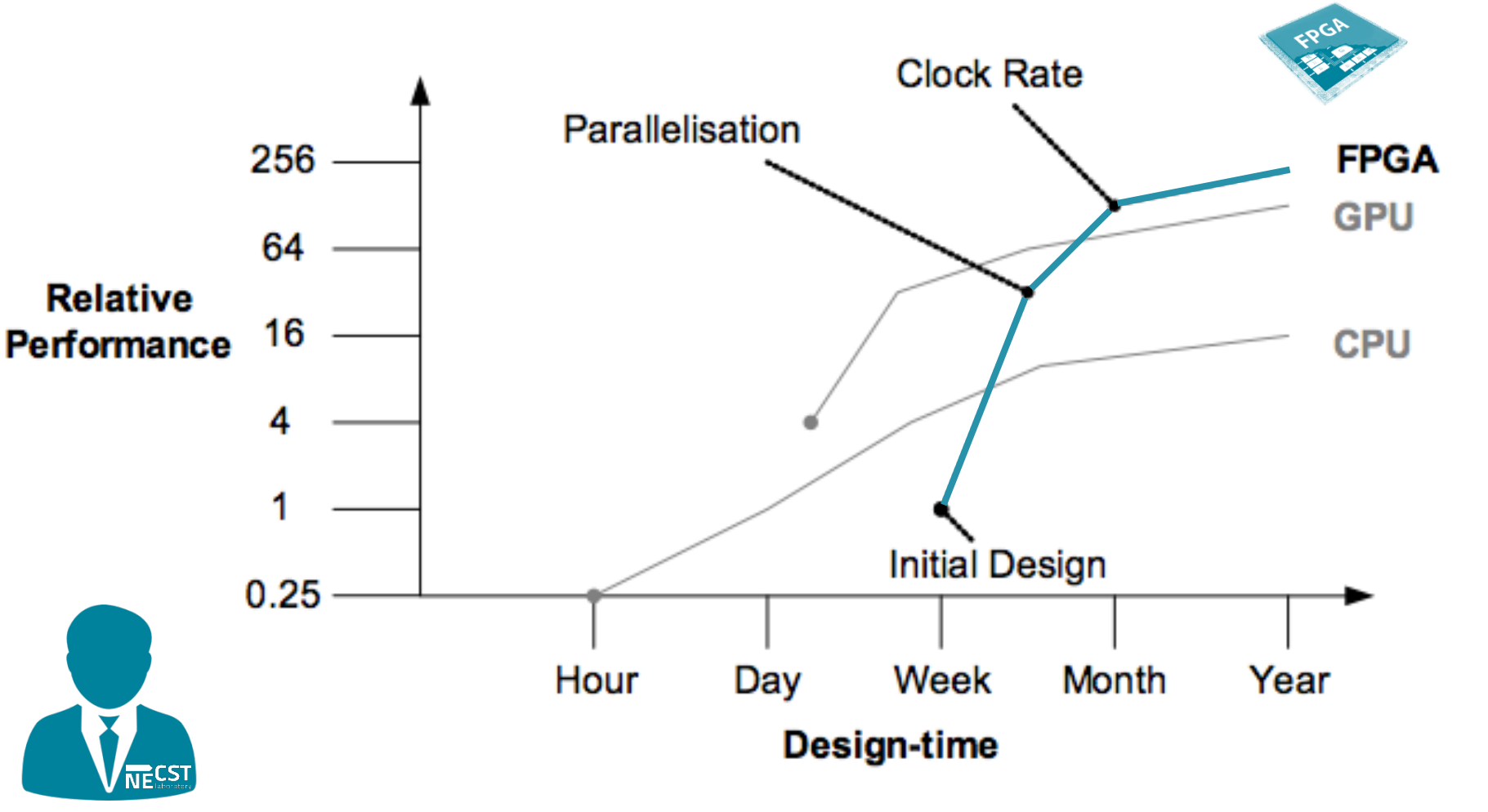
TO GUARANTEE SERVICES OVER POWER CAP AND ENERGY SAVINGS


SYSTEM HAS TO BE ADAPTIVE
TO ADAPT TO

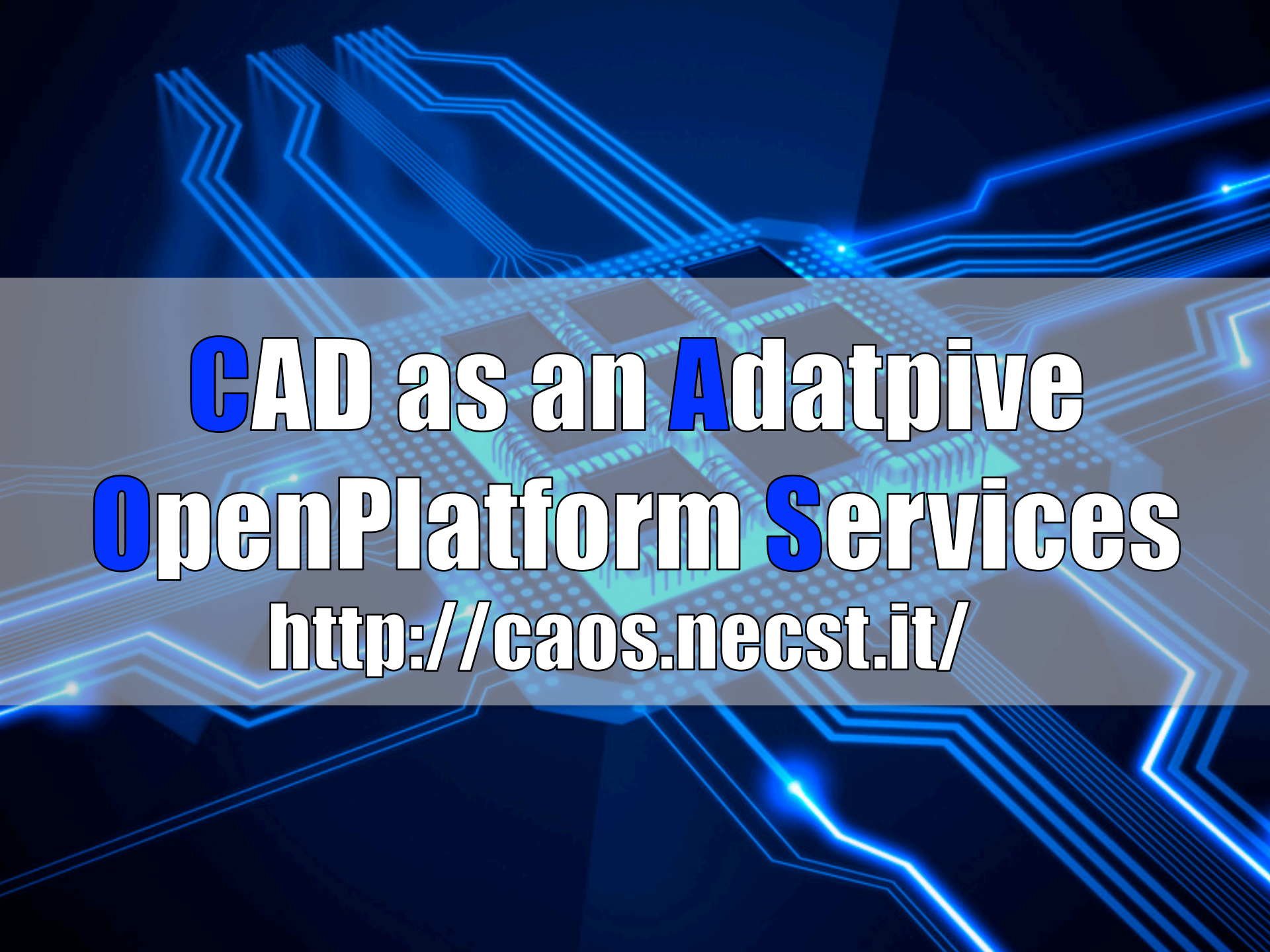

SYSTEM HAS TO BE ADAPTIVE
TO ADAPT TO
UKNOWN CONDITIONS


HOW TO MANAGE/DEAL WITH
THESE HETEROGENOUS SYSTEMS

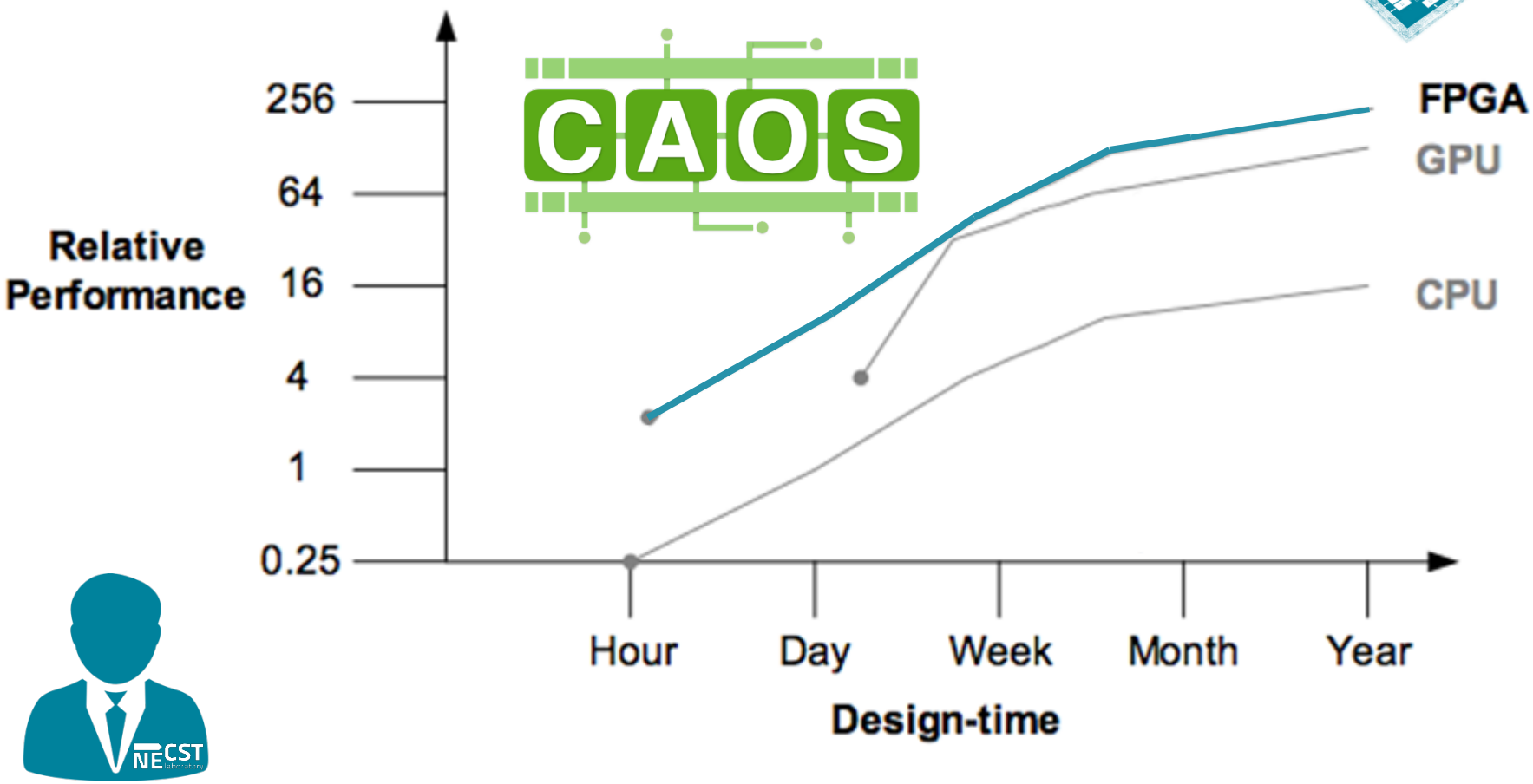
Research Challenge




A RECONFIGURABLE-FRIENDLY
ECOSYSTEM IS NEEDED

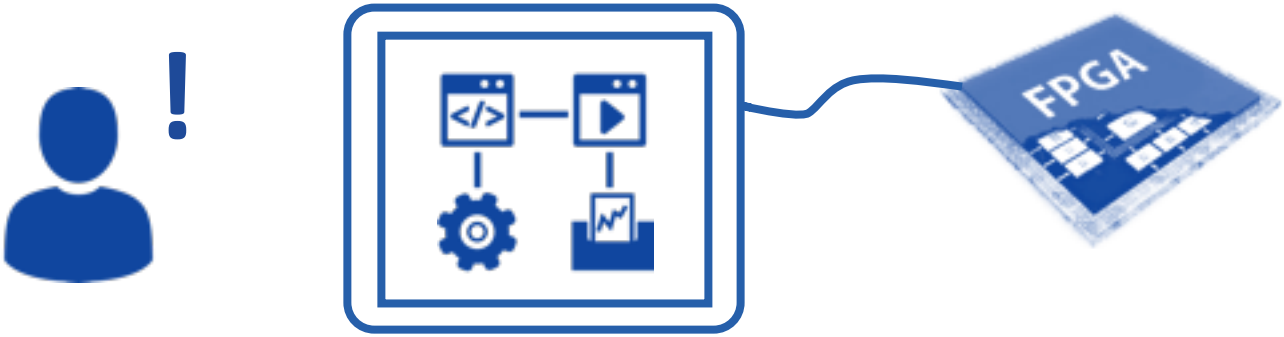

CAD as an Adatpive
OpenPlatform Services
http://caos.necst.it/

CAD as an Adatpive OpenPlatform Services



CAD as an Adatpive OpenPlatform Services

Usability

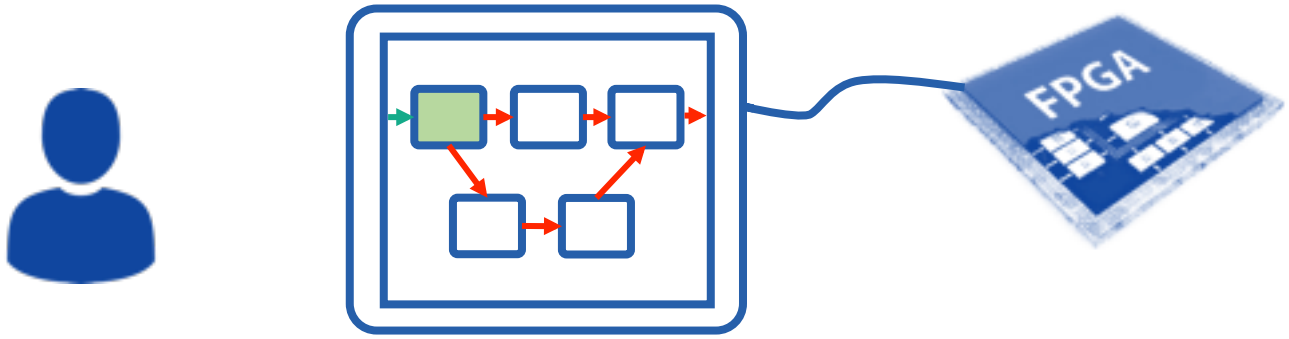


CAD as an Adatpive OpenPlatform Services

Interactivity

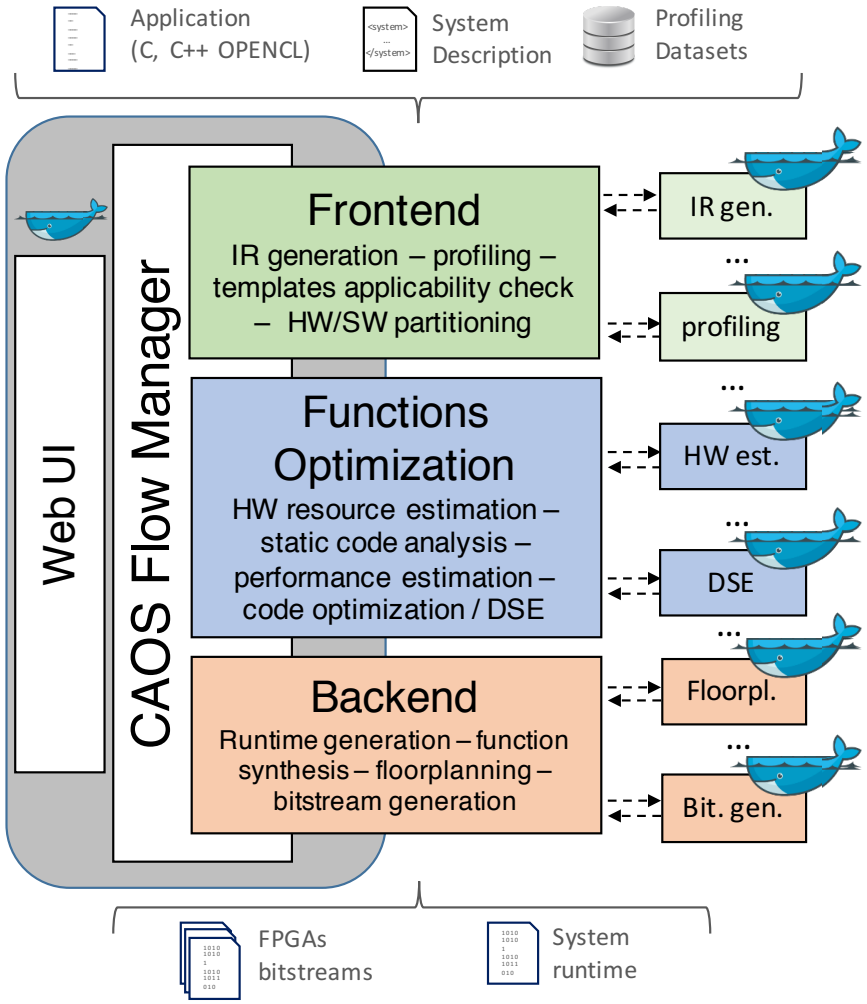
CAD as an Adatpive OpenPlatform Services

