SIPEARL

SiPearl Outlook

Teratec

Jean-Marc Denis
Chief Strategy Officer
jean-marc.denis@sipearl.com

Company Update

SIPEARL

SiPearl corporate overview

The European Server Processor Solution

HQ: Maisons-Laffitte (Paris), France

Incorporated in June 2019

CEO and Founder, Philippe Notton

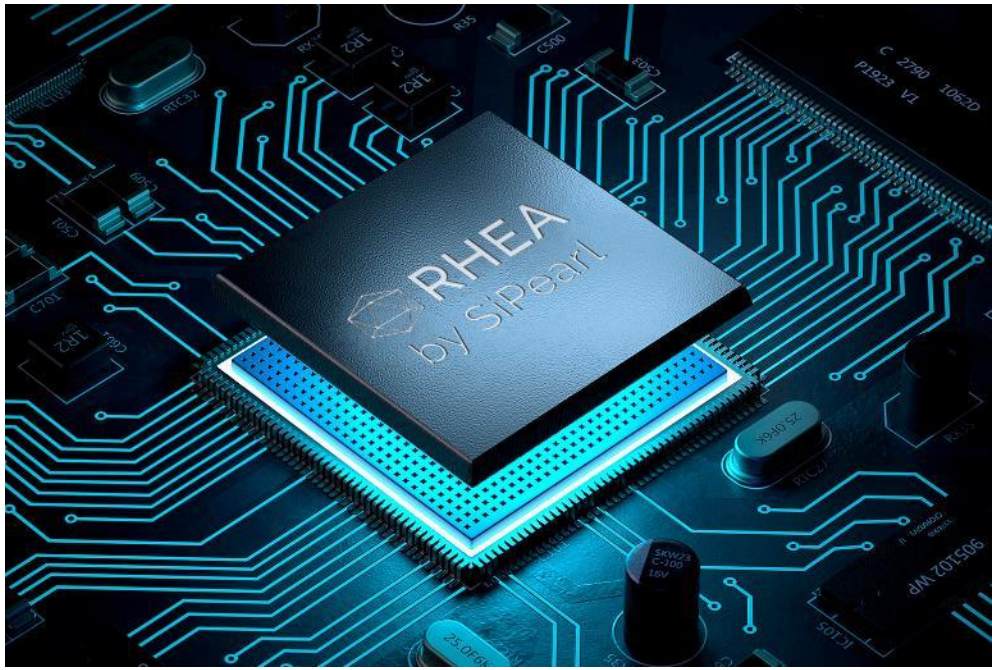
Design centers:

- France: Maisons-Laffitte, Massy Palaiseau, Sophia Antipolis, Grenoble
- Germany: Duisburg (Düsseldorf)
- Spain: Barcelona

Key Personnel from Intel, Atos, ST, Marvell, Nokia, Mstar-Mediatek

HPC Targeted Architecture based on Arm Neoverse V1 cores

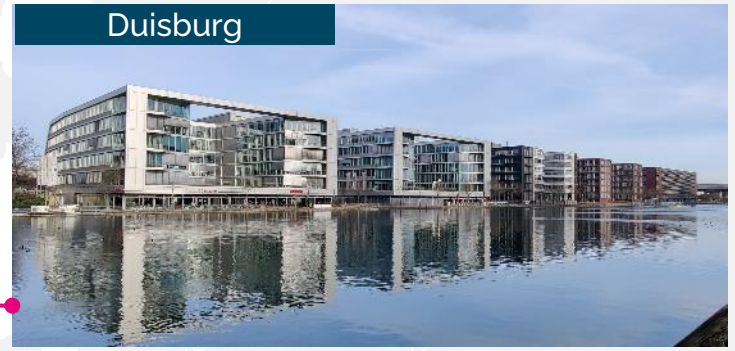
+100 employees today, targeting >1,000 in 2025



Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

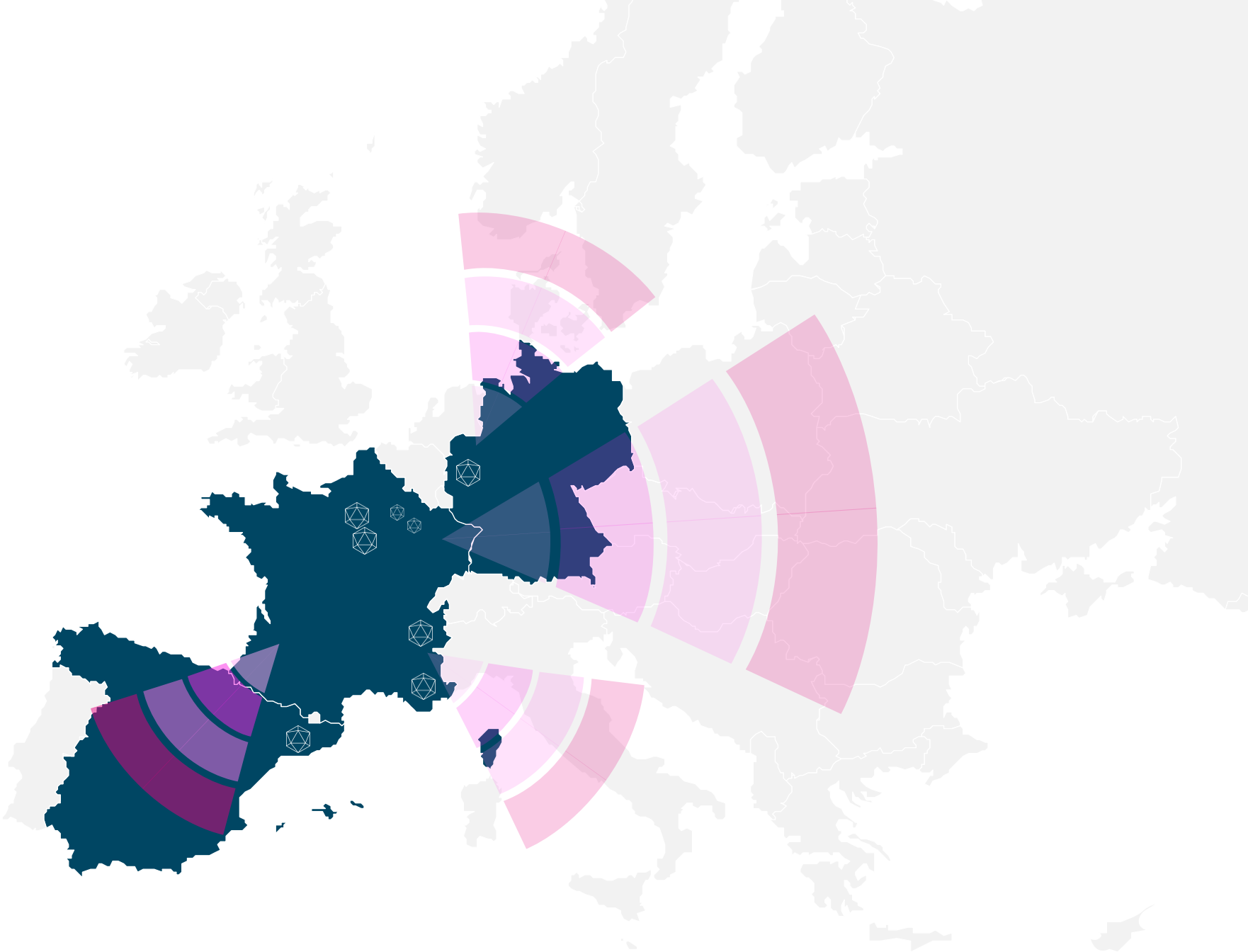
SiPearl offices

We are close to our partners and customers



Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

Sipearl extensions
2022 and 2023



Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

HPC is entering in its Cambrian era

SIPEARL

CPU

intel
AMD
FUJITSU
AMPERE
SIPEARL
NVIDIA
Alibaba Cloud
HUAWEI
amazon

ACC

NVIDIA
AMD
intel
GRAPHCORE
Google
cerebras
ALTERA
now part of Intel
SambaNova SYSTEMS
habana
PASQAL
XILINX
brainchip
IQM
D-Wave
The Quantum Computing Company™

