THE EUROPEAN APPROACH FOR EXSCALE AGES

JEAN-MARC.DENIS@EUROPEAN-PROCESSOR-INITIATIVE.EU

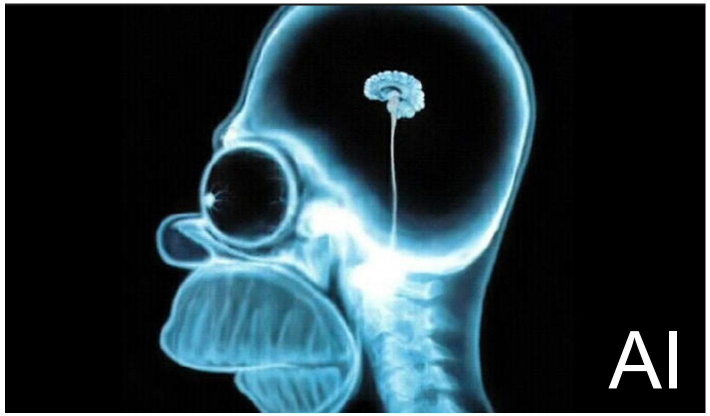
CHAIR OF THE BOARD


Cambrian explosion
Achieving performance through specialization
Courtesy Steve Scott
Cray CTO (2019)

TOP10 OVER THE LAST 10 YEARS

	2009 – Nov.	2014 – Nov.	2019 – Nov.	2020 – Jun.	*(Post)* Exascale
CPU **only**	9	5	2	3	0
CPU + ACC.	1	5	8	7	10

WHY? SOME OBVIOUS REASONS…

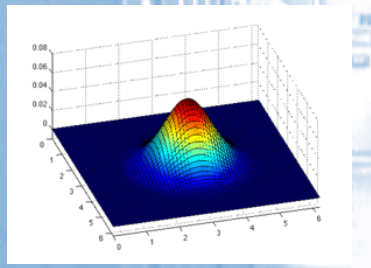


 ESiWACE2 Virtual Workshop on Emerging Technologies for Weather and Climate Modelling