Modularity

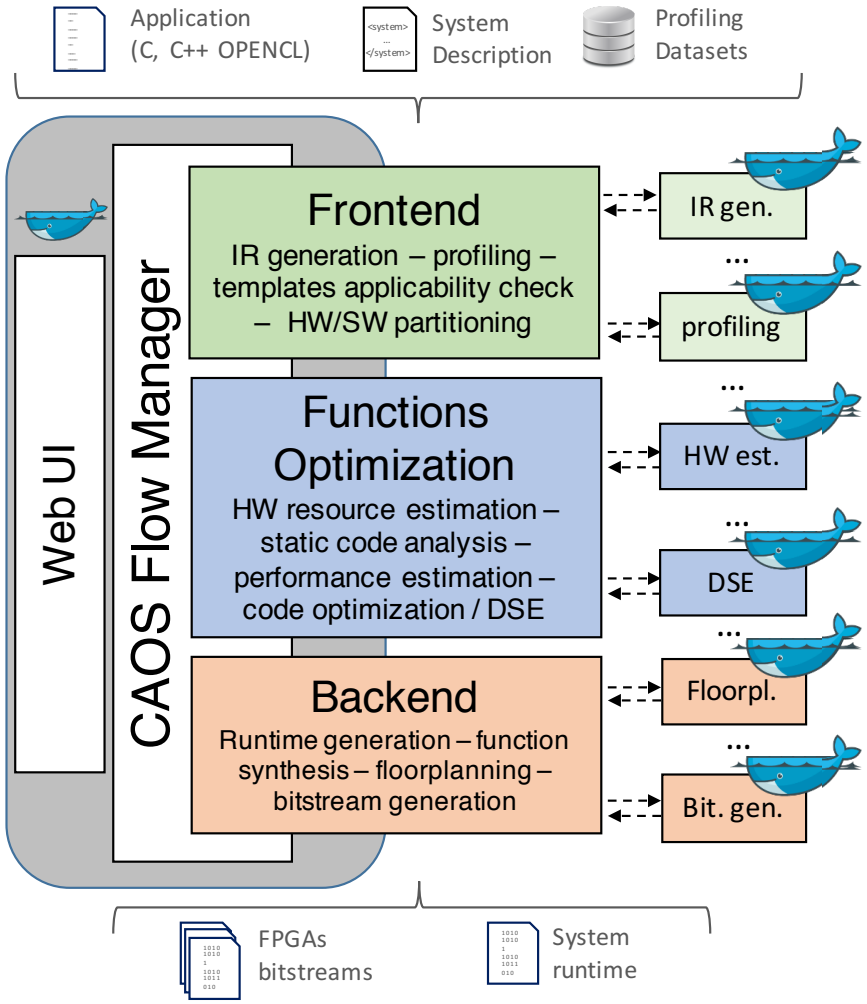


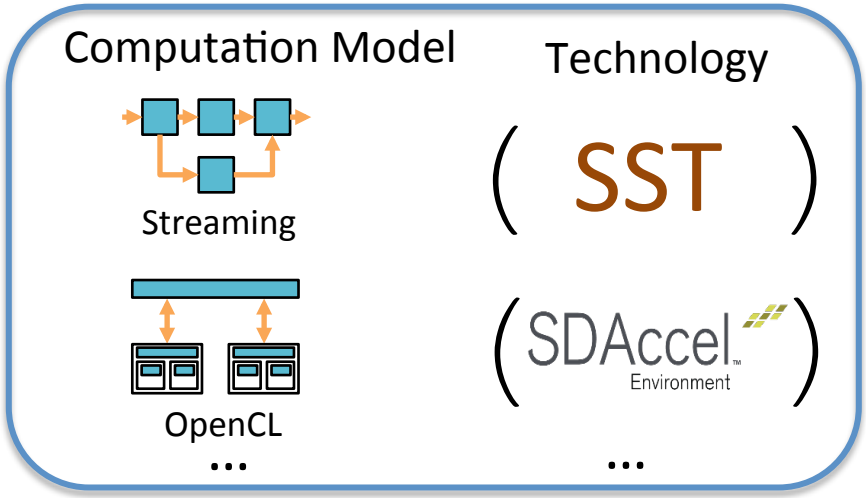

How to
do it?

The proposed CAOS framework



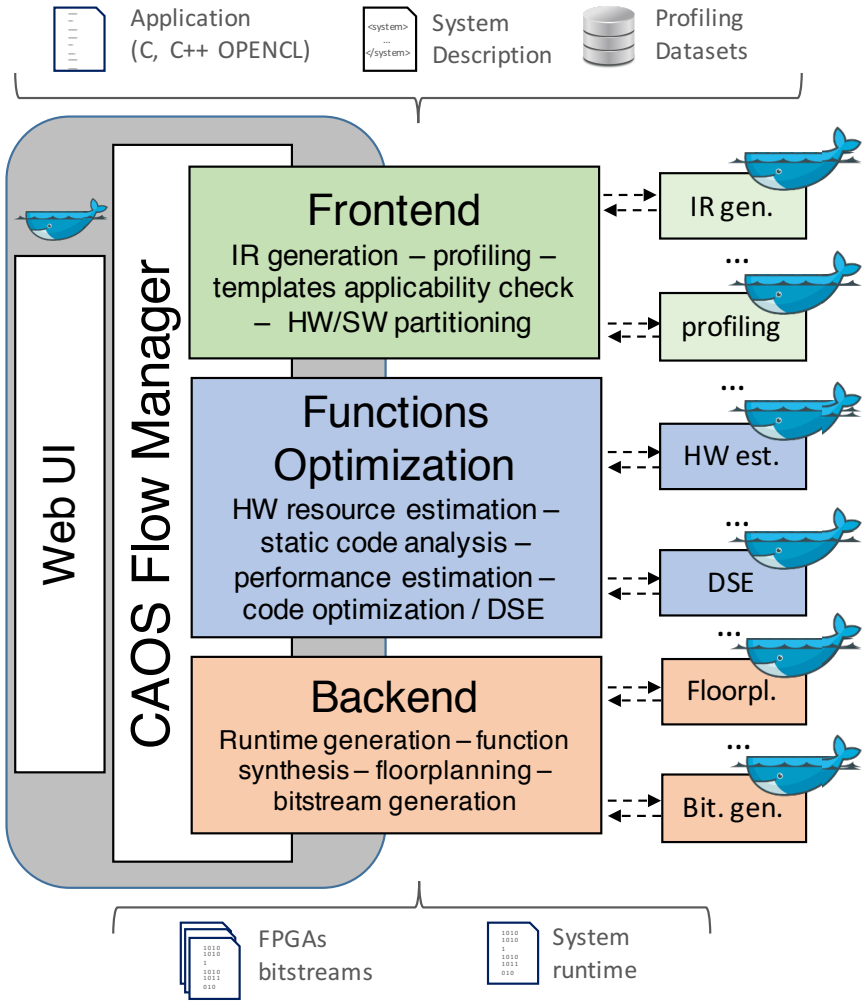
The proposed CAOS framework

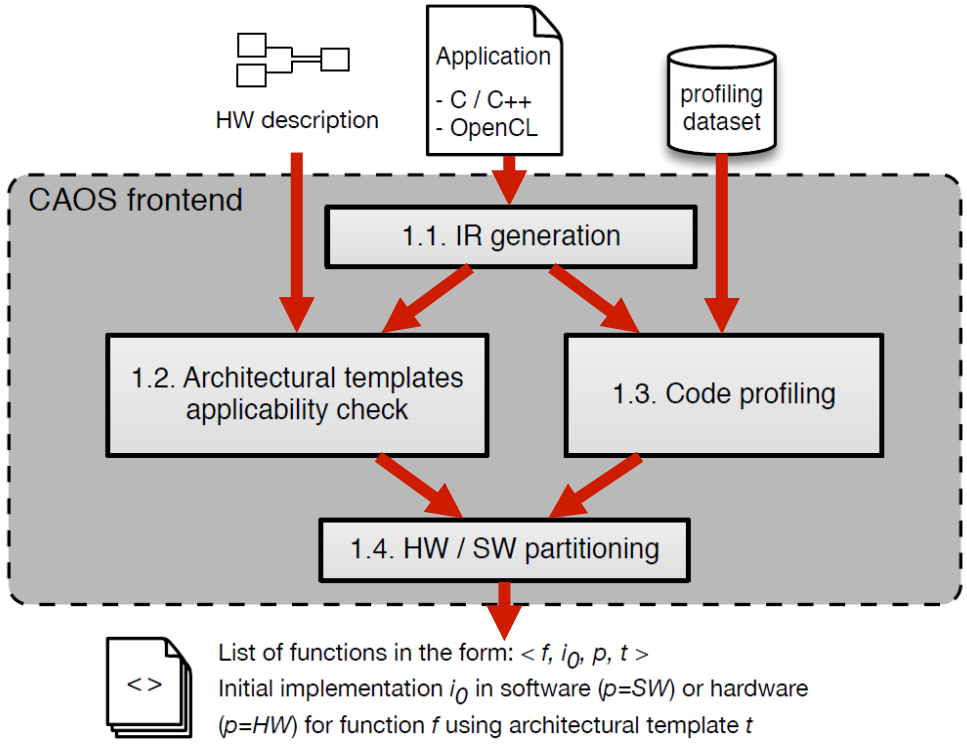




Architectural Templates

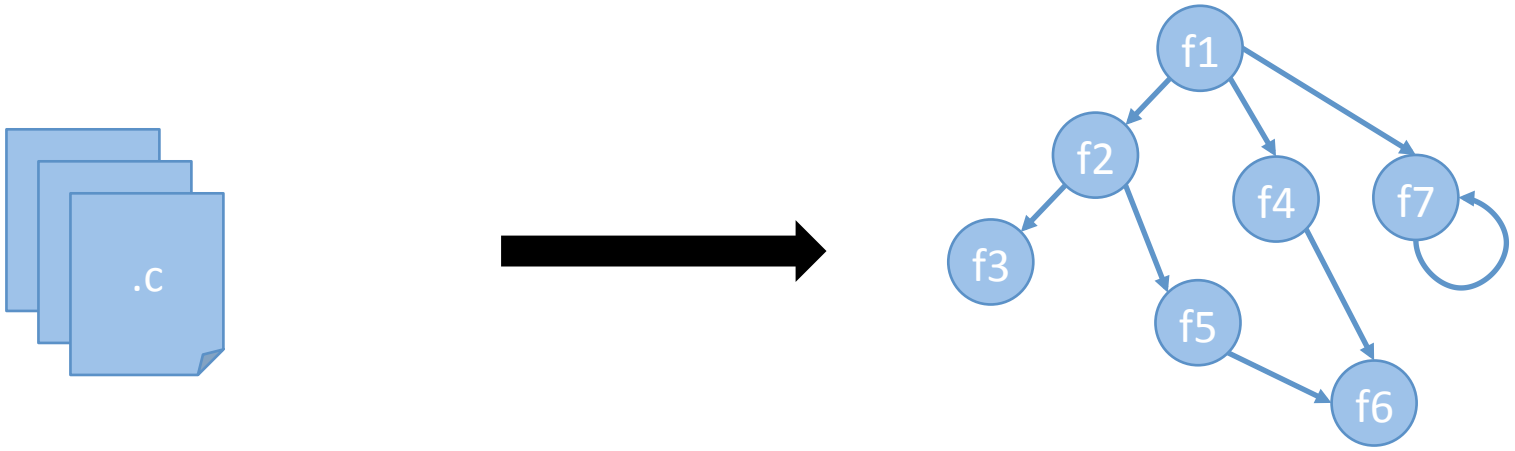
CAOS Frontend





CAOS Frontend – IR Generation

- Functions extraction and generation of the application call graph



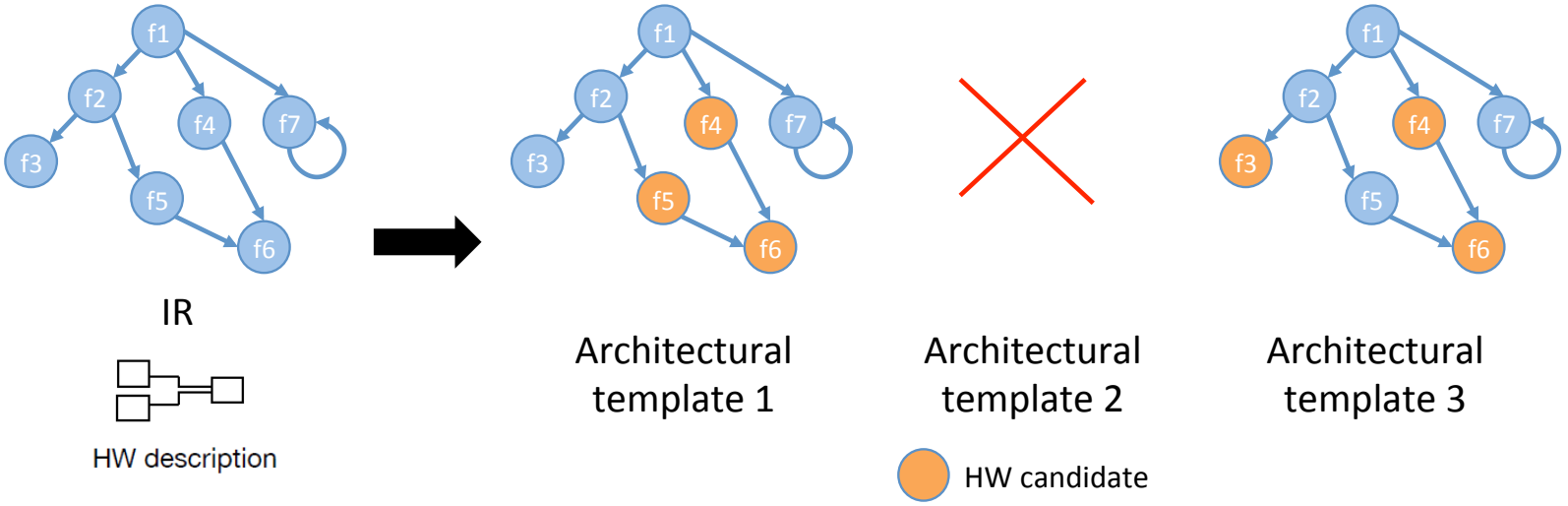
application

IR: call graph + functions description

- Current implementation leverages Doxygen

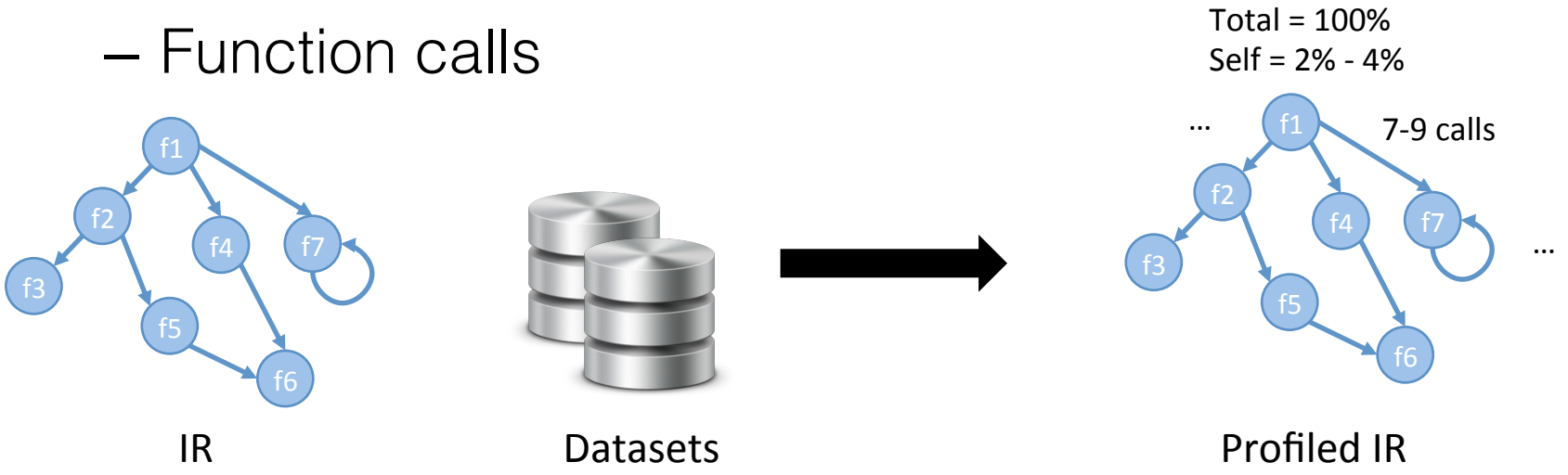
CAOS Frontend – applicability check

- Verifies the applicability of an architectural template w.r.t.:
  - Application
  - System description