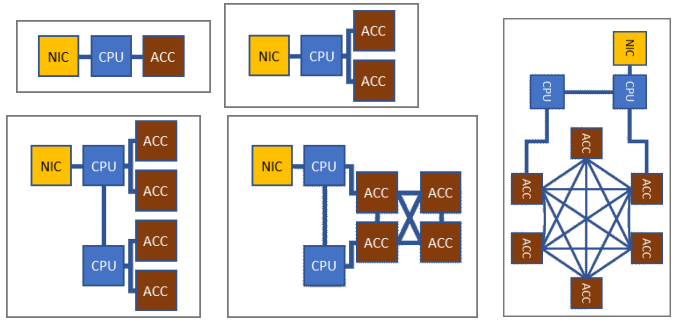
Intranode network

NVIDIA
AMD
intel
CXL Compute Express Link
PCI EXPRESS

HPC network

NVIDIA
Atos
CORNELIS NETWORKS
Hewlett Packard Enterprise

Intranode topology

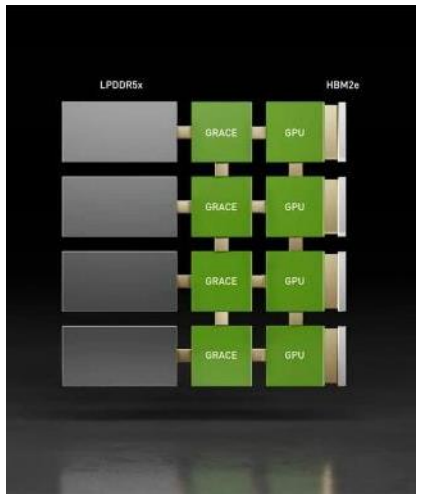


Cloud

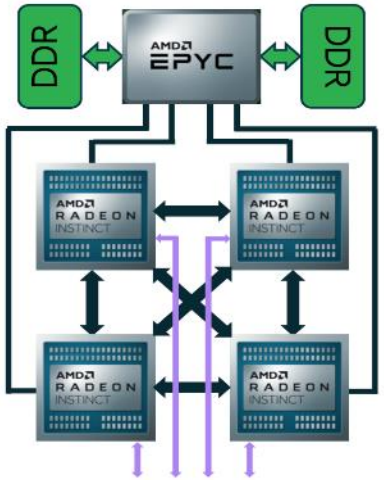
amazon
Microsoft Azure
Google
Lenovo
Hewlett Packard Enterprise
ORACLE
Alibaba Cloud
rackspace
the open cloud company
IBM Cloud

Copyright © SiPearl 2022 - Confidential

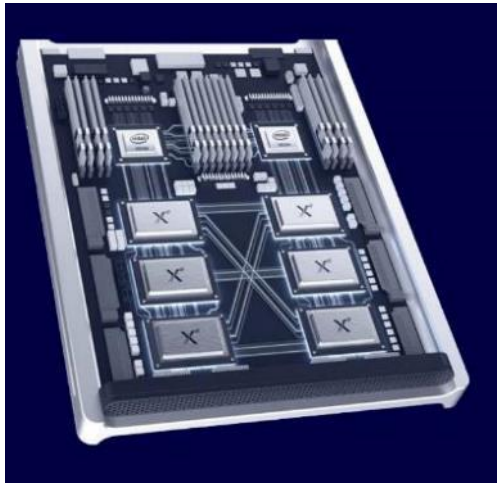
Many announcements - all “Super-SOC” based