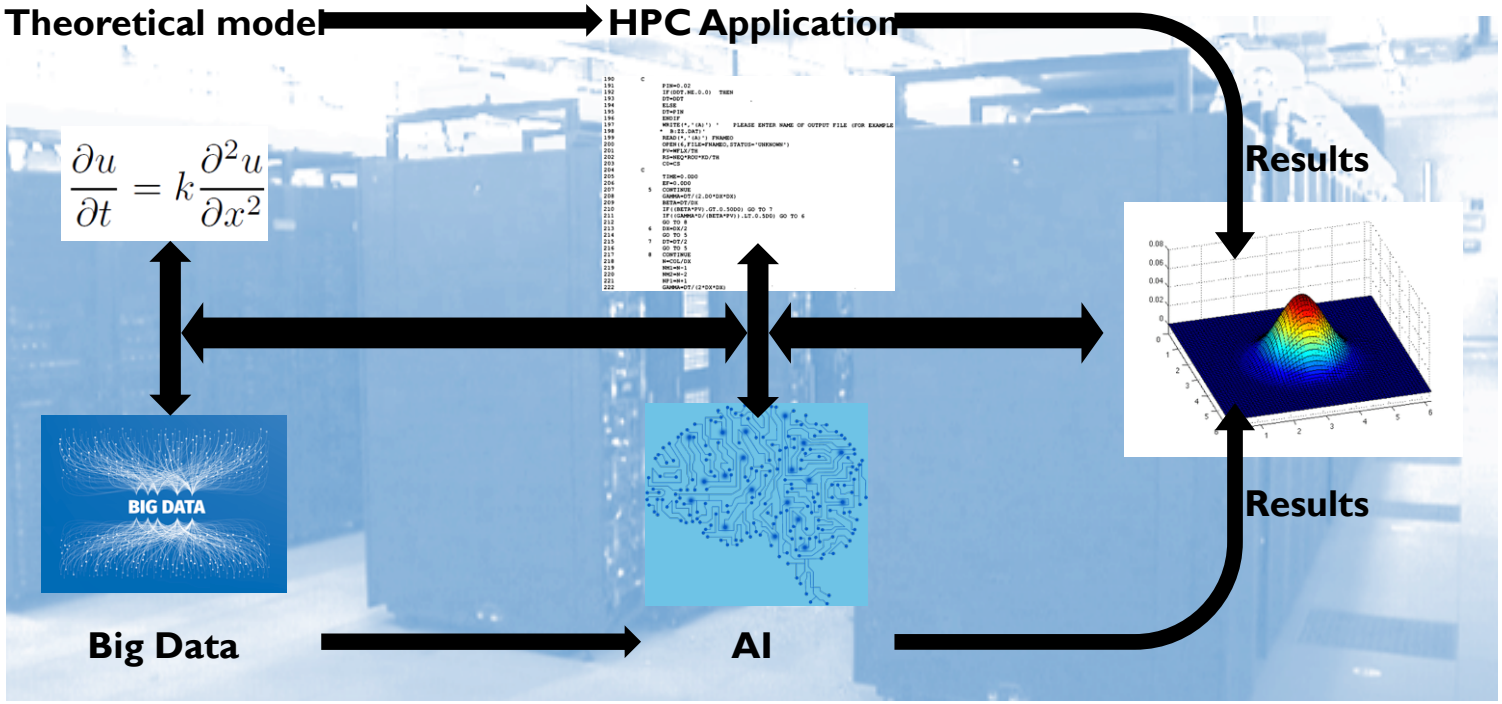
HPC BEFORE ARTIFICIAL INTELLIGENCE

Theoretical model → **HPC Application** → **Results**

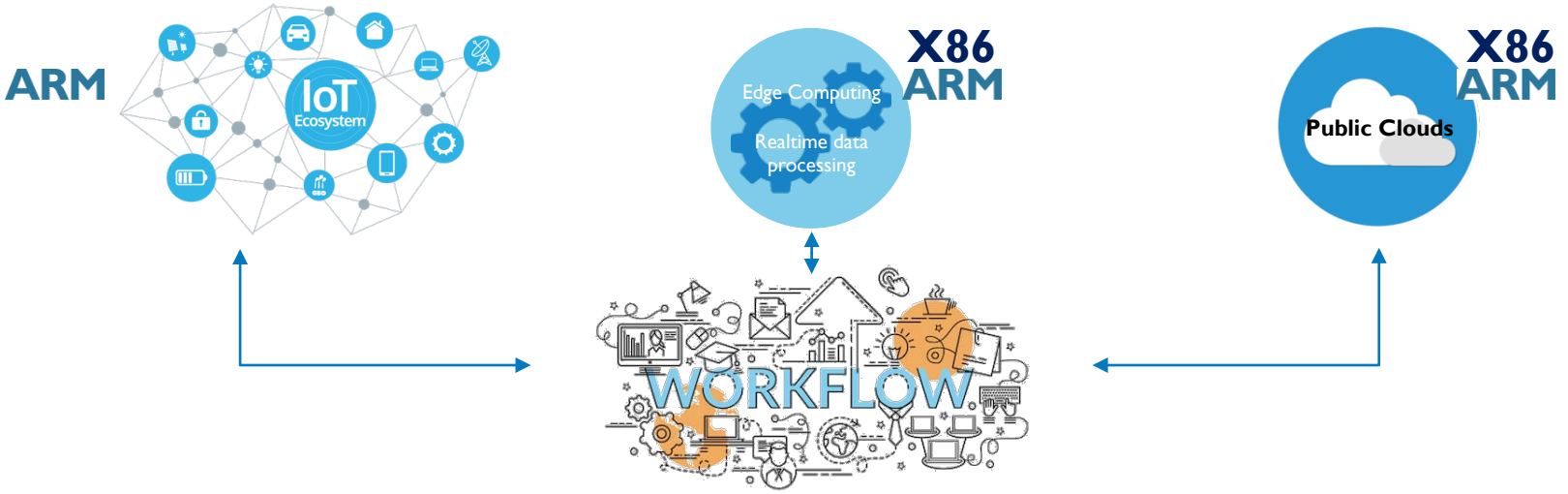
$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}$$



HPC WITH ARTIFICIAL INTELLIGENCE

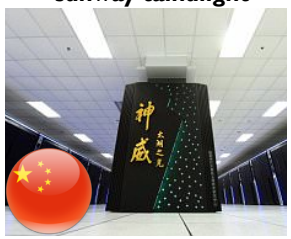


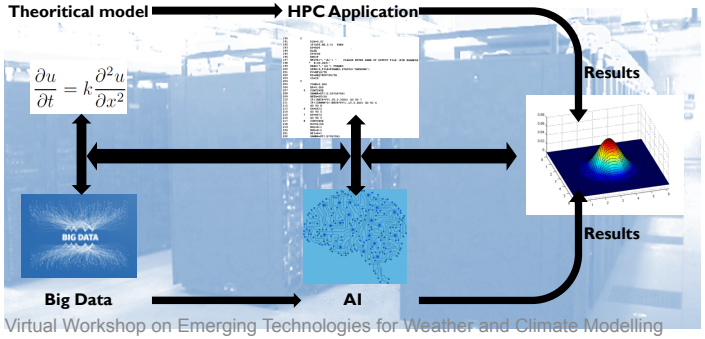
HPC & AI AT EXASCALE: IT'S ALL ABOUT WORKFLOWS (1/2)