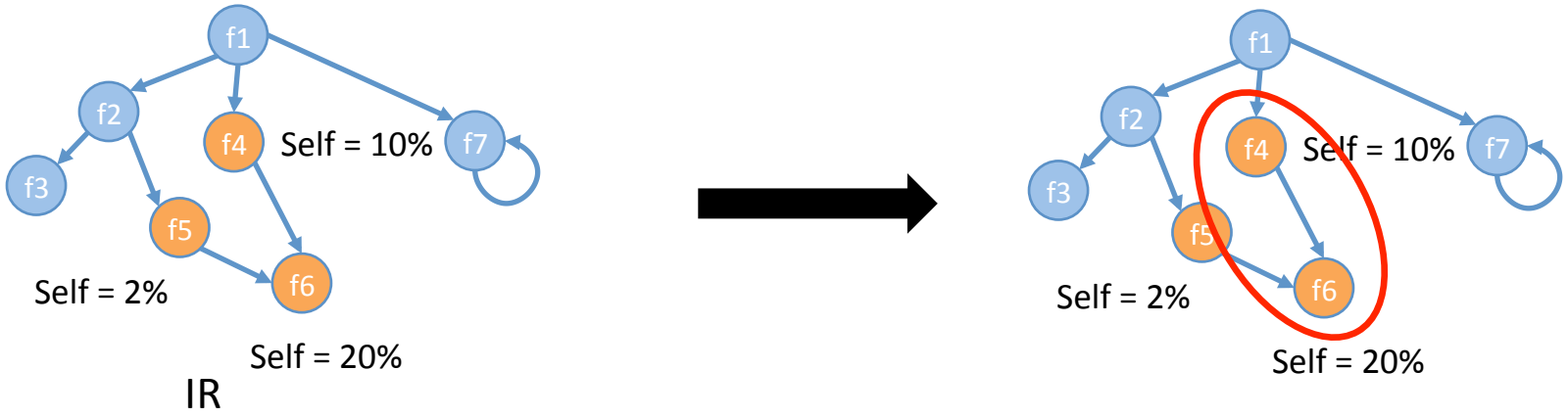
CAOS Frontend – applicability check

- Runs the application against multiple user-defined datasets
- For each functions collects:
  - Self execution time
  - Total execution time
  - Function calls



CAOS Frontend – HW/SW Partitioning

- Identifies the subtree to accelerate for each architectural template
- If needed, translate the identified code for subsequent optimizations (e.g. C to MaxJ)

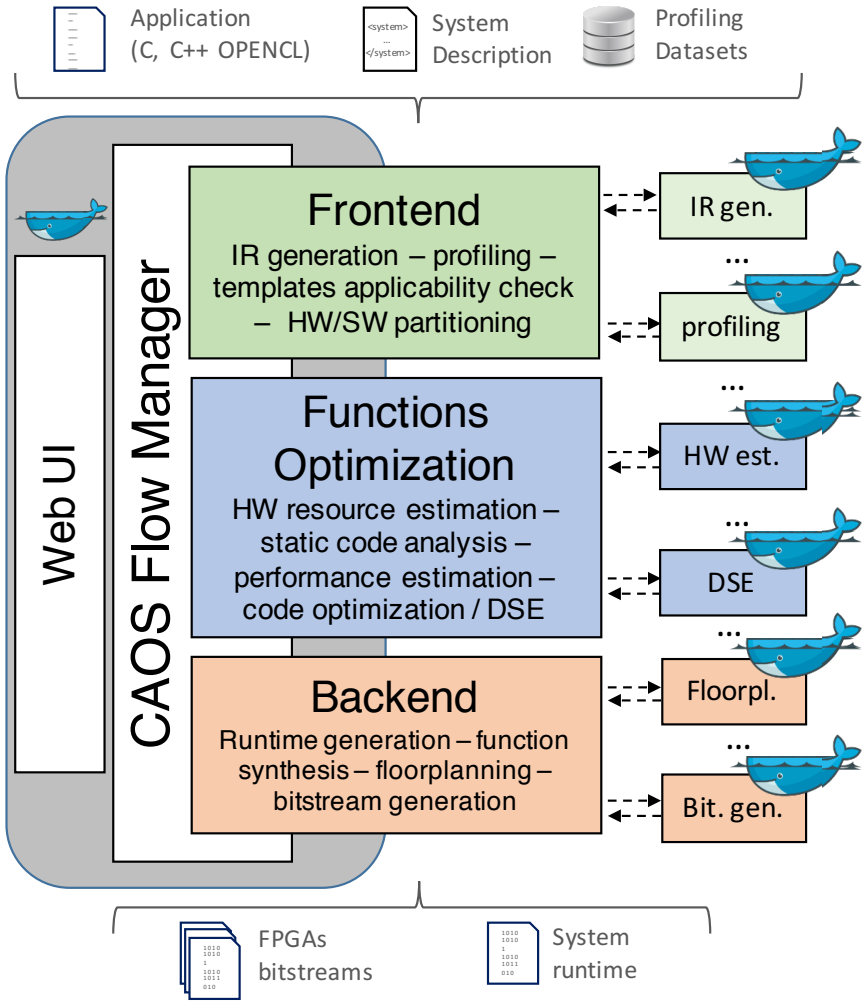


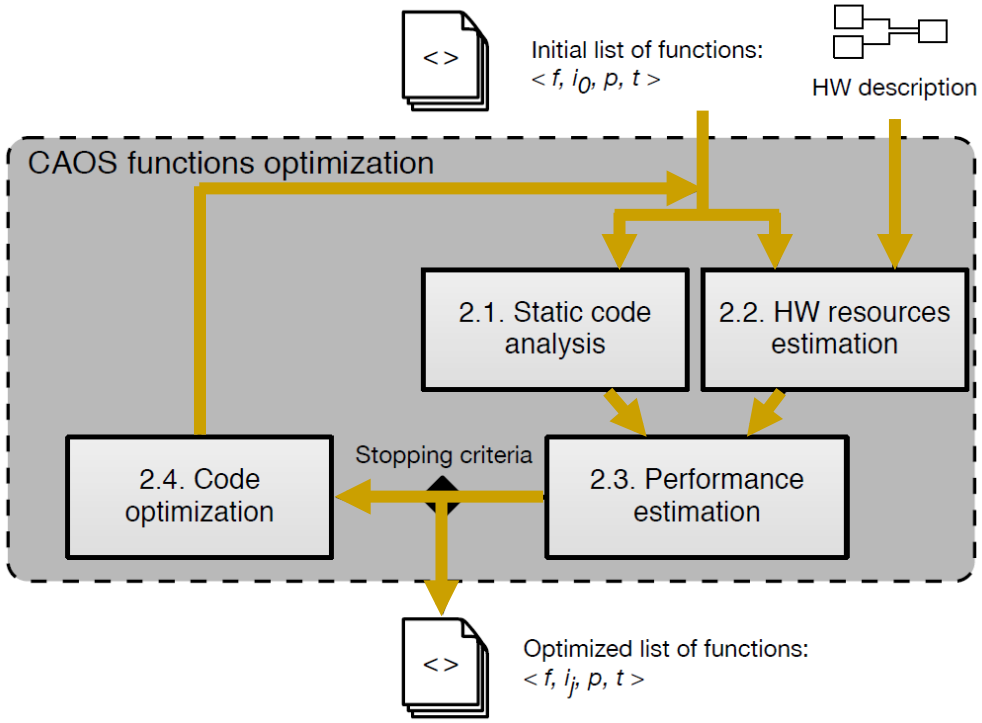
Work done in collaboration with

Imperial College London

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

CAOS Functions Optimization





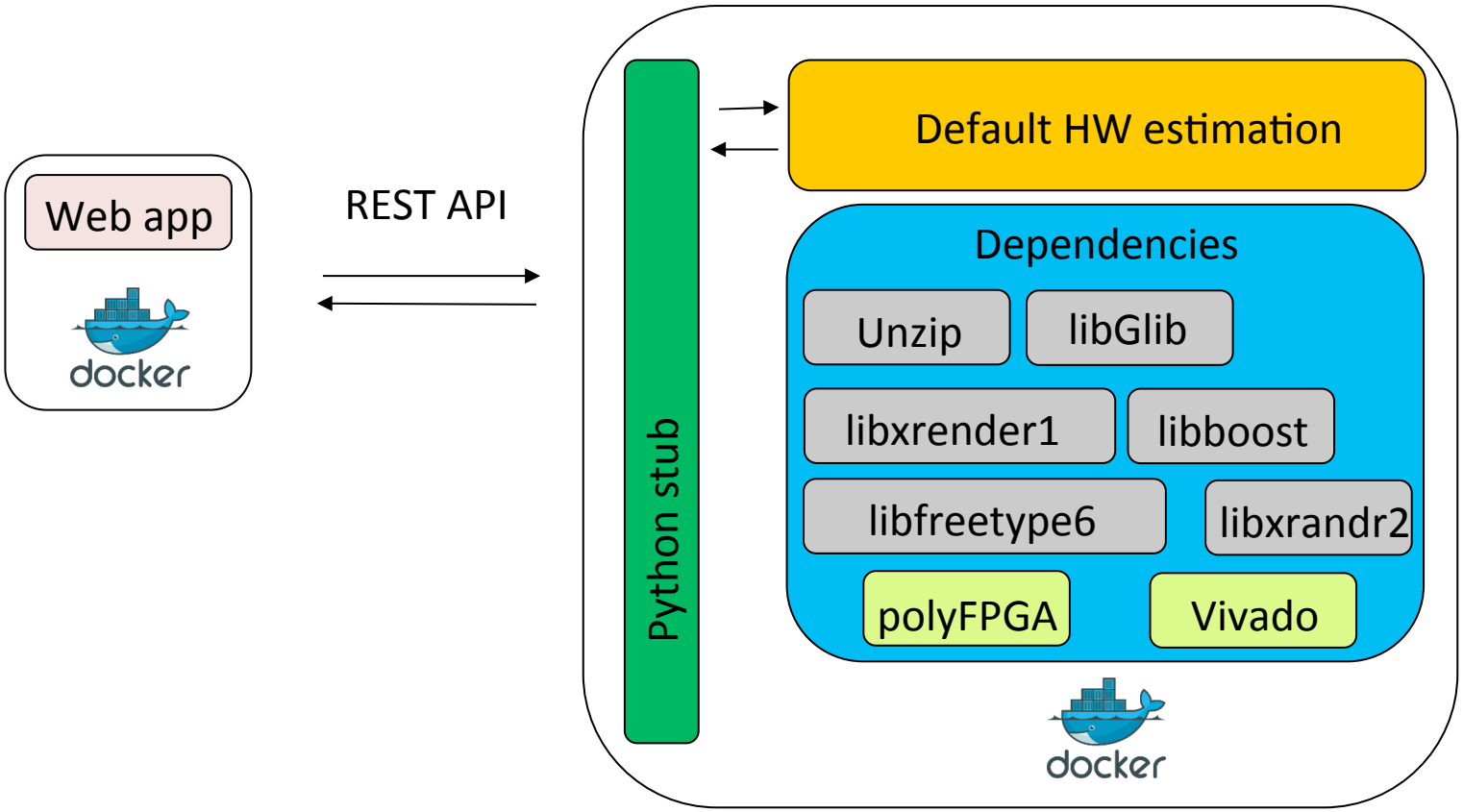
CAOS Functions Optimization Static Code Analysis

- Retrieve metrics on the current implementation for the candidate HW functions
- Metrics are architectural template dependent
  - Produce / consume rate of kernels (Maxeler)
  - Estimated module latency (SST)
  - Computational intensity (OpenCL)

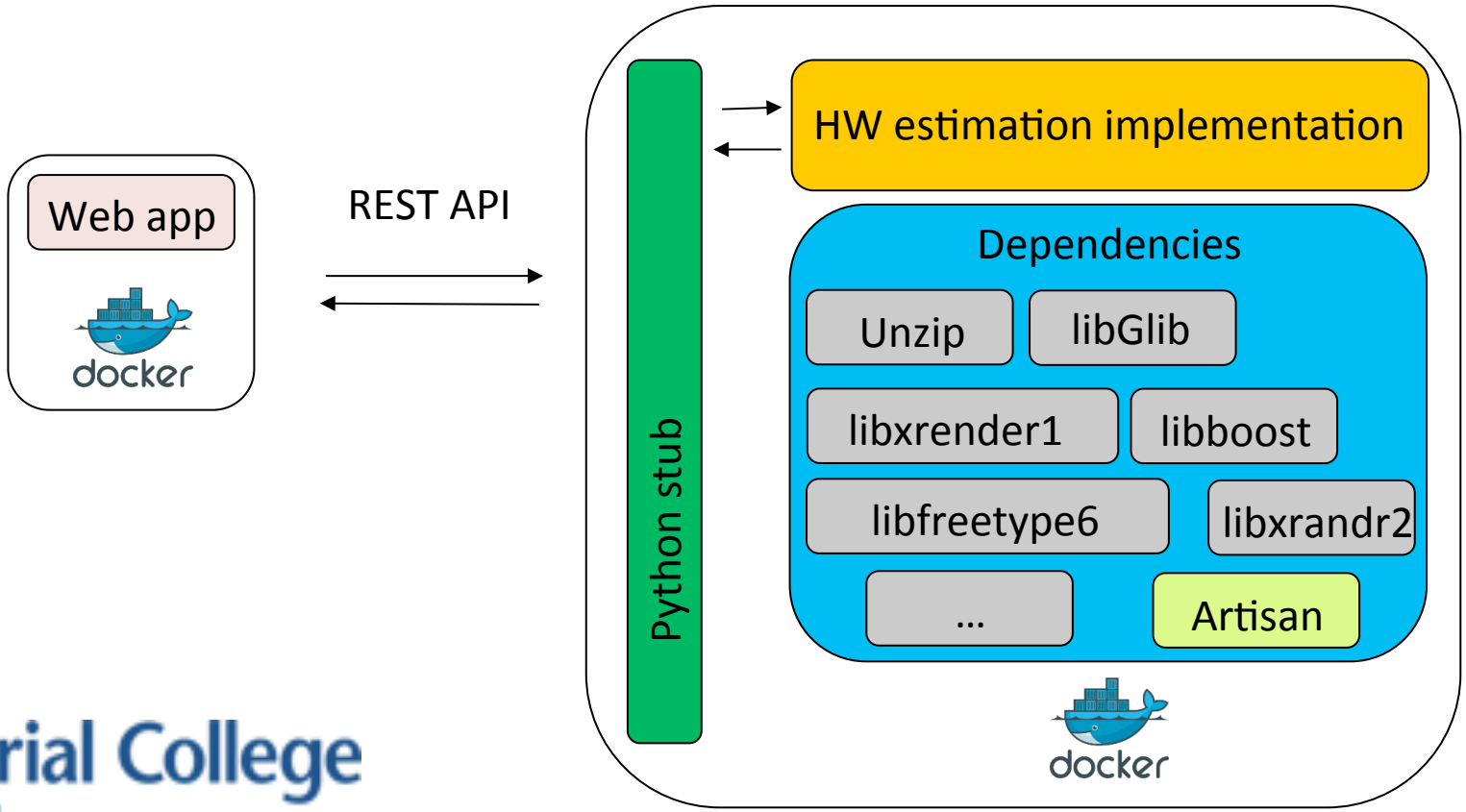
CAOS Functions Optimization Resource Estimantion

- Estimate resource requirements for the entire set of functions to accelerate in HW
- Multiple resource estimation modules:
  - Default HW estimation module (Vivado HLS based)
    - Might require a high execution time
    - Accurate estimation
  - Artisan HW estimation module
    - Operations count-based estimation
      - Fast execution time
      - Coarse grain estimation
      - MaxJ code support