NVIDIA



AMD



intel

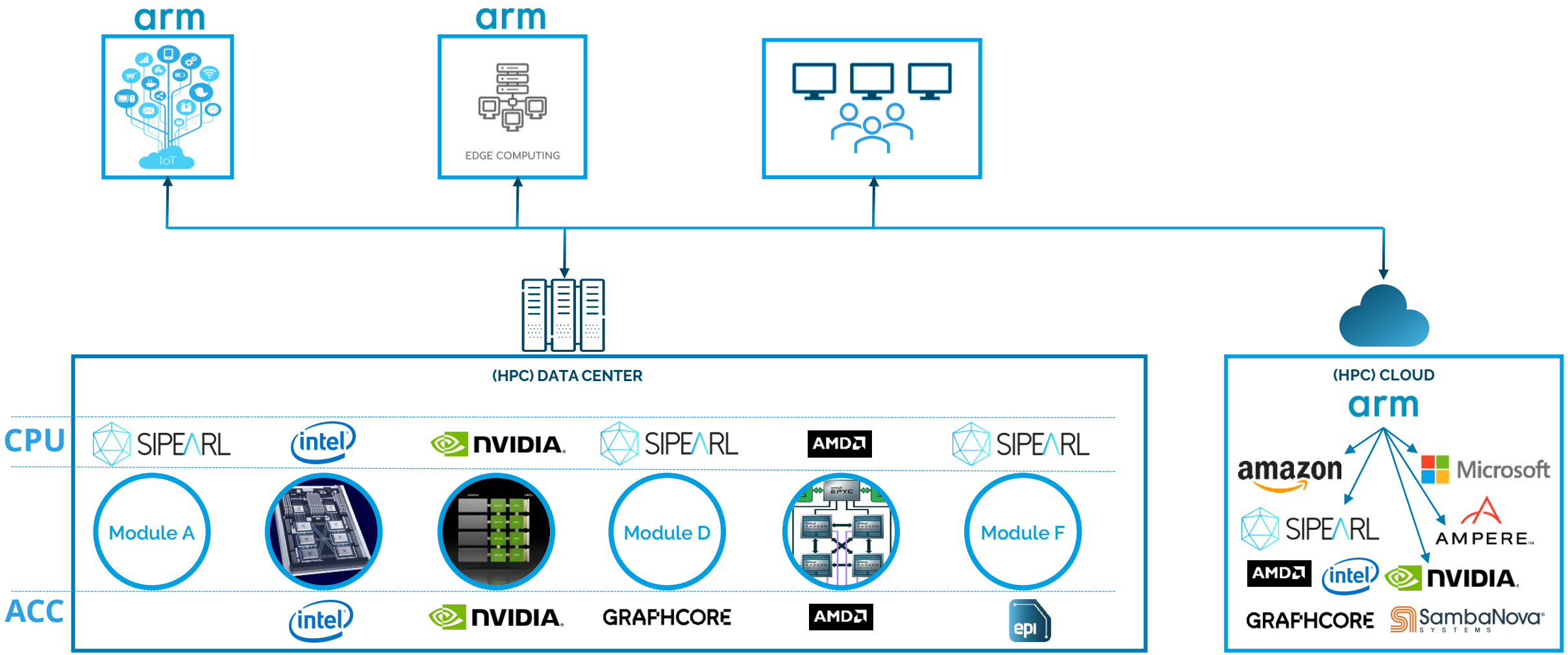
GRAPHCORE

cerebras

SambaNova SYSTEMS

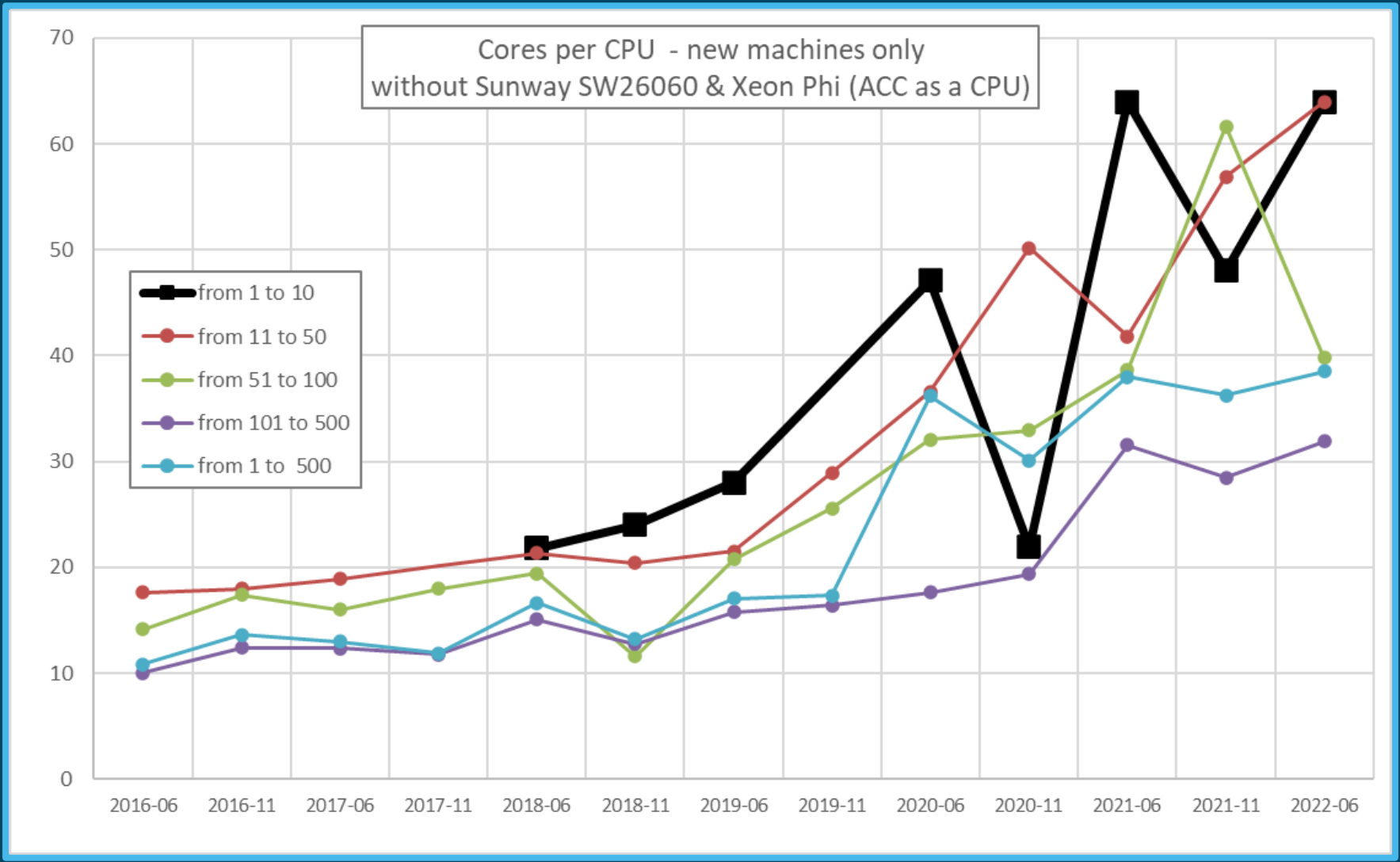
Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