SUMMIT

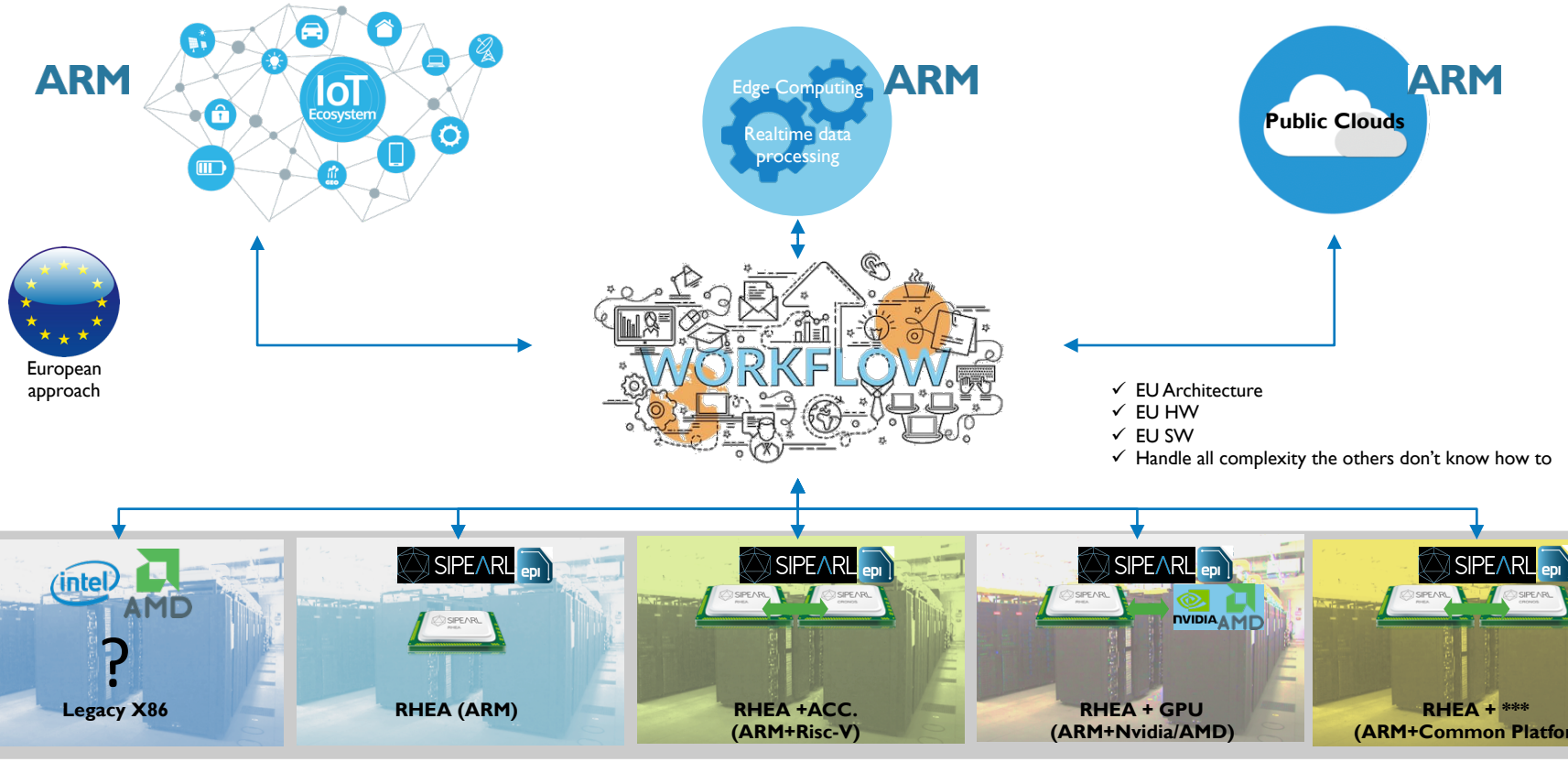
sunway taihulight



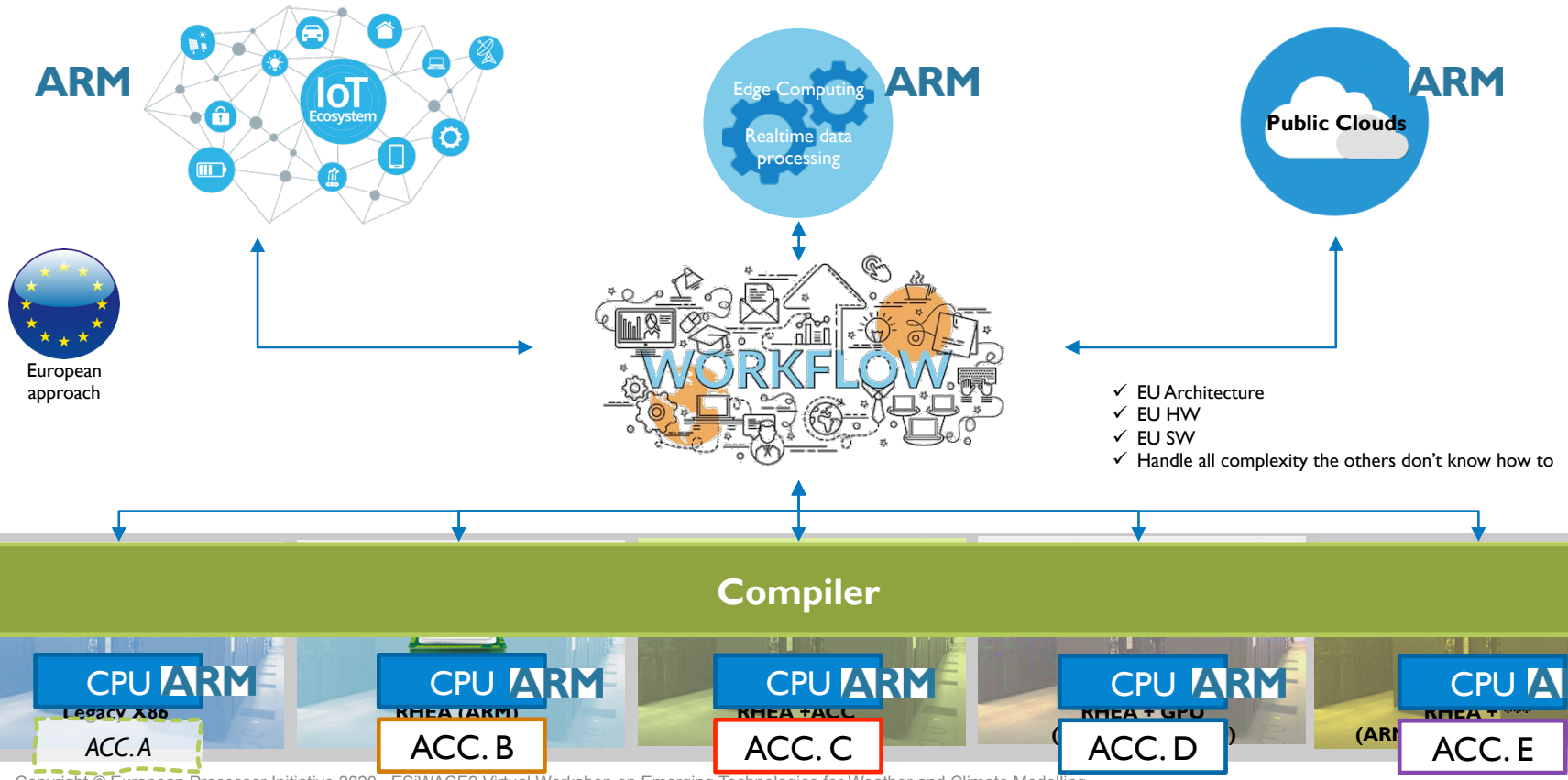


FUGAKU

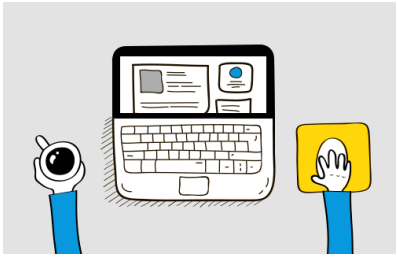
HPC & AI AT EXASCALE: IT'S ALL ABOUT WORKFLOWS (2/2)