CAOS default implementation of HW estimation

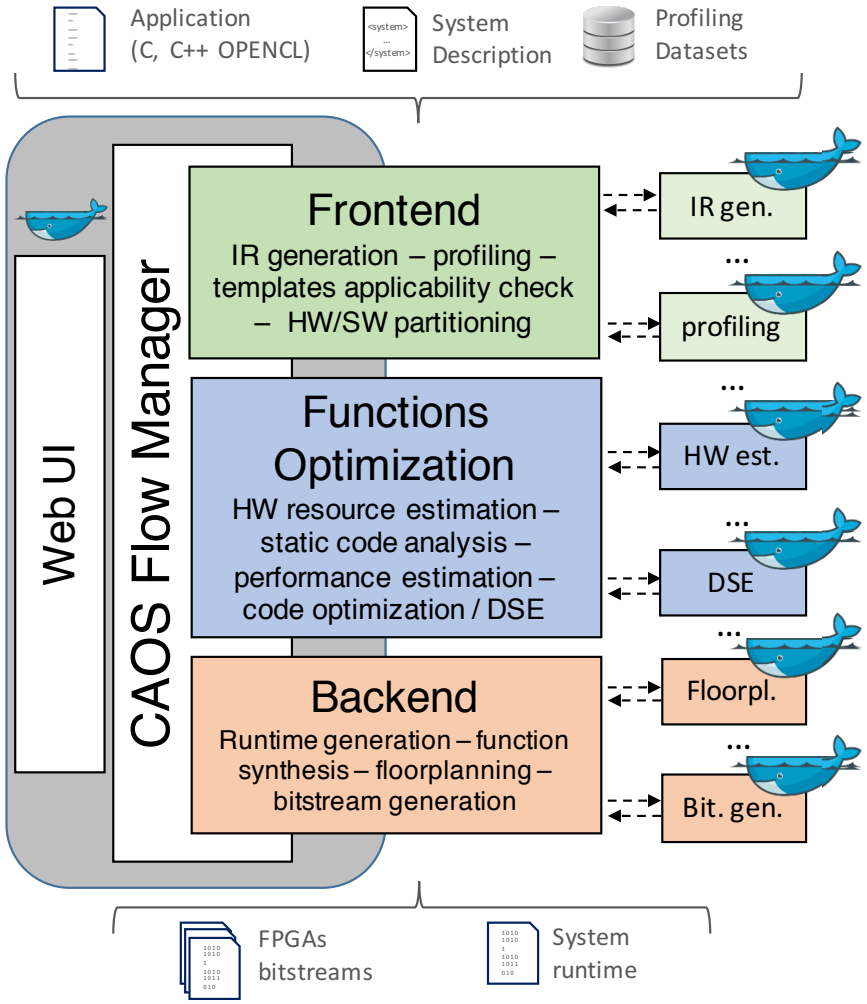


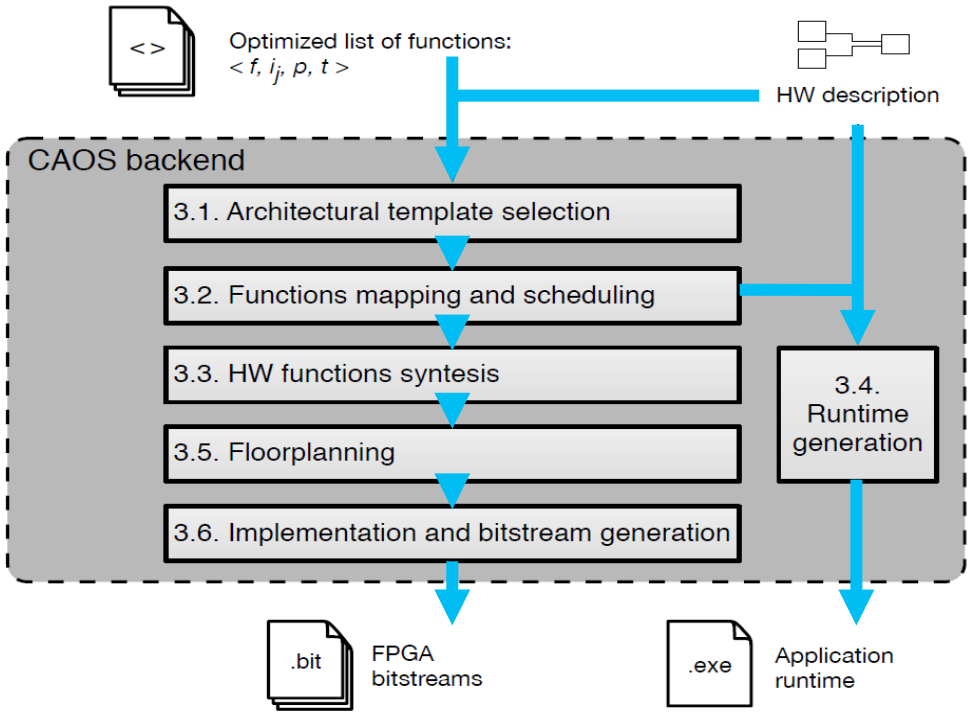
Artisan HW estimation



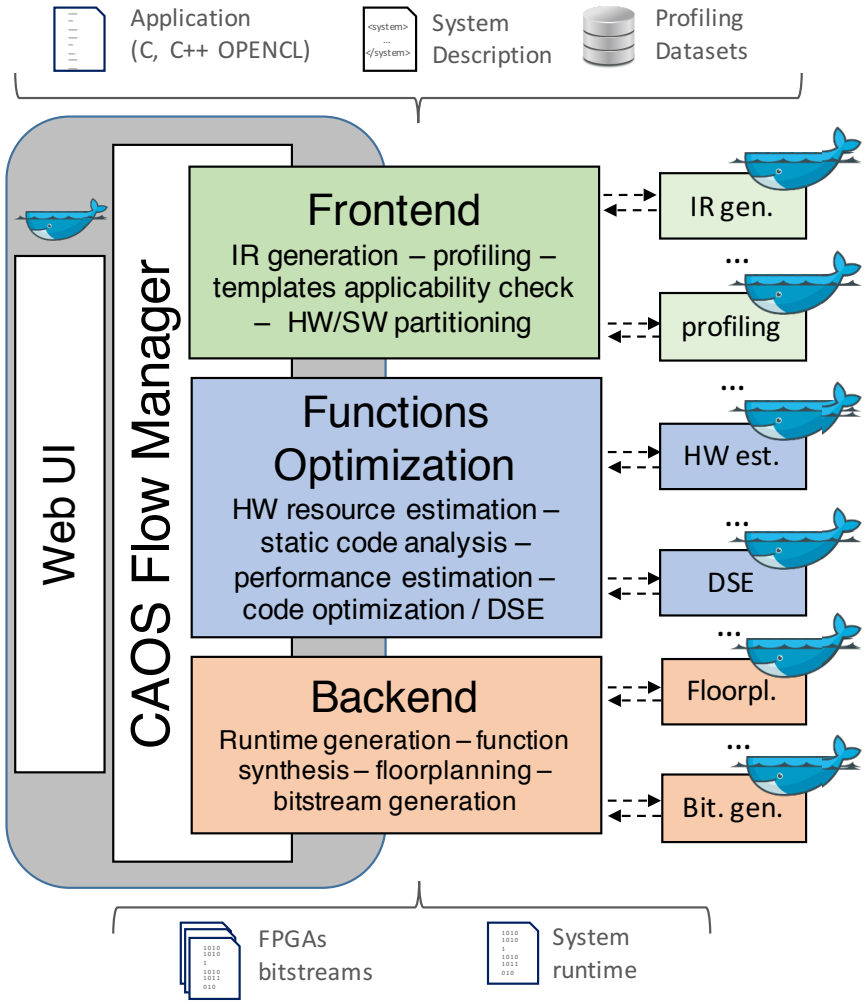
Imperial College London

CAOS Backend





CAOS Backend



SST → VIVADO

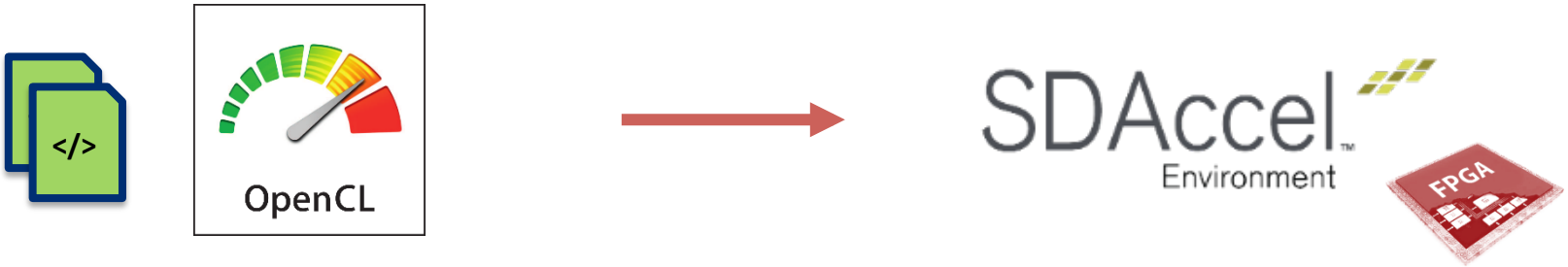
MAXELER Technologies Maximum Performance Computing → **MaxCompiler**

OpenCL → SDAccel Environment

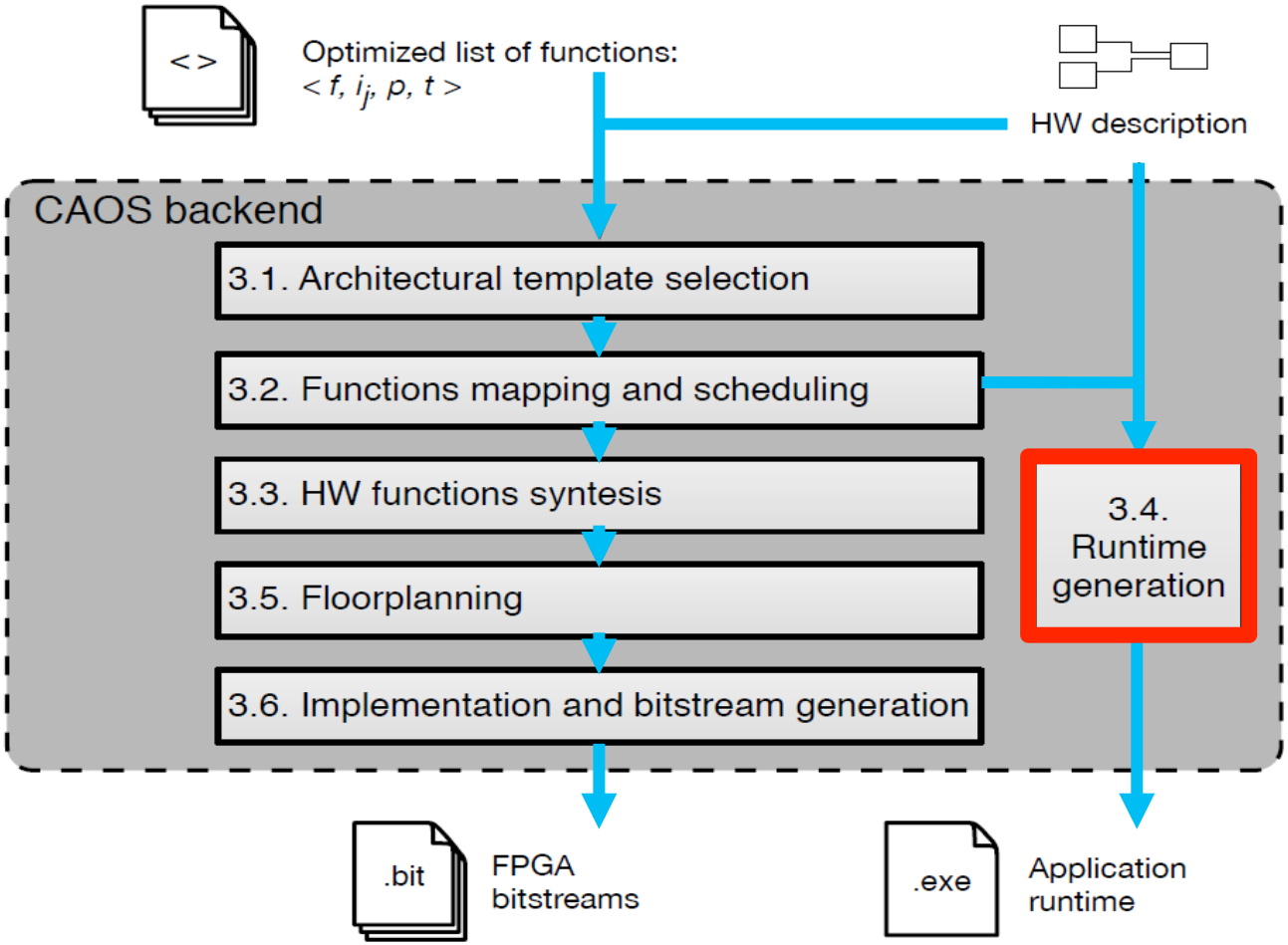
Output Generation

CAOS: OpenCL and SDAccel

- CAOS Frontend supports OpenCL code:
  - Intermediate representation support
  - Template applicability check for SDA
  - Code profiling through LTPV (OpenCL profiler)
  - Function optimization:
    - Static code analysis and HW resource estimation within SDA
  - Backend support for SDAccell



CAOS Backend for SDAccel



SDAccel generates & provides:

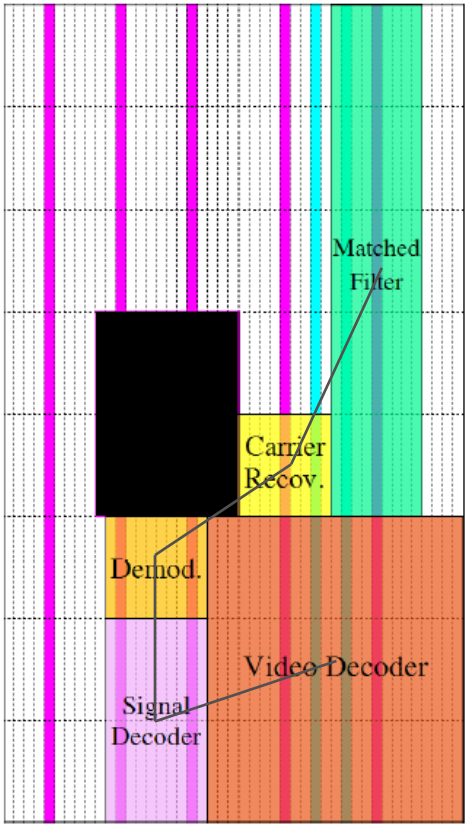
- **XCLBIN** containing the bitstream
- **OpenCL Runtime** to manage kernel execution

CAOS Integrates SDAccel:

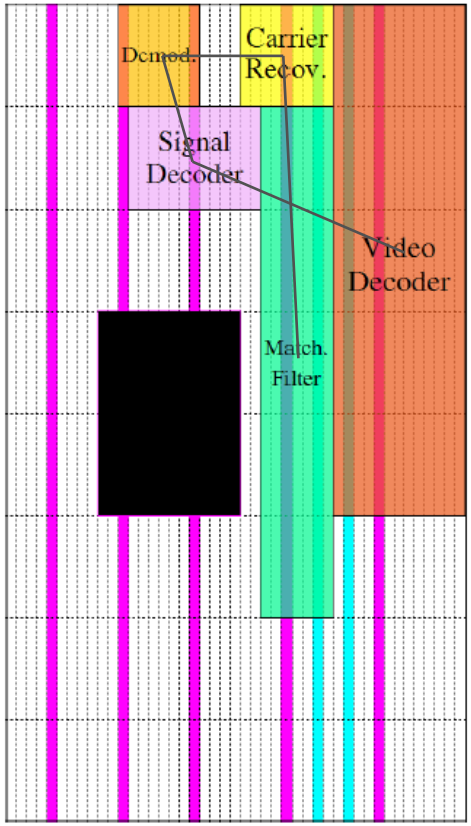
- Identifying I/O Variables
- Generating a specific OpenCL Host code for the application


RATIONALE BEHIND CAOS
A PRACTICAL EXAMPLE
22.3.3.55
85.3.66.1
TECHNOLOGY

Hints on the problem…



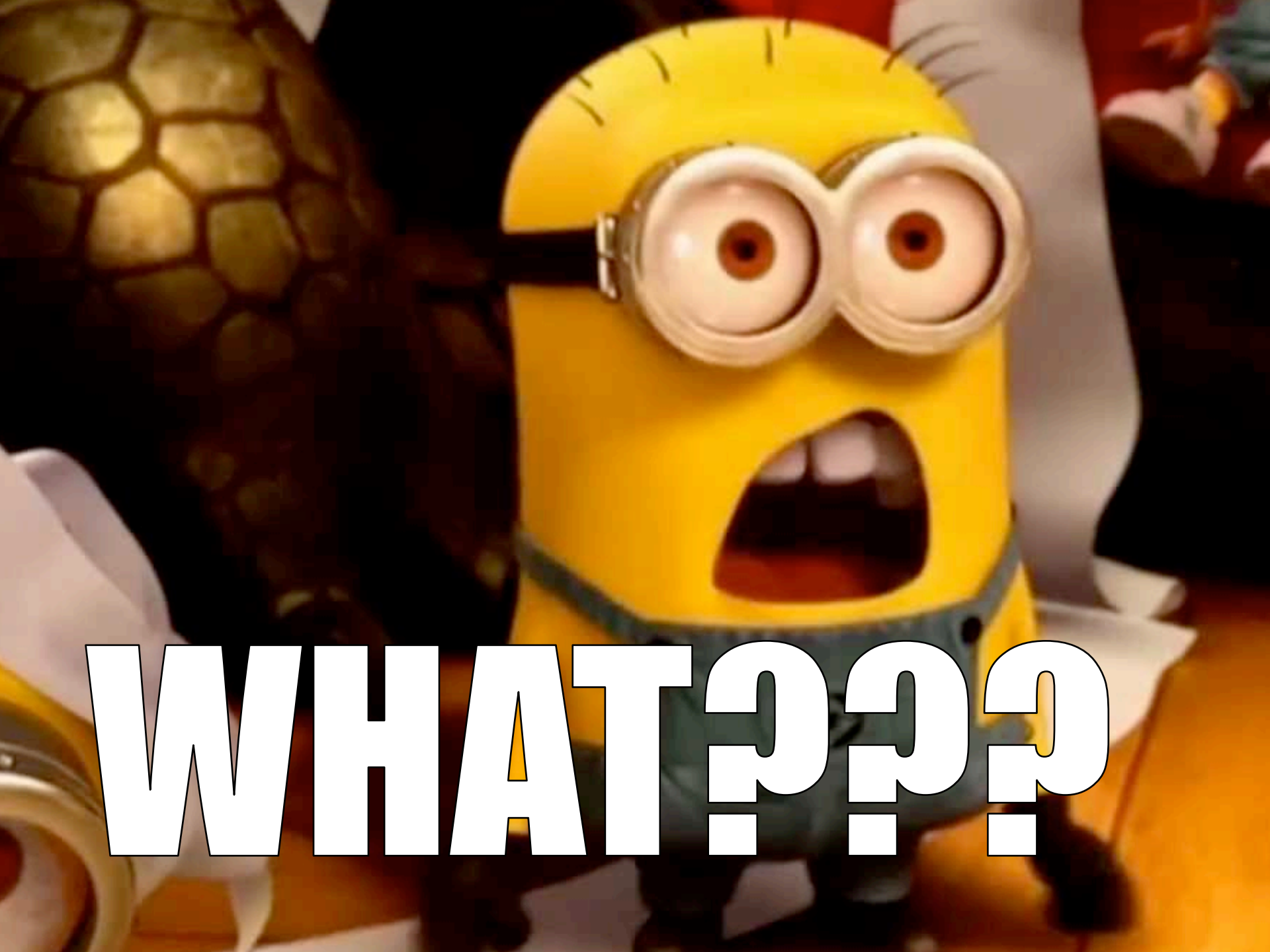
Floorplan produced by [*].