HPC modular and hybrid architecture



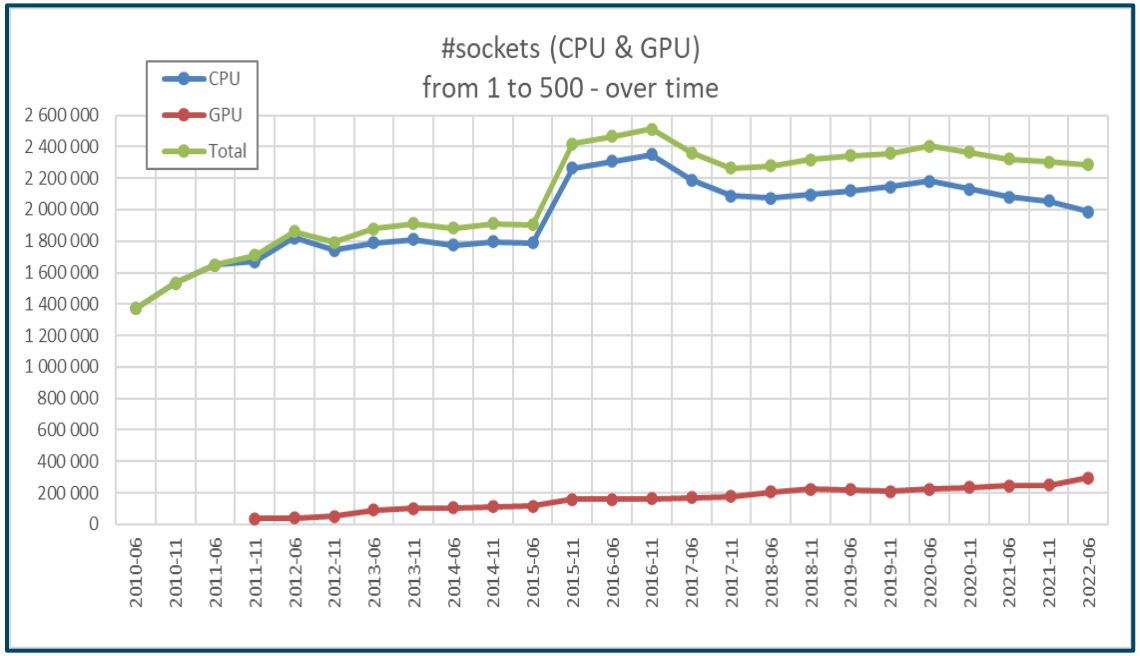
Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

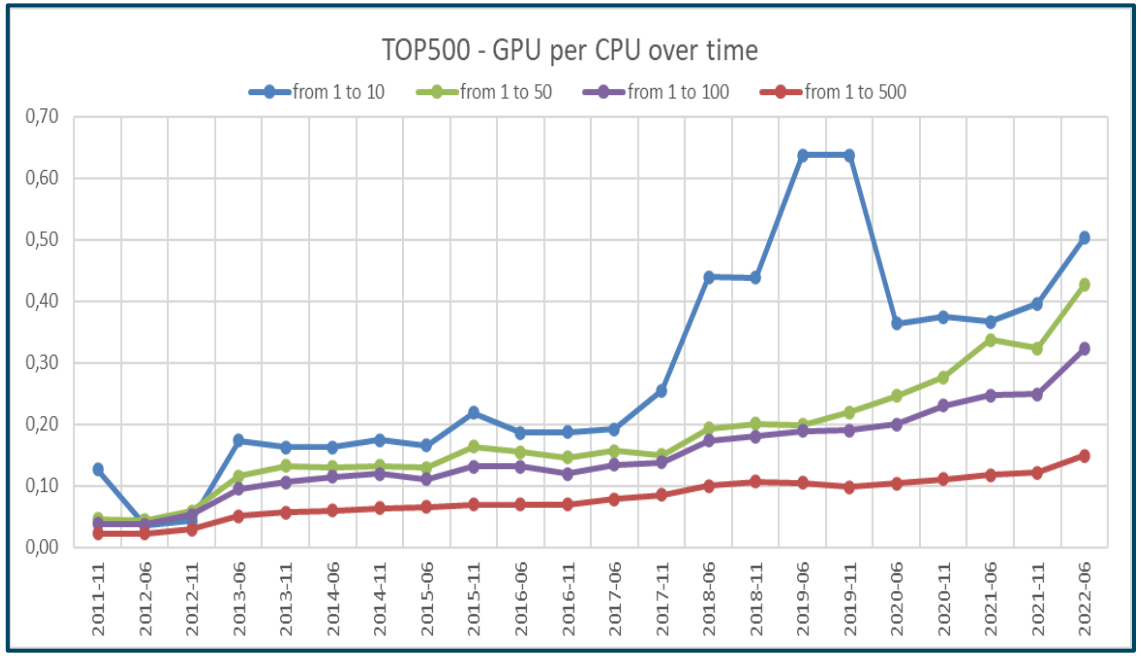
Core count per CPU socket over time *(from Top500)*