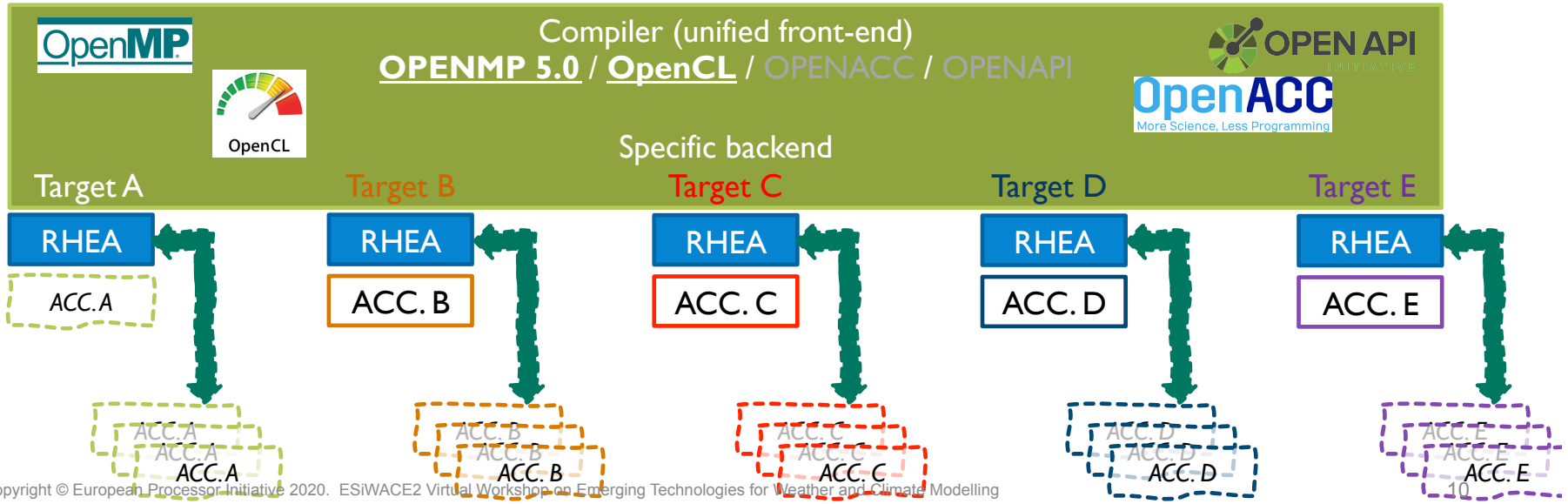


HPC & AI AT EXASCALE: IT'S ALL ABOUT WORKFLOWS (2/2)



THE DEVELOPER / USER STANDPOINT





EVIDENCE: INTEL OVERALL STRATEGY IN HPC, CLOUD, EDGE

Intel own accelerator developments:
- GPU (Artic Sound)
- CSA

Accelerator companies acquired by Intel or with intel in their Capital, in the last years:
- Altera
- Habana
- Mobileye
- Untether AI* (Toronto, Ontario, Canada)
- SambaNova Systems* (Palo Alto, California, U.S.)
- Zhuhai EEasy Technology Co. Ltd.* (Zhuhai, China)
- …

CPU - Accelerator interface:
- CXL

User environment:
- Intel Compiler
- OneAPI

LESSONS LEARNED
PROFILE FOR EXASCALE SOLUTIONS

Main changes
- Holistic view of data from IoT to Supercomputers.
- Hybrid in-house / cloud
- Workflow everywhere

Modularity is a must have. One does not fit all

Several accelerators, typically one per module

Performance comes from accelerators

The CPU has to be well balanced
- peak performance is not important
- Agility (FP64 for HPC, BF16 for deep learning) is crucial
- Data transfer is crucial

➔ Cover day to day needs and for all compute not fitting well in ACC

Keep overall architecture simple
➔ one CPU to unify all accelerators

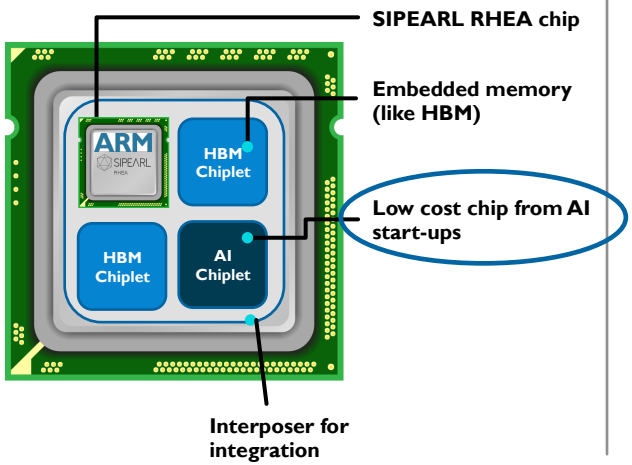
Keep end user life simple

- 1 CPU only
- LLVM + GCC + OPENMP 5.0
- Keep it open!