Floorplan produced by OF.

Which one is the best?

[*] Vipin, K. and Fahmy, S. A.: Architecture-aware reconfiguration-centric floorplanning for partial reconfiguration. In ARC, pages 13-25, 2012.

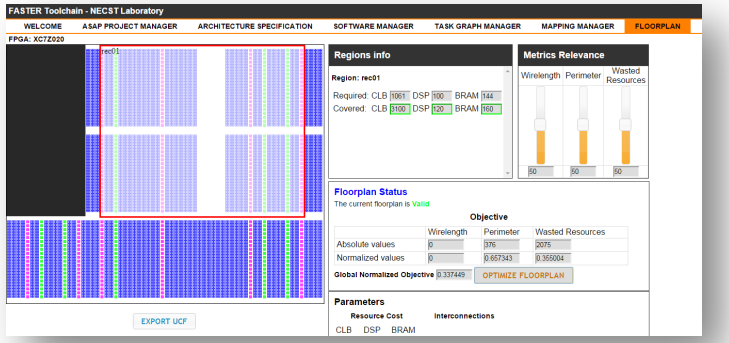

WHAT???

Objective function

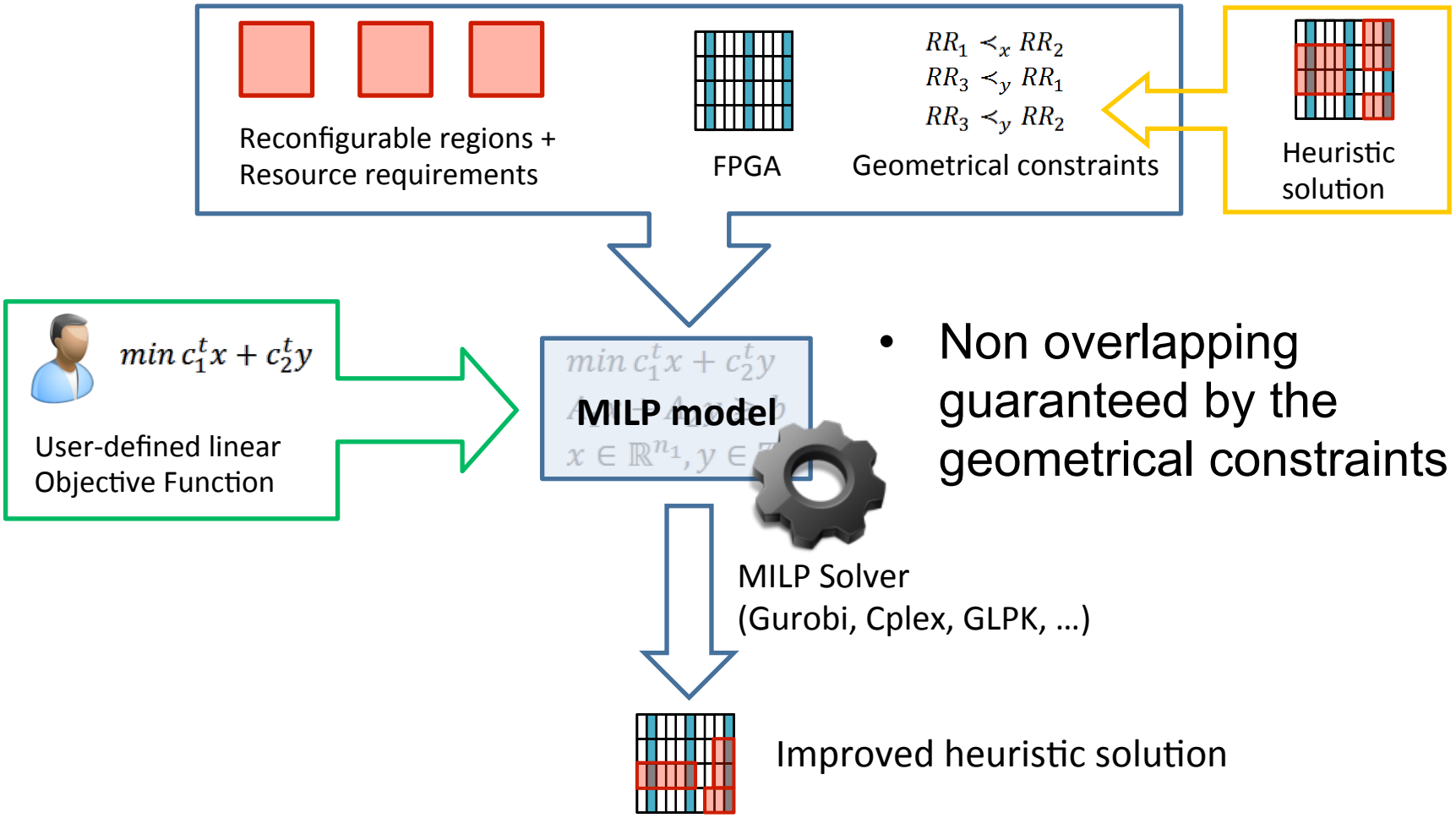
- Cost function can be defined starting from the variables and parameters of the MILP model

- Implemented metrics:
  - Global wirelength measured using HPWL (WL_{cost})
  - Regions perimeter (P_{cost})
  - Wasted resources (R_{cost})

$$min\left\{q_1 \cdot \frac{WL_{cost}}{WL_{max}} + q_2 \cdot \frac{P_{cost}}{P_{max}} + q_3 \cdot \frac{R_{cost}}{R_{max}}\right\}$$

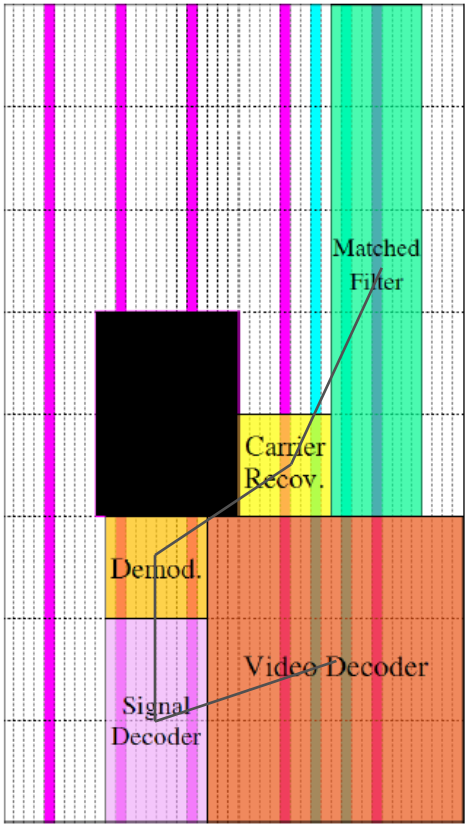


Heuristic-Optimal Floorplanner

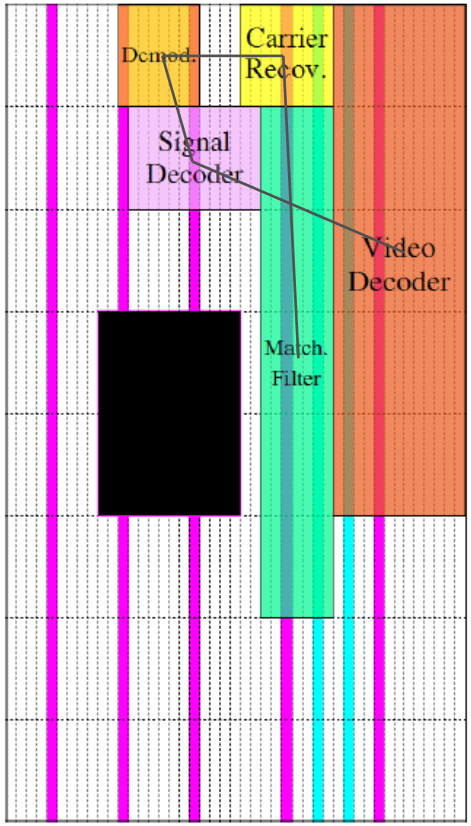


- Non overlapping guaranteed by the geometrical constraints

Hints on the problem…



Floorplan produced by [*].

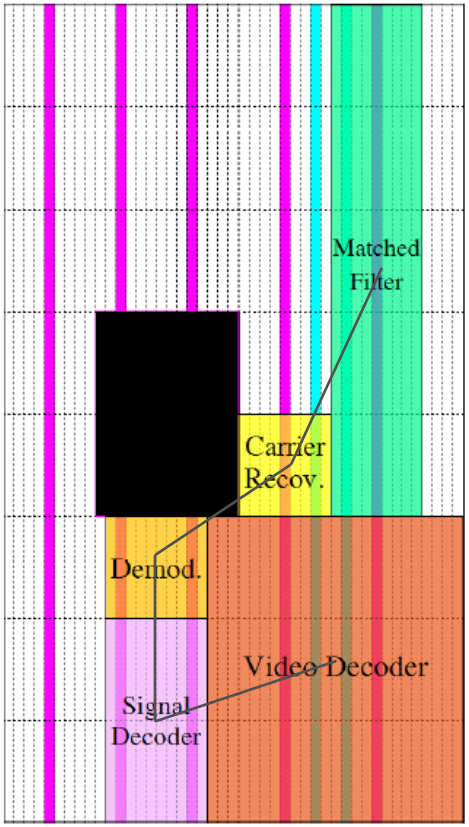


Floorplan produced by OF.

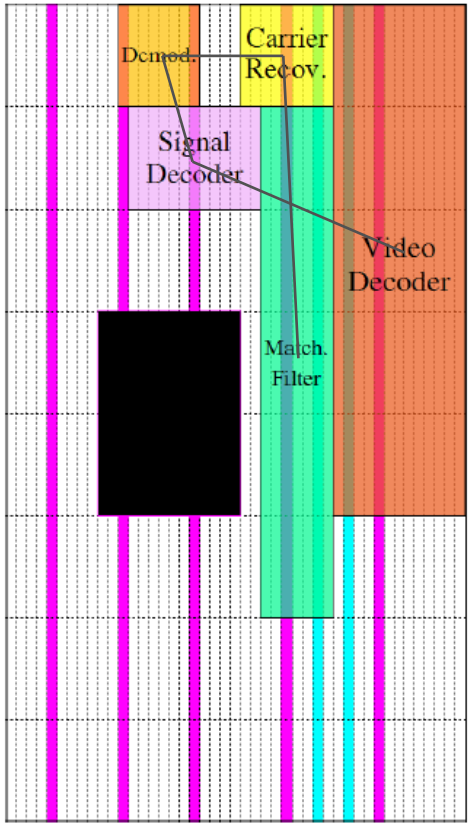
Which one is the best?

[*] Vipin, K. and Fahmy, S. A.: Architecture-aware reconfiguration-centric floorplanning for partial reconfiguration. In ARC, pages 13-25, 2012.

Hints on the problem…



Floorplan produced by [*].



Floorplan produced by OF.

- Optimal solution in 29s
- 34% wasted frames reduction
  - No DSP and CLB wasted by the Video Decoder RR
  - No BRAM wasted by the Signal Decoder RR
- Approximately same wirelength

[*] Vipin, K. and Fahmy, S. A.: Architecture-aware reconfiguration-centric floorplanning for partial reconfiguration. In ARC, pages 13-25, 2012.

Evaluations

[1, 2] Streaming Stencil Time-step (SST)
[3] Pearson Correlation Coefficient, Asian Option Pricing
[5] Protein Folding
[4] Smith Waterman and Vessels Segmentation

Case Study	Board	Improvement wrt CPU[*]	
		Performance	Energy Efficiency
[4]	Virtex 7	3.68x	11.8x
[4]	Kintex	14.15x	45x
[5]	Virtex 7	1.61x	15.29x
[3]	Virtex 7	3.1x	2.2x
[1]	Virtex 7	1.09x	12.9x
[2]	Virtex 7	0.22x	2.46x

[*] intel Xeon E5 1410
32 GB RAM


CAOS DEMOS
AND CASE STUDIES
YouTube
https://tinyurl.com/extrahpc

Some Applicative Domains for FPGA Acceleration

- Image and Video Processing
- Security
- Machine Learning
- Genomics
- Financial Analytics
- Big Data Analytics

edico genome

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

RYFT™
ACTIONABLE INTELLIGENCE FROM COMPLEX DATA

Who Victor is

How a Genetic Test Changed Victor's Life

History of Personalized Medicine Brings Future Hope to Lung Cancer Patients

After feeling a tickle in his throat for about a month, Victor visited the University of Chicago Medicine campus in June 2010 for a check-up. It had only been a very quick tickle, which caused him to clear his throat a half dozen or so times a day, but he wanted to make sure his health remained stable.

Science News *from research organizations*

Targeting breast cancer through precision medicine

The protein RYBP could make cancer cells more sensitive to DNA damage

Date: January 9, 2018

Source: University of Alberta Faculty of Medicine & Dentistry

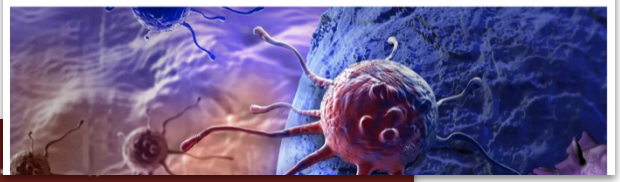
Summary: Researchers have discovered a mechanism that may make cancer cells more susceptible to treatment. The research team found that the protein RYBP prevents DNA repair in cancer cells, including breast cancer.

Genetic test helps pick the right drugs for mental health

Updated: OCTOBER 25, 2017 — 12:59 PM EDT

Personalized cancer vaccines successful in first-stage human trials

Rich Haridy | July 10th, 2017



nature

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

News & Comment > News > 2018 > February > Article

NATURE | NEWS

Personalized cancer vaccines show glimmers of success

Treatments tailored to a person's individual cancer mutations train immune system to attack tumours.