Copyright © SiPearl 2022 - Confidential

Supercomputers (top500) – how many GPU per CPU? The historical perspective





In top10: bias because of Fugaku and Tianhe-2 machines. GPU part is artificially low. Next release (Nov.2022) should show >0.5GPU per CPU. Very variable ratio over time because Top10 is where new, innovative and sometimes non conventional technologies are introduced in first place.

In top50: growing very fast (x2 during the last 3Y)

In top100: growing fast

Over the complete list: 1) CPU still massively dominant. But GPU part is growing. Inertia is dominant. GPU cannot do everything everywhere.
2) even with Fugaku bias, total #socket is decreasing Reason: more powerful CPUs & GPUs

Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

Cambrian era – Consequences

Must have

Open/Standard HW interfaces (UCIe, CXL, PCIe)

Must Have

Open/Standard SW interfaces (UCIe, CXL, PCIe)

Must have

On-permise & in the Cloud

CPUs still matter. GPUs and ACCs don't fit all needs

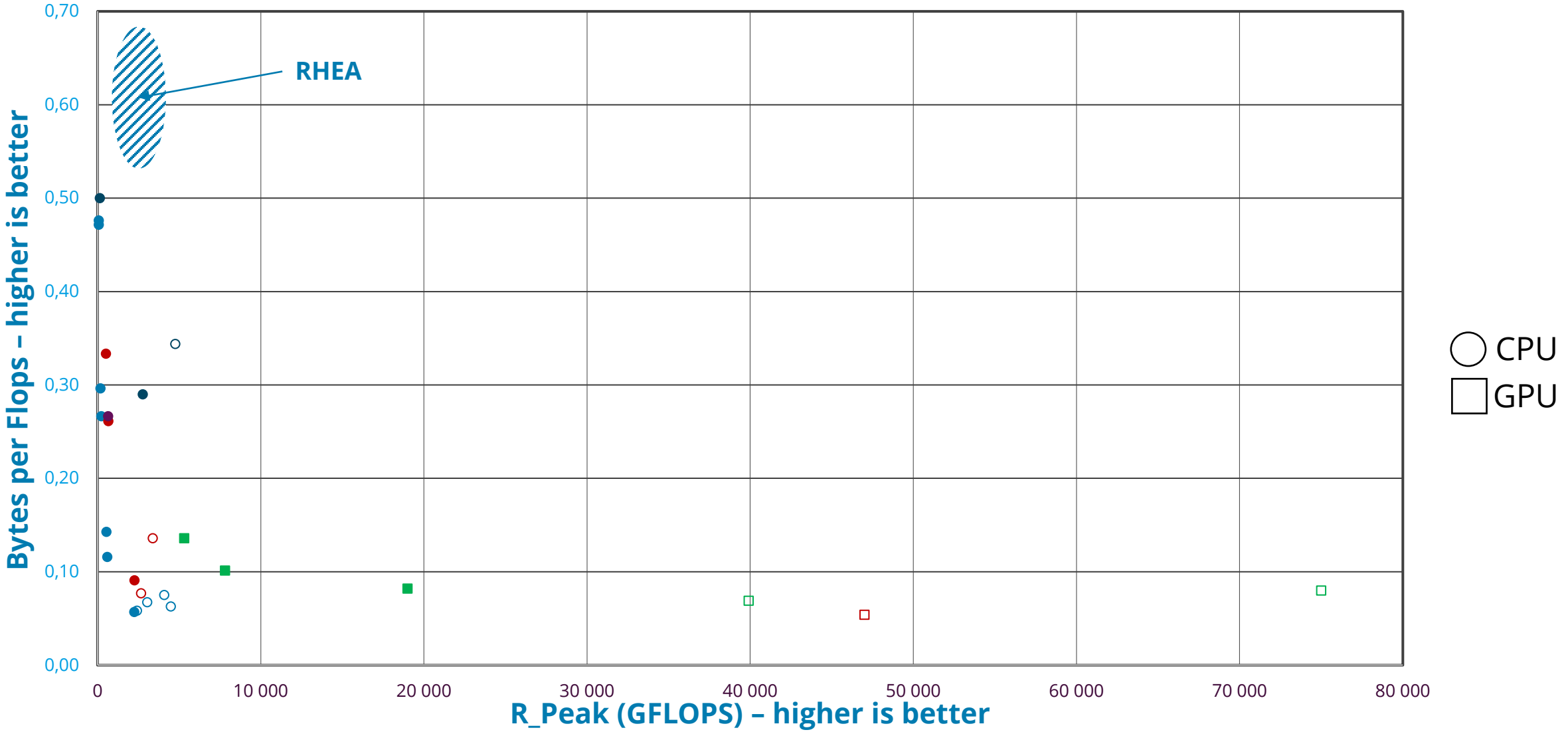
CPU Core count is growing (fast).