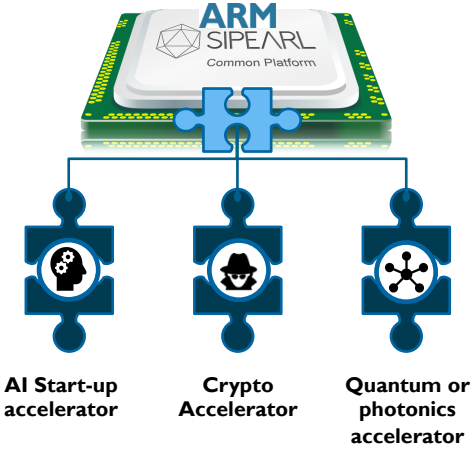
TECHNOLOGY & ROADMAP

COMMON PLATFORM VISION: FEDERATE ACCELERATORS

Interposer function



Value proposition

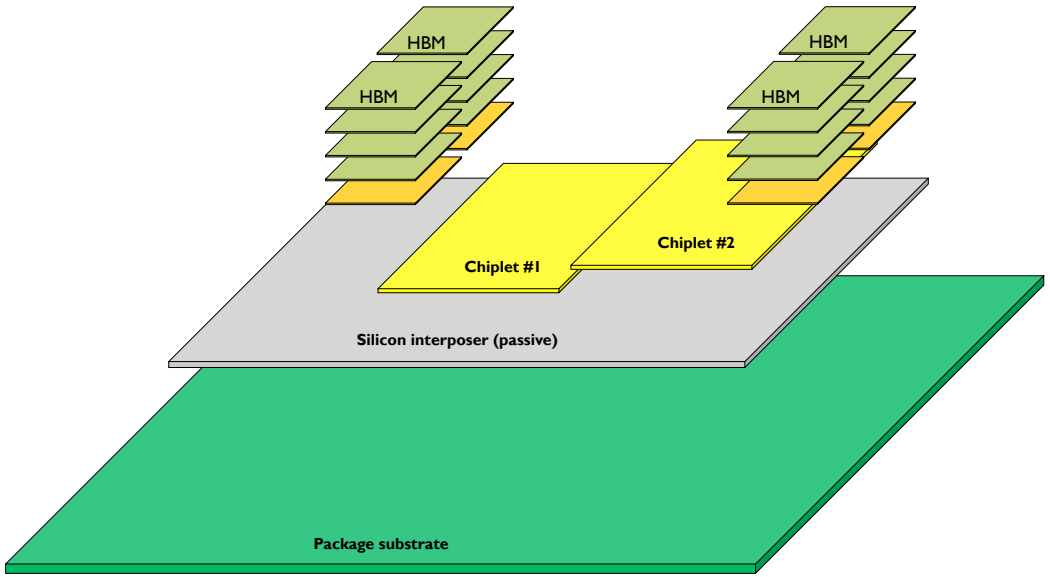


+60 potential ACC. (non exhaustive)

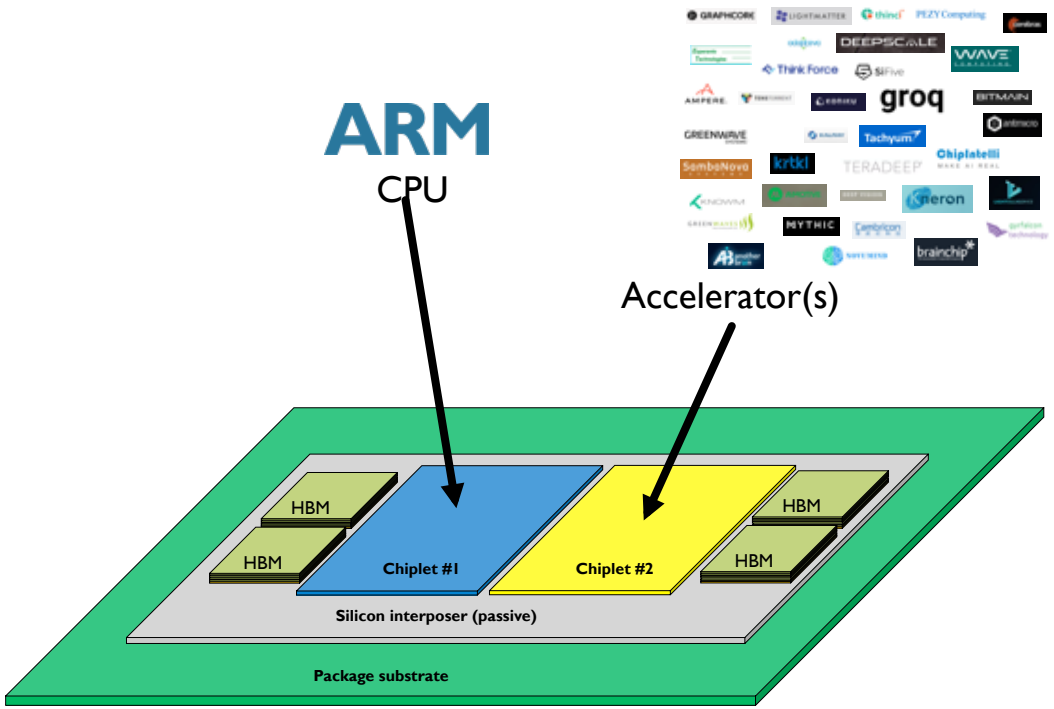
GRAPHCORE · LIGHTMATTER · thinci · PEZY Computing · cerebras · Esperanto Technologies · adapteva · DEEPSCALE · WAVE COMPUTING · Think Force · SiFive · AMPERE · TENSTORRENT · KONIKU · groq · BITMAIN · antmicro · GREENWAVE SYSTEMS · KALRAY · Tachyum · SambaNova SYSTEMS · krtkl · TERADEEP · ChipIntelli MAKE AI REAL · KNOWM · AIMOTIVE · DEEP VISION · Kneron · LIGHTELLIGENCE · GREENWAVES TECHNOLOGY · MYTHIC · Cambricon · gyrfalcon technology · another brain · NOVUMIND · brainchip