Heidi Ledford

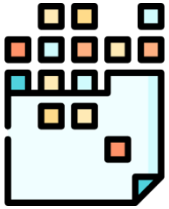
05 July 2017

Open Challenges

- It is necessary to keep-up with continuous development of biological research

Open Challenges

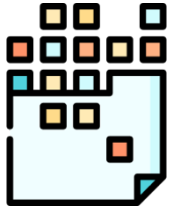
- It is necessary to keep-up with continuous development of biological research



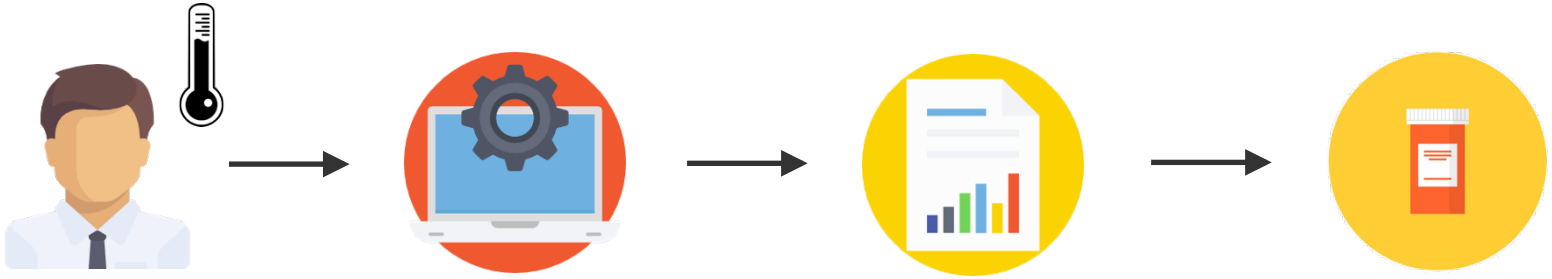
- Each individual DNA provides huge amount of data

Open Challenges

- It is necessary to keep-up with continuous development of biological research



- Each individual DNA provides huge amount of data
- To produce a tailor-made drug, for each DNA:



Personalized Medicine Today

- FPGA-based acceleration
  - optimal ratio performance/power consumption
  - reconfigurability
- Possibility to use pre-accelerated biological pipelines
- Available on-site or for AWS cloud

edico genome

HUGenomics

An advanced **support for genomic research** that,

by means of **reconfigurable hardware accelerators**,

is capable of delivering massive performance
for fast-changing algorithms, letting researchers

to focus on delivering **best-in-class results** in the **least amount of time**

Rationale Behind HUG

NO hardware competences required

Possibility to handle **massive amount of data**

Possibility to integrate **custom code**

Reduction in research time

Genome Assembly

Smith-Waterman
(Software)

30h

Haplotype Caller - PairHMM
(Software)

10h

Smith-Waterman
(HUGenomics)

2,5 mins

Haplotype Caller - PairHMM
(HUGenomics)

17s

---

up to

658x

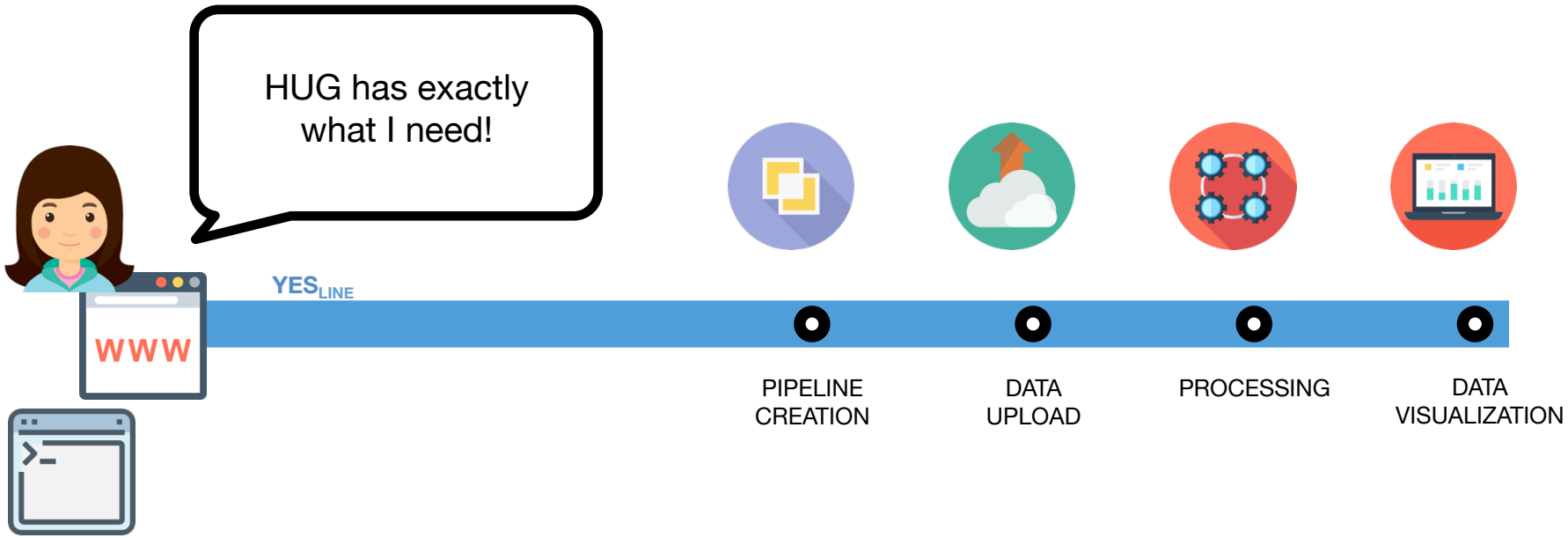
Performance Improvement

up to

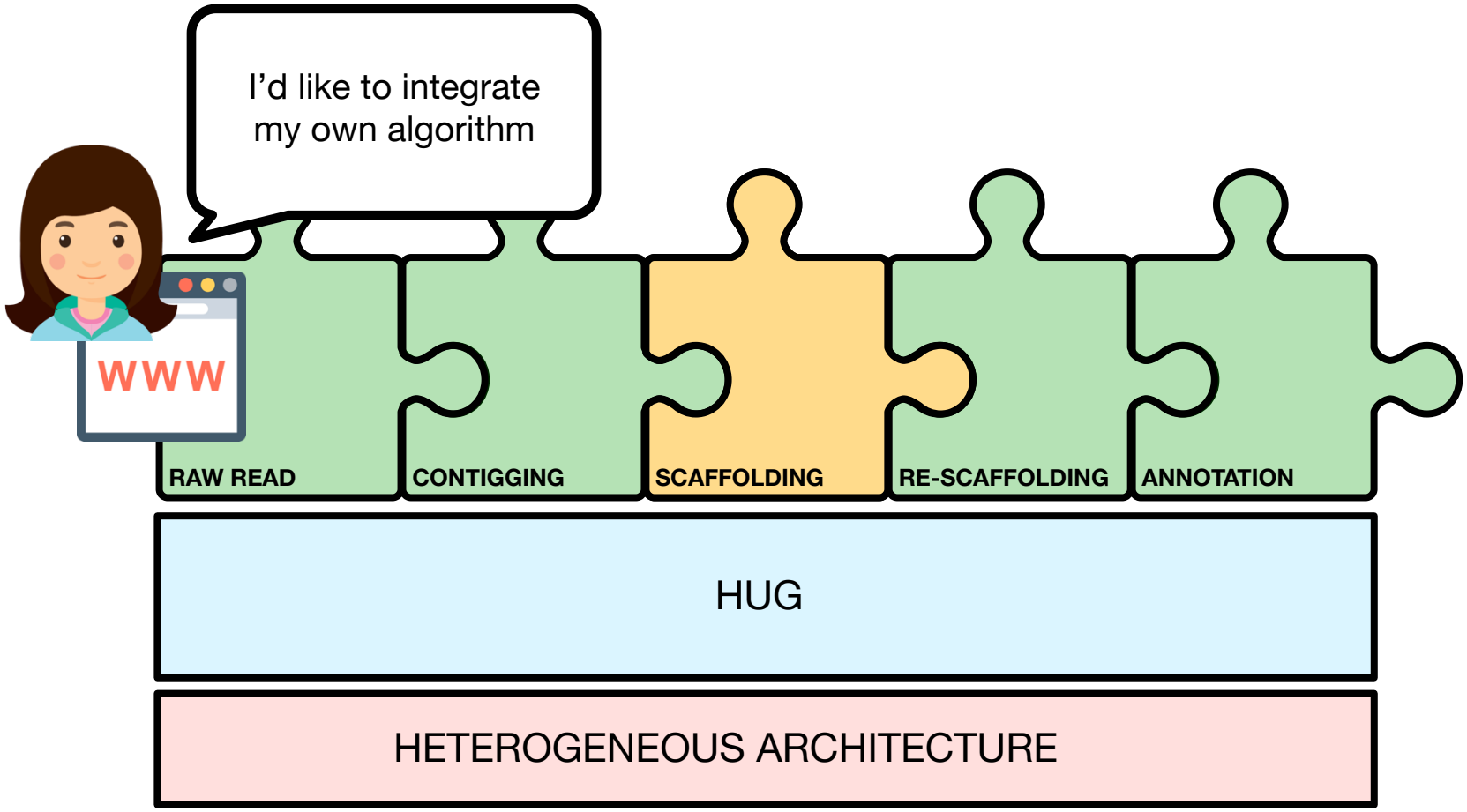
2160x

Performance Improvement

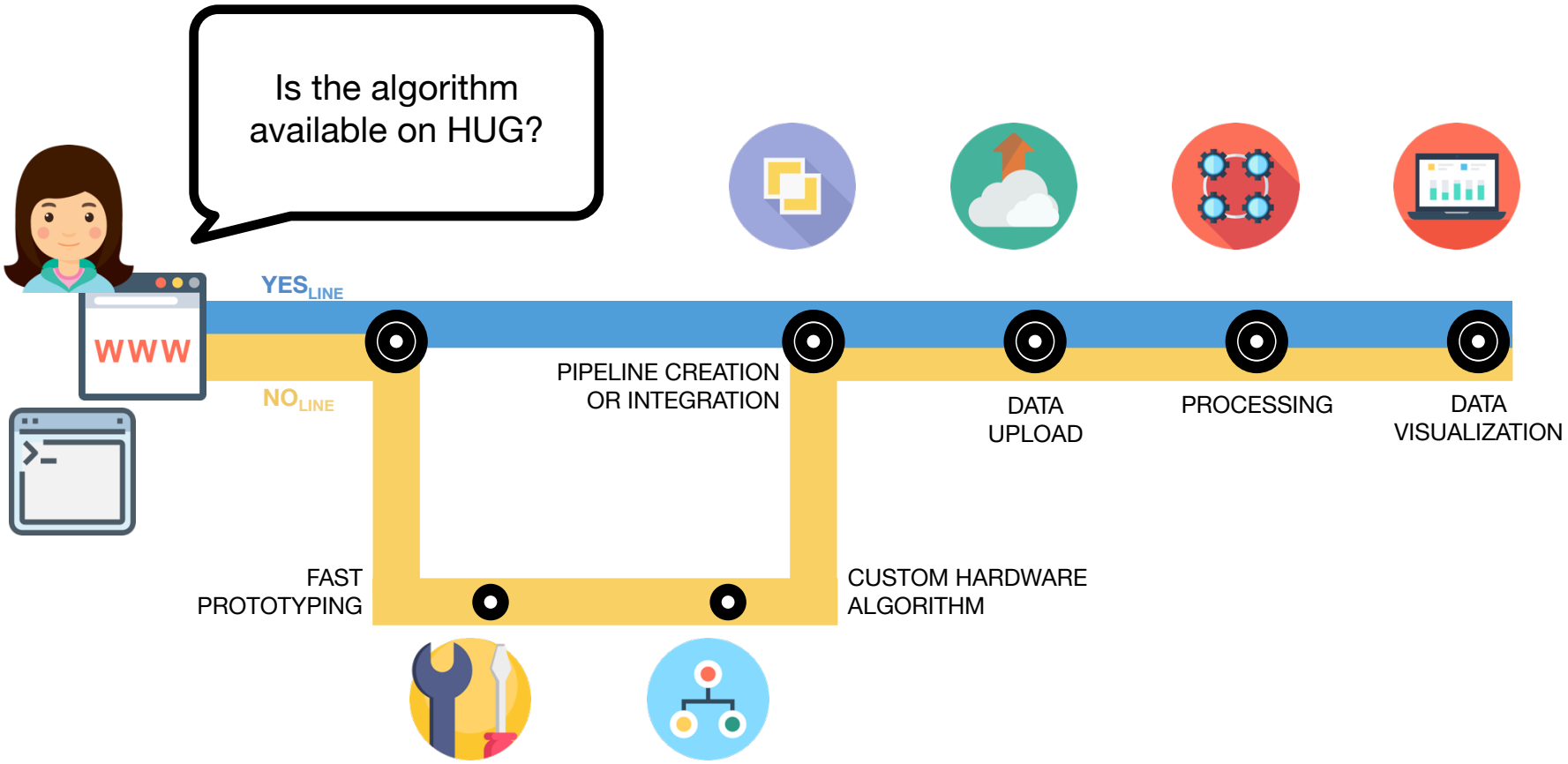
Genomics HW Pipeline



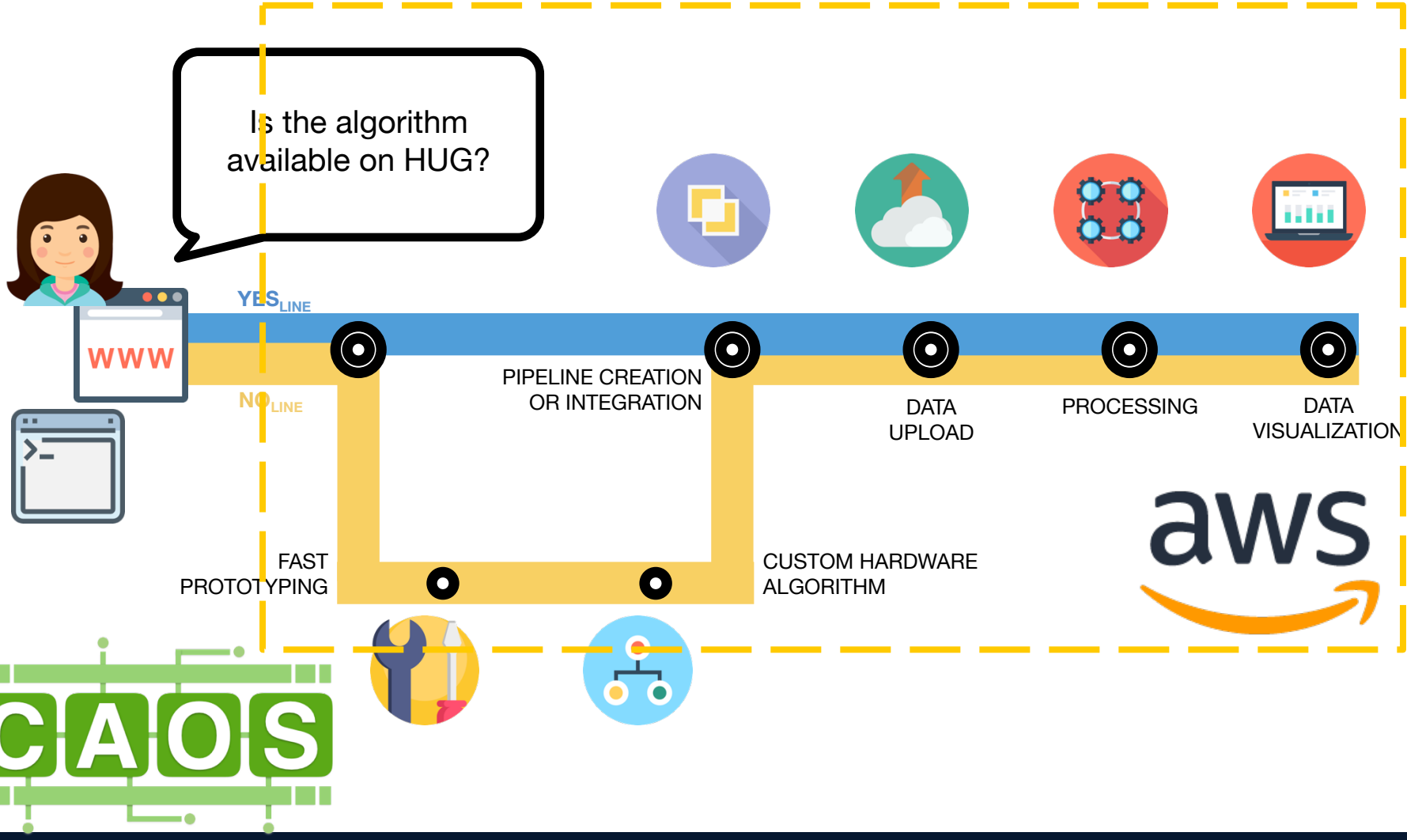
Custom Code Integration



Genomics HW Pipeline



Genomics HW Pipeline



Benefits of the AWS F1 Cloud Compute Platform

- Makes FPGA acceleration available to a large community of developers, and to millions of potential AWS users
- Provides dedicated and large amounts of FPGA logic with elasticity to scale to multiple FPGAs
- Simplifies the development process by providing cloud-based FPGA development tools
- Provides a Marketplace for FPGA applications, giving more choice, secure and easy access to millions of AWS users