What's the trade-off?

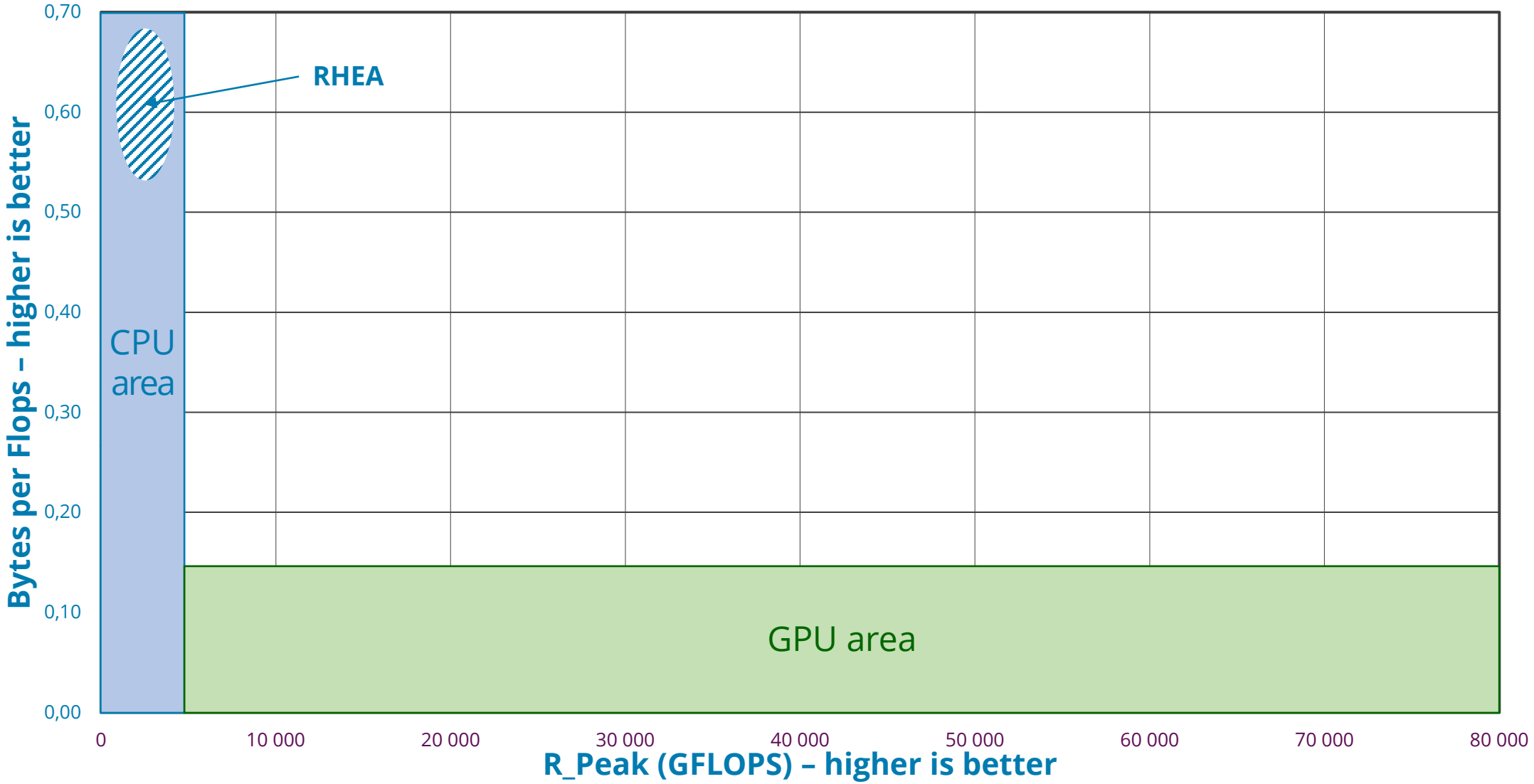
Copyright © SiPearl 2022 - Confidential

Performance?