THE COMMON **OPEN** PLATFORM IS
THE EUROPEAN STANDARD FOR MANAGING EXTREME SPECIALIZATION

CONCEPT OF COMMON PLATFORM : INTERPOSER & MULTI-CHIPLET

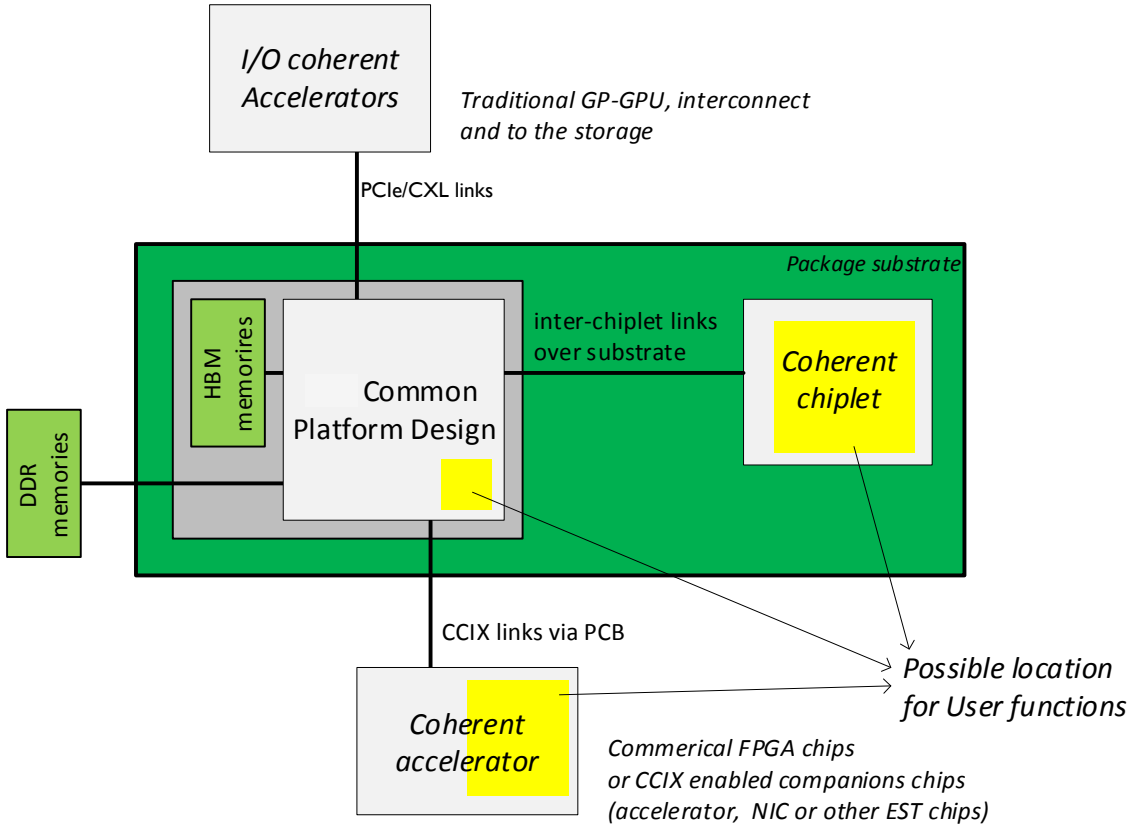


CONCEPT OF COMMON PLATFORM : INTERPOSER & MULTI-CHIPLET

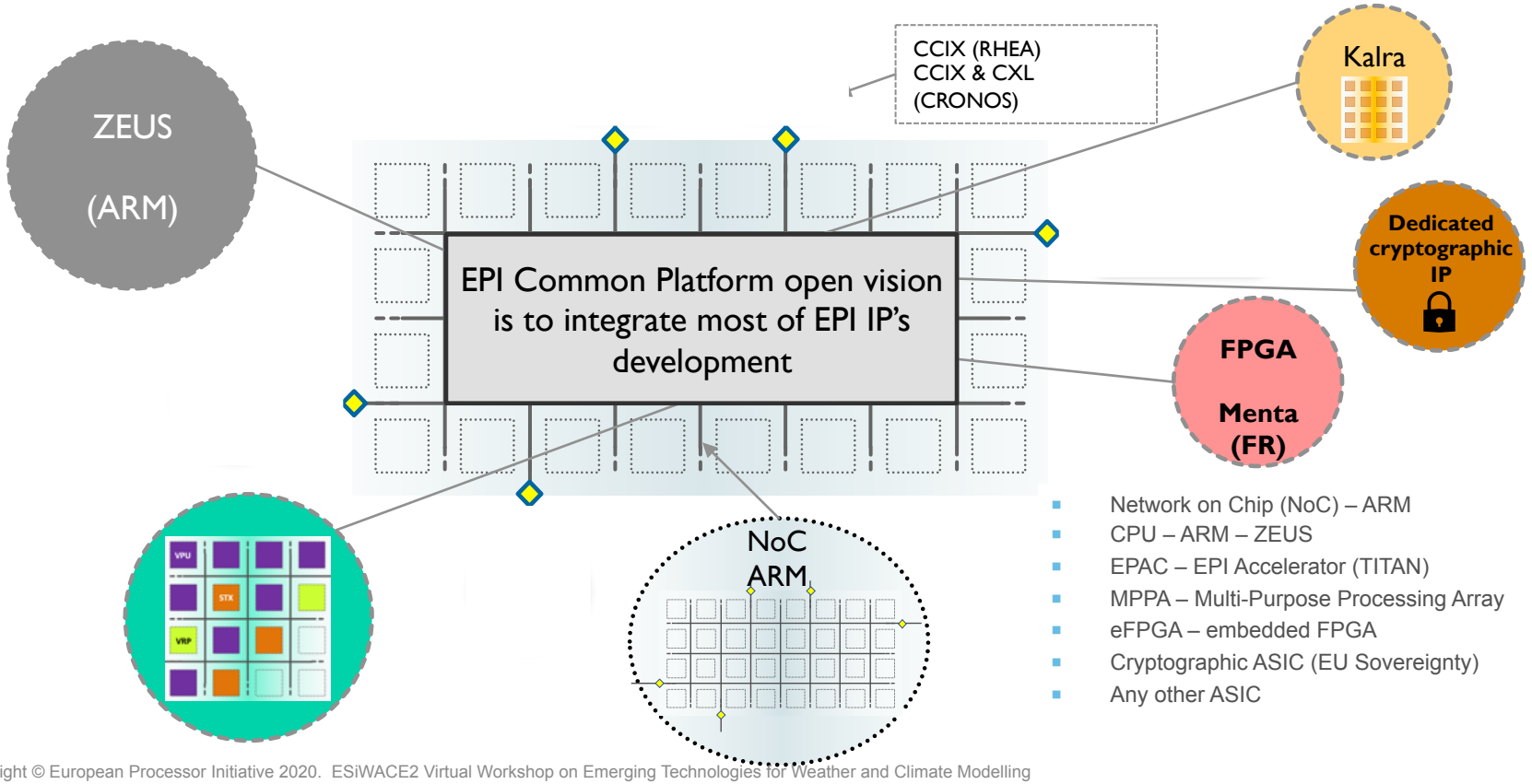


HETEROGENEOUS INTEGRATION

- Integrating customized functions at different levels
- EPI accelerator IPs today are integrated in Rhea design



GENERAL PURPOSE PROCESSOR (GPP) AND COMMON OPEN ARCHITECTURE

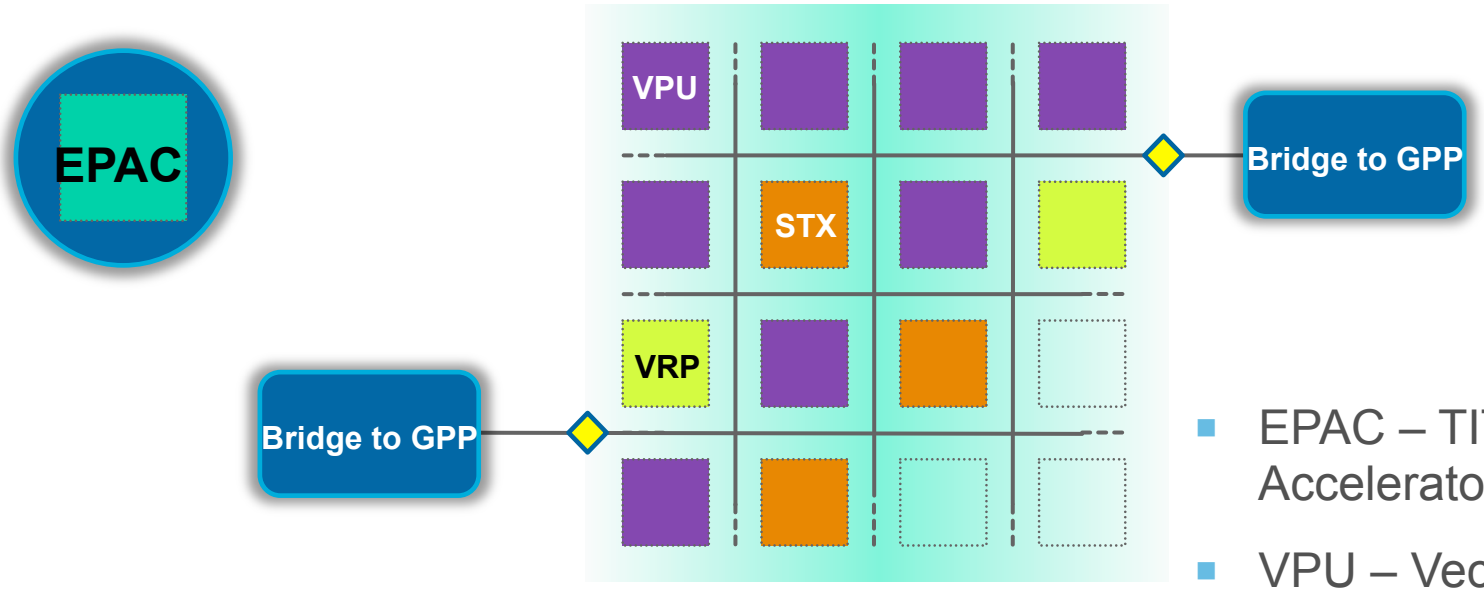


- Network on Chip (NoC) – ARM
- CPU – ARM – ZEUS
- EPAC – EPI Accelerator (TITAN)