Life is sharing

Bringing the Right People Together

coursera

XILINX

amazon web services

Bringing the Right People Together

coursera

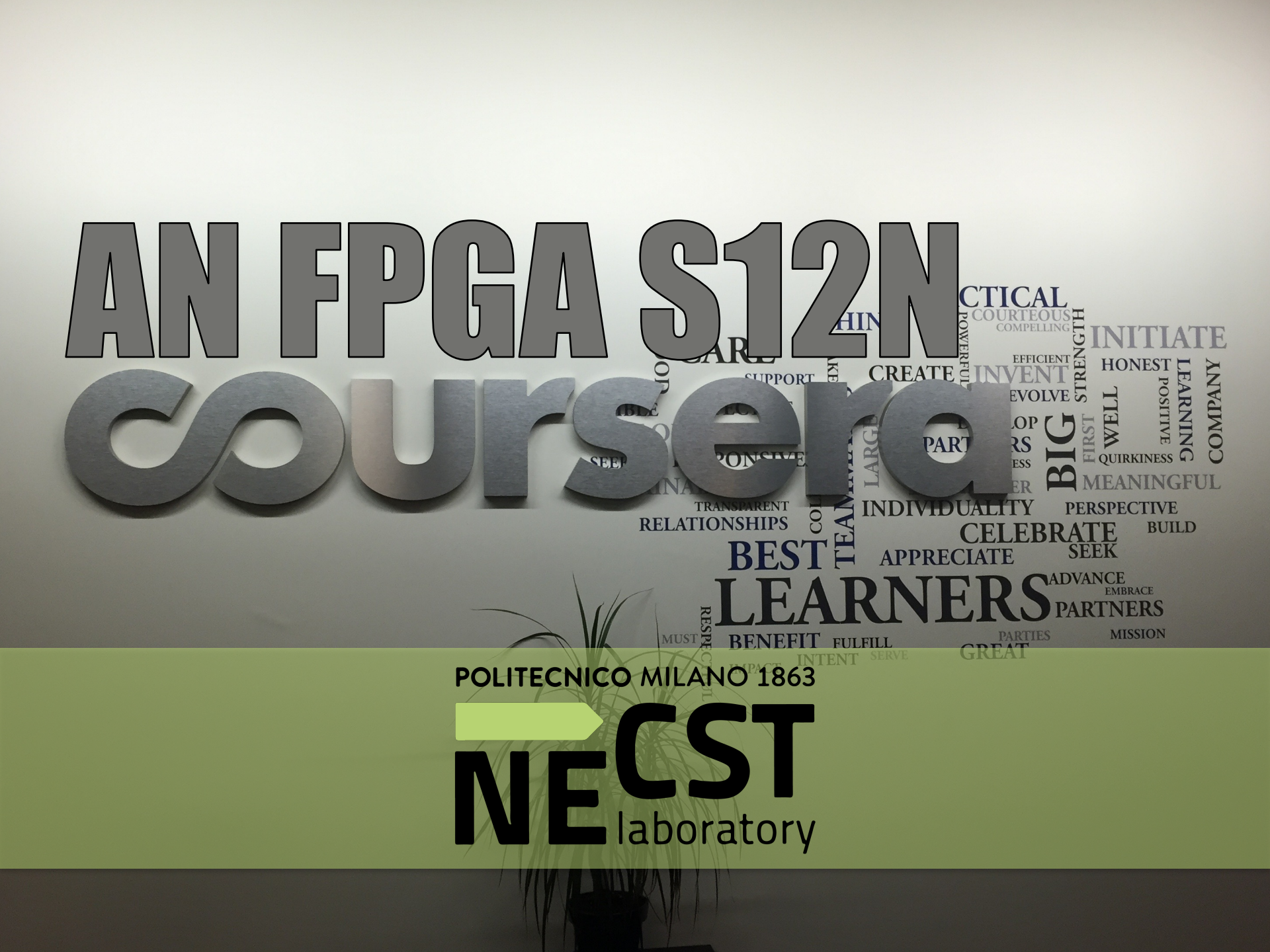
XILINX

amazon web services

POLITECNICO MILANO 1863

NECST laboratory

AN FPGA S12N



POLITECNICO MILANO 1863

NECST laboratory

FPGA computing systems: Background knowledge and introductory materials

Go to Course

Already enrolled

Apply for F

review Course Materials

About this course: This course is for anyone passionate in learning how a hardware component can be adapted at runtime to better respond to users/environment needs. This adaptation can be provided by the designers, or it can be an embedded characteristic of the system itself. These runtime adaptable systems will be implemented by using FPGA technologies.

More

Who is this class for: Anyone with moderate computer experience should be able to master the materials in this course. This is an introductory course to FPGA, therefore within this context no specific background

included in this course or as websites/handbooks that can be easily found/accessed.

Created by: Politecnico di Milano

POLITECNICO
MILANO 1863

Taught by: Marco Domenico Santambrogio, Associate Professor
DEIB - Dept. of Electronics, Information and Bioengineering

intrigano

3.243 iscritti

ISCRIVITI 3.243

HOME PAGE | VIDEO | PLAYLIST | CANALI | DISCUSSIONE | INFORMAZIONI

Video caricati

RIPRODUCI TUTTI

ncorporating abstraction

visualizzazioni • 5 giorni fa

The objectivity of Swiss Design

7 visualizzazioni • 5 giorni fa

Introduction to Graphic Design - Radicalism

9 visualizzazioni • 5 giorni fa

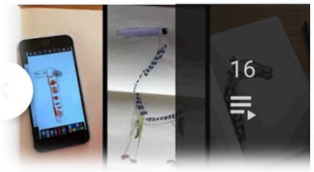
Modernism - Wolliam Golden - A clear identity

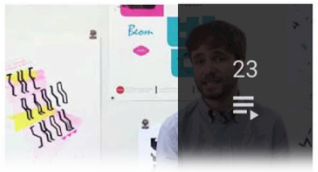
5 visualizzazioni •
1 settimana fa

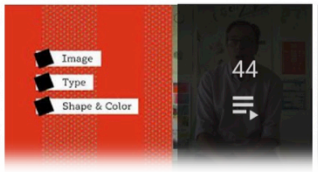
Modernism - Paul Rand - Play and humor

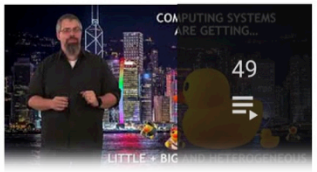
3 visualizzazioni •
1 settimana fa

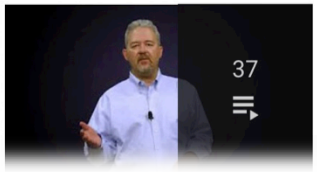
Playlist create











CANALI CORRELATI

Last moment tuitions
ISCRIVITI

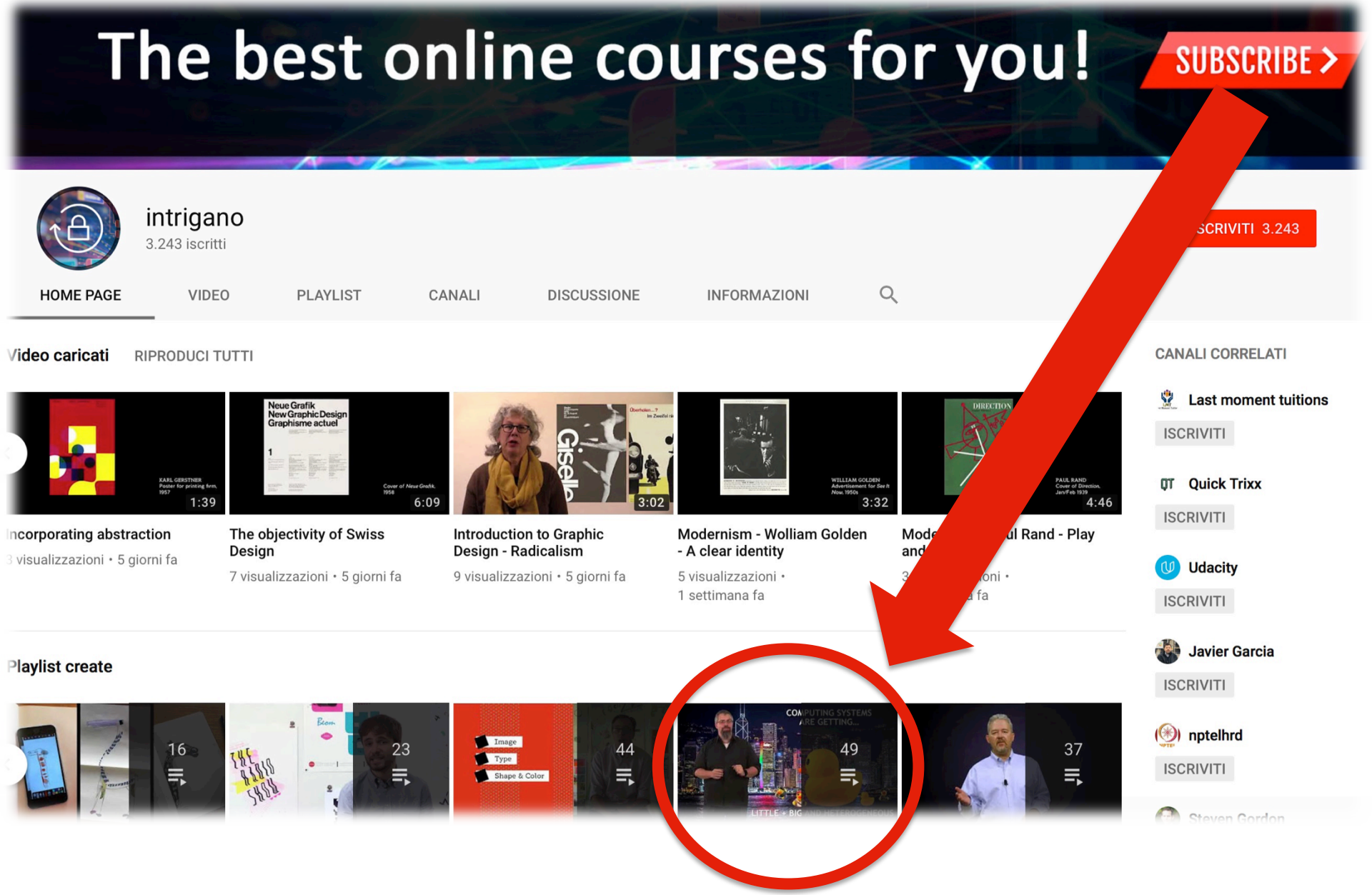
Quick Trixx
ISCRIVITI

Udacity
ISCRIVITI

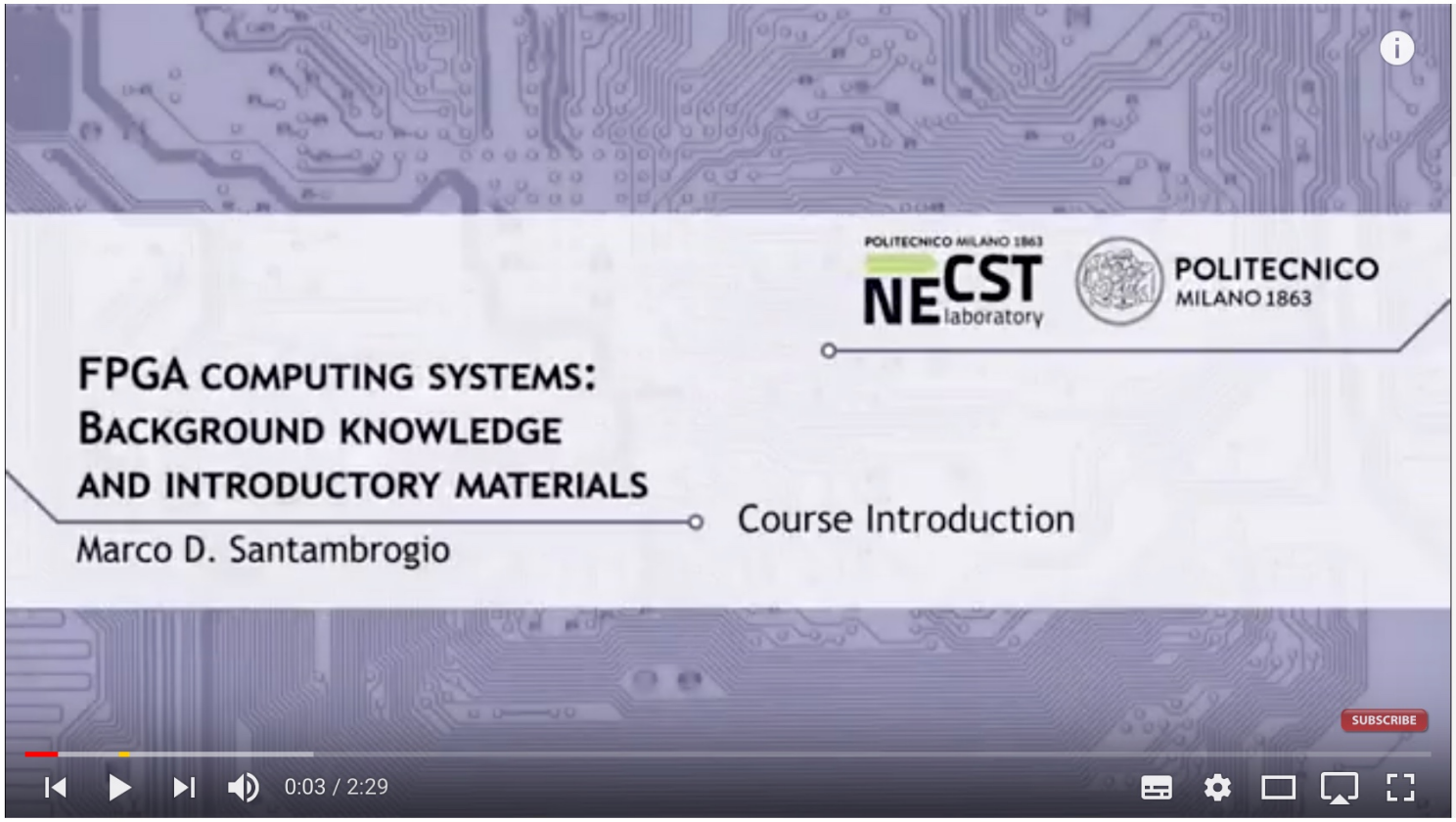
Javier Garcia
ISCRIVITI

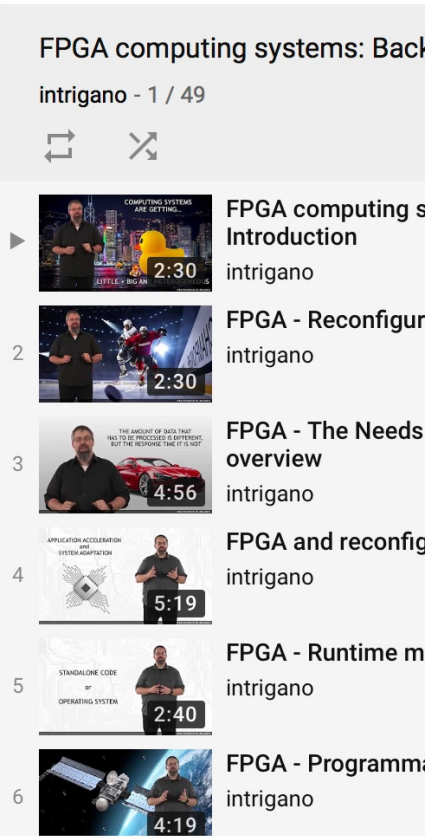
nptelhrd
ISCRIVITI

Steven Gordon

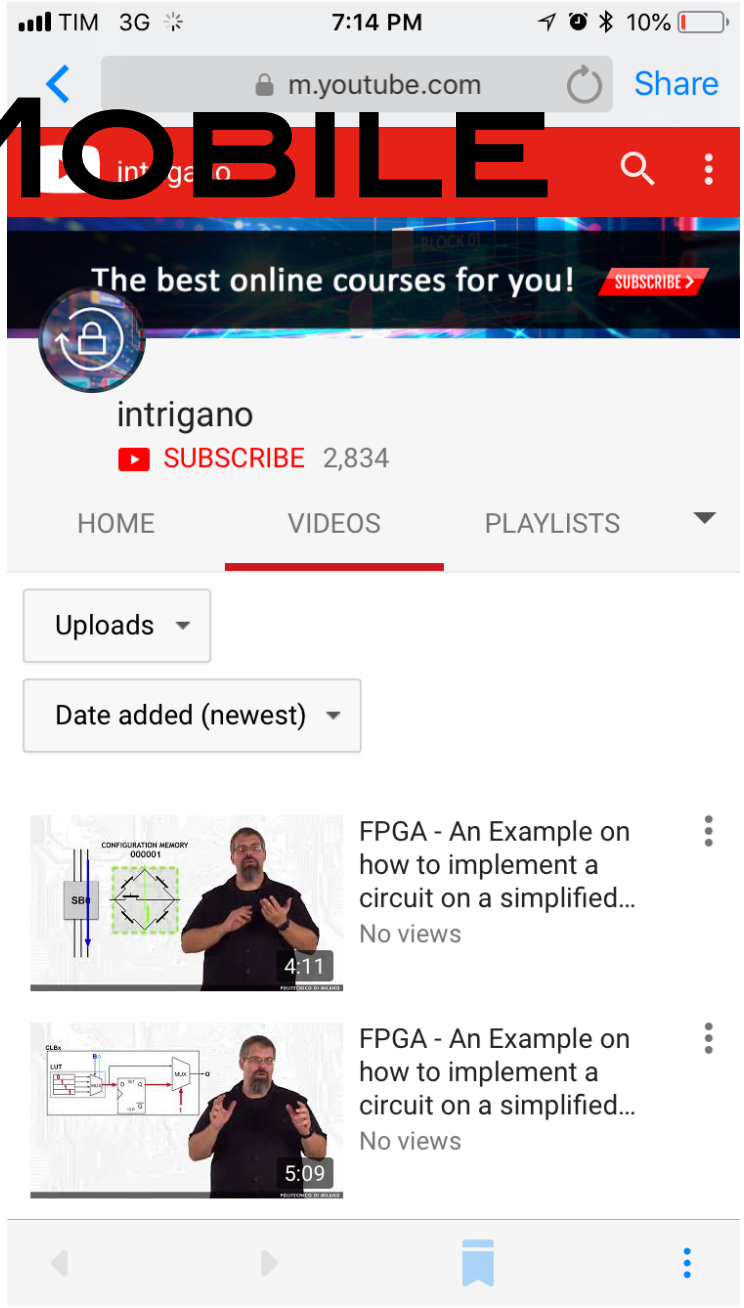

The best online courses for you!
SUBSCRIBE
intrigano
3.243 iscritti
SCRIVITI 3.243
HOME PAGE
VIDEO
PLAYLIST
CANALI
DISCUSSIONE
INFORMAZIONI
Video caricati
RIPRODUCI TUTTI
1:39
Incorporating abstraction
visualizzazioni • 5 giorni fa
Neue Grafik
New Graphic Design
Graphisme actuel
6:09
The objectivity of Swiss Design
7 visualizzazioni • 5 giorni fa
Gisela
3:02
Introduction to Graphic Design - Radicalism
9 visualizzazioni • 5 giorni fa
3:32
Modernism - Wolliam Golden - A clear identity
5 visualizzazioni •
1 settimana fa
4:46
Playlist create
16
23
Image
Type
Shape & Color
44
49
37
CANALI CORRELATI
Last moment tuitions
ISCRIVITI
Quick Trixx
ISCRIVITI
Udacity
ISCRIVITI
Javier Garcia
ISCRIVITI
nptelhrd
ISCRIVITI