- MPPA – Multi-Purpose Processing Array
- eFPGA – embedded FPGA
- Cryptographic ASIC (EU Sovereignty)
- Any other ASIC

CPU (RHEA) DESIGNS

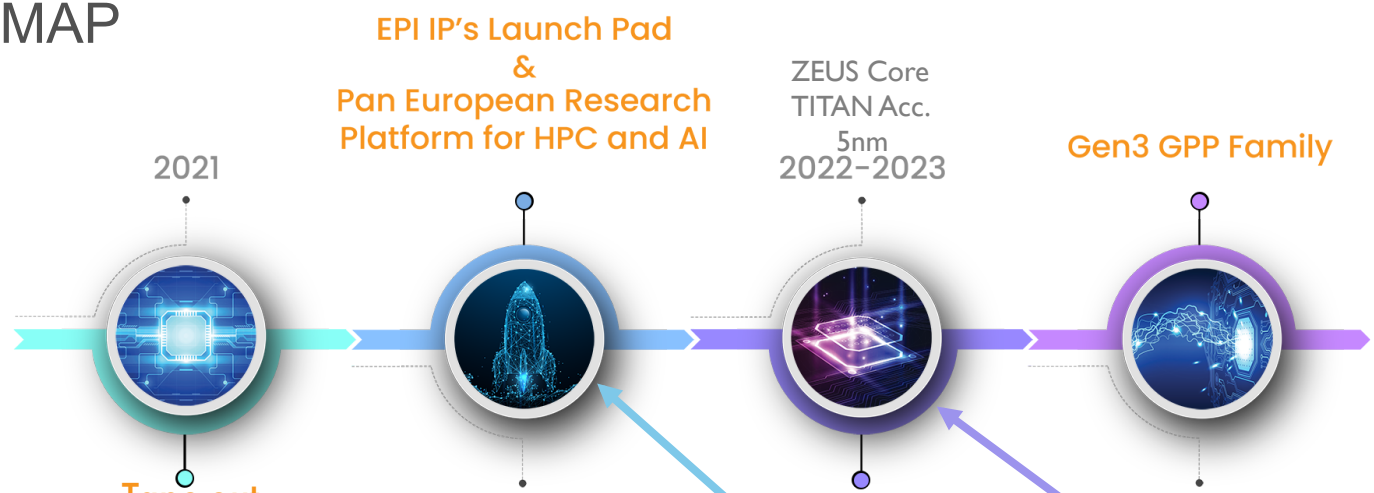
- Multi-core Armv8.x processor with vectorial extensions for both computing and control flow
- **Very high Byte/FLOP ratio**
- EPI Accelerators work in I/O coherent mode and share the same memory view
- Coherent NoC with system level cache to keep the data local
- HBM2e, DDR5 and PCIe gen5
- High-speed links for SMP and for functional extensions
- Low voltage to improve the energy efficiency / N6 process

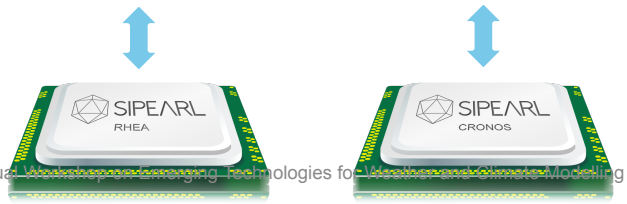
EPAC – RISC-V ACCELERATOR FOUNDATIONS



- EPAC – TITAN = EPI Accelerator
- VPU – Vector Processing Unit
- STX – Stencil/Tensor accelerator

ROADMAP





2022 – H2
EU Exascale Supercomputer
Edge-HPC (autonomous vehicle)
with CRONOS & TITAN

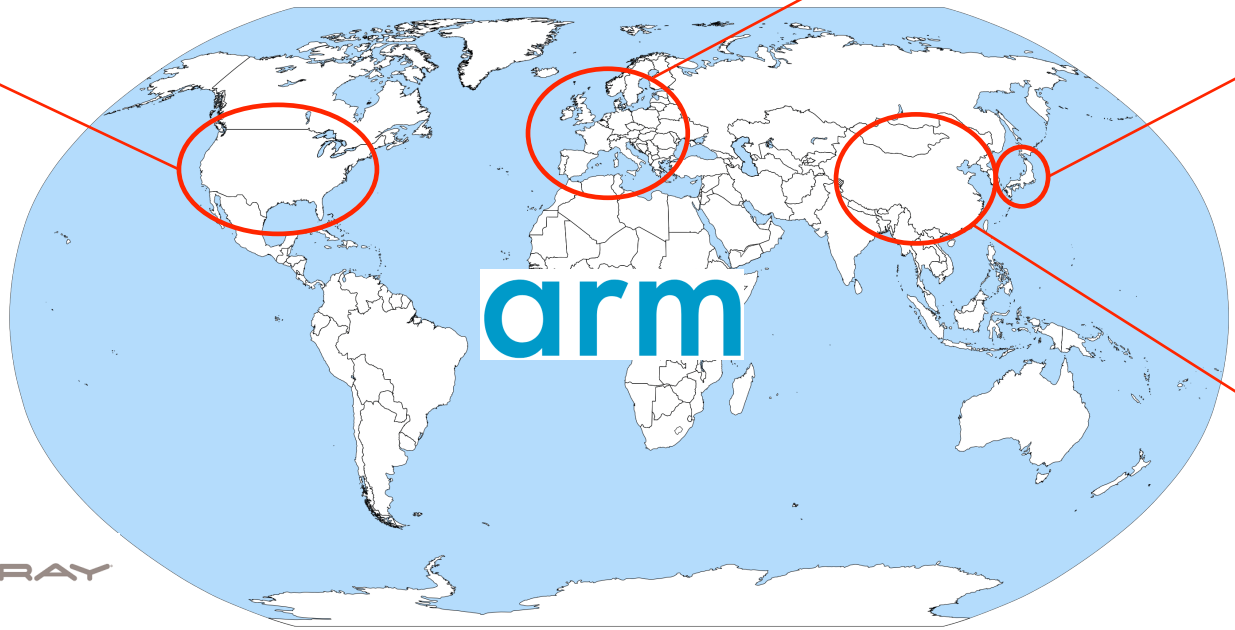
2021 – H2
E4 - PCIe board (WS compatible)
ATOS - BullxSequana Board
with RHEA β version

ARM WW HPC

Upcoming CPU manufacturer, Server manufacturer, system at CEA, UK, MontBlanc, EPI

CPU manufacturer, 2 server manufacturers, "Astra" system at Sandia, upcoming system at Stony Brook

CPU & server manufacturer, upcoming "Fugaku" system



CPU & server manufacturer [geopolitical issues regarding the license]

CONCLUSION

EUROPEAN APPROACH FOR (POST) EXASCALE CHALLENGES

Technology	Open	Ecosystem (holistic)
▪ One CPU to rule all accelerators ▪ ARM is the best choice: performances, openness, unique IoT to Supercomputer ecosystem ▪ Chiplet based approach ▪ Common Open Platform	▪ Common Open platform ▪ Open programming model ▪ Aim open hardware	▪ ARM from IoT to HPC ▪ GCC and LLVM ▪ OPENMP 5 ▪ OPENCL ▪ …

THANK YOU FOR YOUR ATTENTION

- www.european-processor-initiative.eu
- @EuProcessor
- European Processor Initiative
- European Processor Initiative