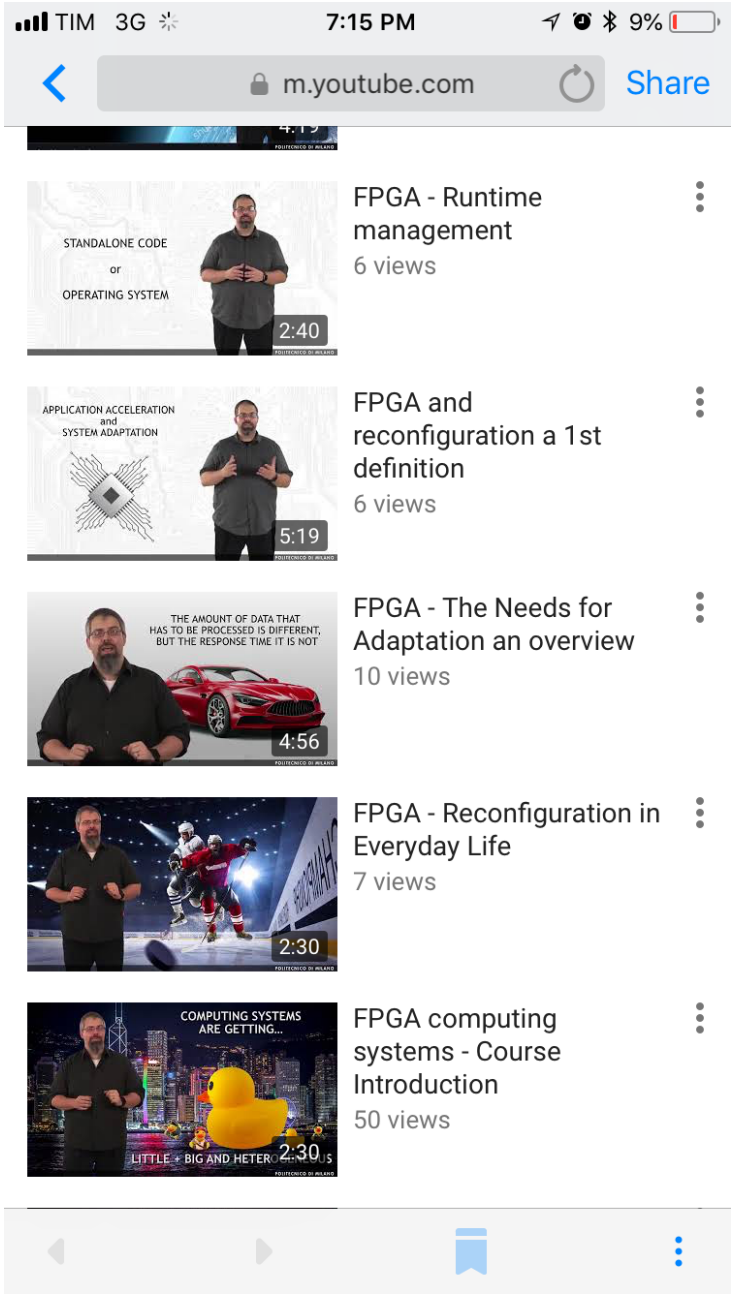
PC/Laptop


POLITECNICO MILANO 1863
NECST laboratory
POLITECNICO MILANO 1863
FPGA COMPUTING SYSTEMS:
BACKGROUND KNOWLEDGE
AND INTRODUCTORY MATERIALS
Course Introduction
Marco D. Santambrogio
SUBSCRIBE
0:03 / 2:29


FPGA computing systems: Back
intrigano - 1 / 49
FPGA computing s
Introduction
intrigano
2:30
2
FPGA - Reconfigur
intrigano
2:30
3
FPGA - The Needs
overview
intrigano
4:56
4
FPGA and reconfig
intrigano
5:19
5
FPGA - Runtime m
intrigano
2:40
6
FPGA - Programma
intrigano
4:19

MOBILE


TIM 3G 7:14 PM 10%
m.youtube.com
Share
The best online courses for you!
SUBSCRIBE
intrigano
SUBSCRIBE 2,834
HOME
VIDEOS
PLAYLISTS
Uploads
Date added (newest)
FPGA - An Example on how to implement a circuit on a simplified...
No views
4:11
FPGA - An Example on how to implement a circuit on a simplified...
No views
5:09


TIM 3G 7:15 PM 9%
m.youtube.com
Share
STANDALONE CODE
or
OPERATING SYSTEM
2:40
FPGA - Runtime management
6 views
APPLICATION ACCELERATION
and
SYSTEM ADAPTATION
5:19
FPGA and reconfiguration a 1st definition
6 views
THE AMOUNT OF DATA THAT HAS TO BE PROCESSED IS DIFFERENT, BUT THE RESPONSE TIME IT IS NOT
4:56
FPGA - The Needs for Adaptation an overview
10 views
2:30
FPGA - Reconfiguration in Everyday Life
7 views
COMPUTING SYSTEMS ARE GETTING...
2:30
FPGA computing systems - Course Introduction
50 views

COMING SOON

coursera FEB 2019

POLITECNICO MILANO 1863
NECST laboratory

POLITECNICO MILANO 1863

DEVELOPING FPGA-ACCELERATED CLOUD APPLICATIONS WITH SDACCEL

Marco D. Santambrogio

An introduction to the AWS EC2 F1 instances

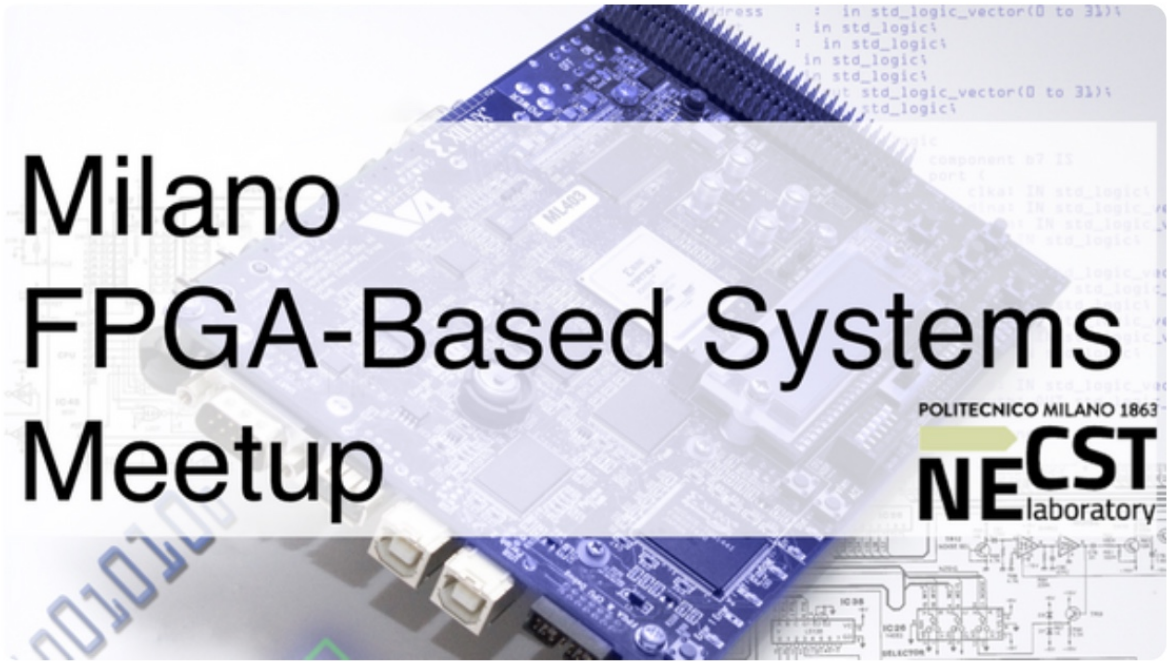
NOT 1, BUT 2 COURSES

WORK TOGETHER AND CREATE A COMMUNITY

WITH ALSO LOCAL MEETINGS AND LIVE EVENTS

Meetup

Start a new group | Log in



Milano FPGA-Based Systems Meetup

Milano, Italy

41 members · Public group

Organized by **Marco D. S.**

Share:

What We Have

What We Have

What We Have

What We Have

coursera

A worldwide class to share knowledge and preview game changing technologies

What We Have

coursera

A worldwide class to share knowledge and preview game changing technologies

CAD for efficient HW/SW solutions for high performance FPGA-based systems

What We Have

An advanced support to genomic research by heterogeneous HW architectures

CAD for efficient HW/SW solutions for high performance FPGA-based systems

coursera

A worldwide class to share knowledge and preview game changing technologies

What We Have

An advanced support to genomic research by heterogeneous HW architectures

CAD for efficient HW/SW solutions for high performance FPGA-based systems

coursera

A worldwide class to share knowledge and preview game changing technologies

What We Have

An advanced support to genomic research by heterogeneous HW architectures

CAD for efficient HW/SW solutions for high performance FPGA-based systems

coursera

A worldwide class to share knowledge and preview game changing technologies

What We Have

An advanced support to genomic research by heterogeneous HW architectures

coursera

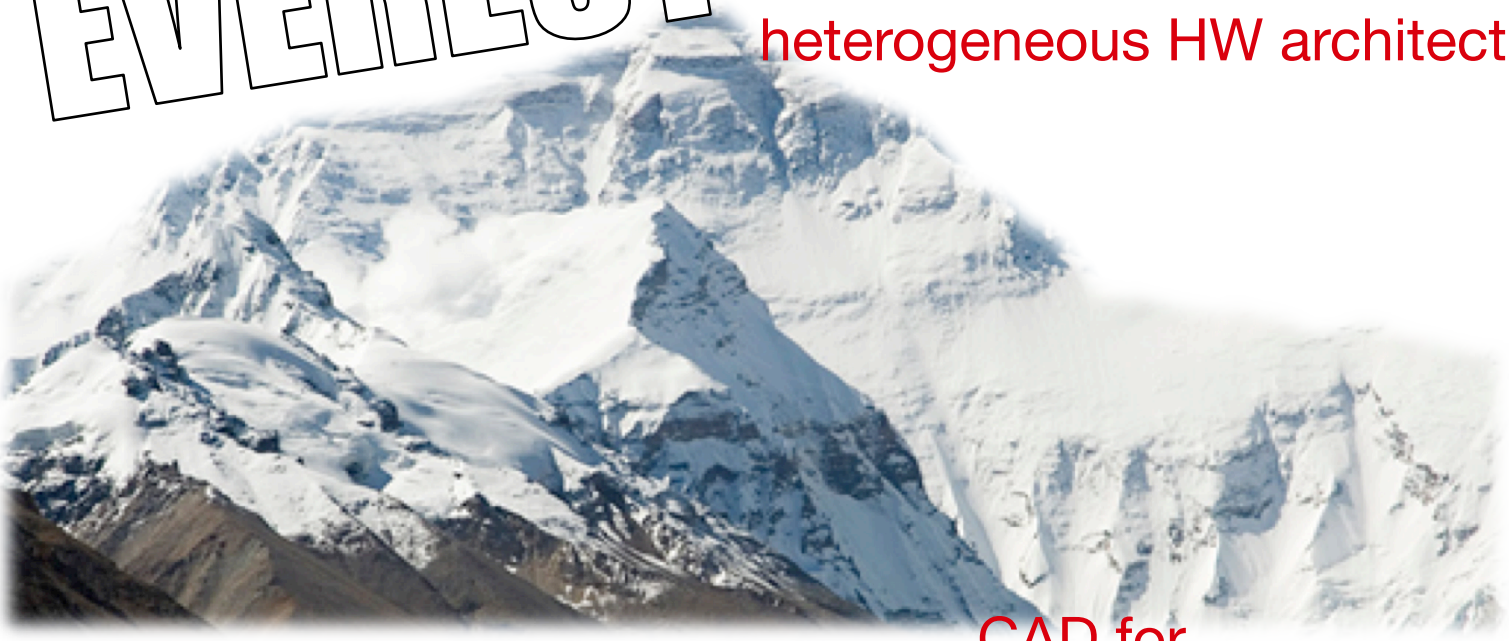
A worldwide class to share knowledge and preview game changing technologies

CAD for efficient HW/SW solutions for high performance FPGA-based systems

What We Have

EVEREST

heterogeneous HW architectures



CAD for

high performance

FPGA-based systems

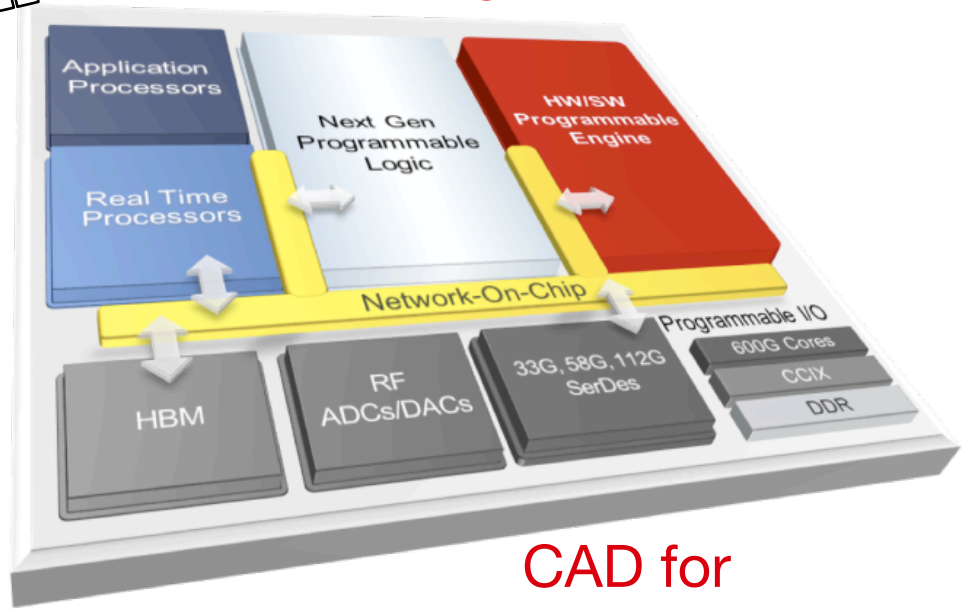
preview game

changing technologies

What We Have

Adaptive **C**ompute **A**cceleration **P**latform

heterogeneous HW architectures



https://www.anandtech.com/show/12509/xilinx-announces-project-everest-fpga-soc-hybrid

CAD for

high performance

FPGA-based systems

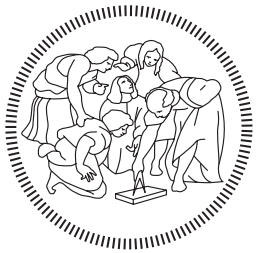
preview game

changing technologies

THANKS TO...

https://www.extrahpc.eu/

UNIVERSITY OF AMSTERDAM



POLITECNICO
MILANO 1863


THANKS TO...
R. Cattaneo, G. Guidi, L. Di Tucci, E. Del Sozzo, G. Durelli, M. Rabozzi, L. Stornaiuolo, S. Notargiacomo...
... and seveveral MSc Students in CS @ PoliMI

The NECSTLab Multi-Faceted Experience with AWS F1

Teaching, Research, Framework and Application stack

POLITECNICO MILANO 1863
NECST laboratory

Marco D. Santambrogio
<marco.santambrogio@polimi.it>
Politecnico di Milano

POLITECNICO
MILANO 1863