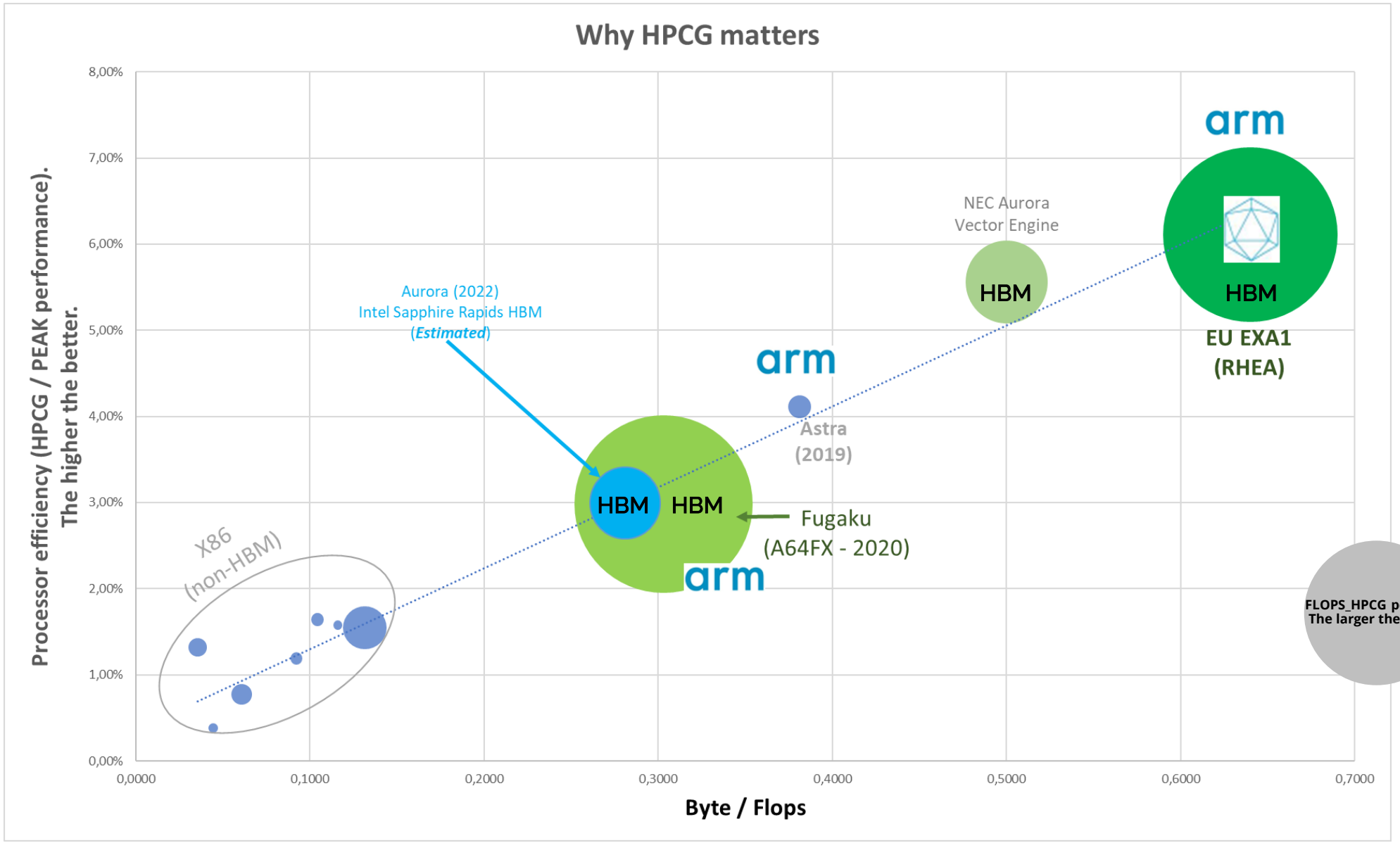
SIPEARL



Copyright © SiPearl 2022 - Forum Teratec, 14-15 June 2022

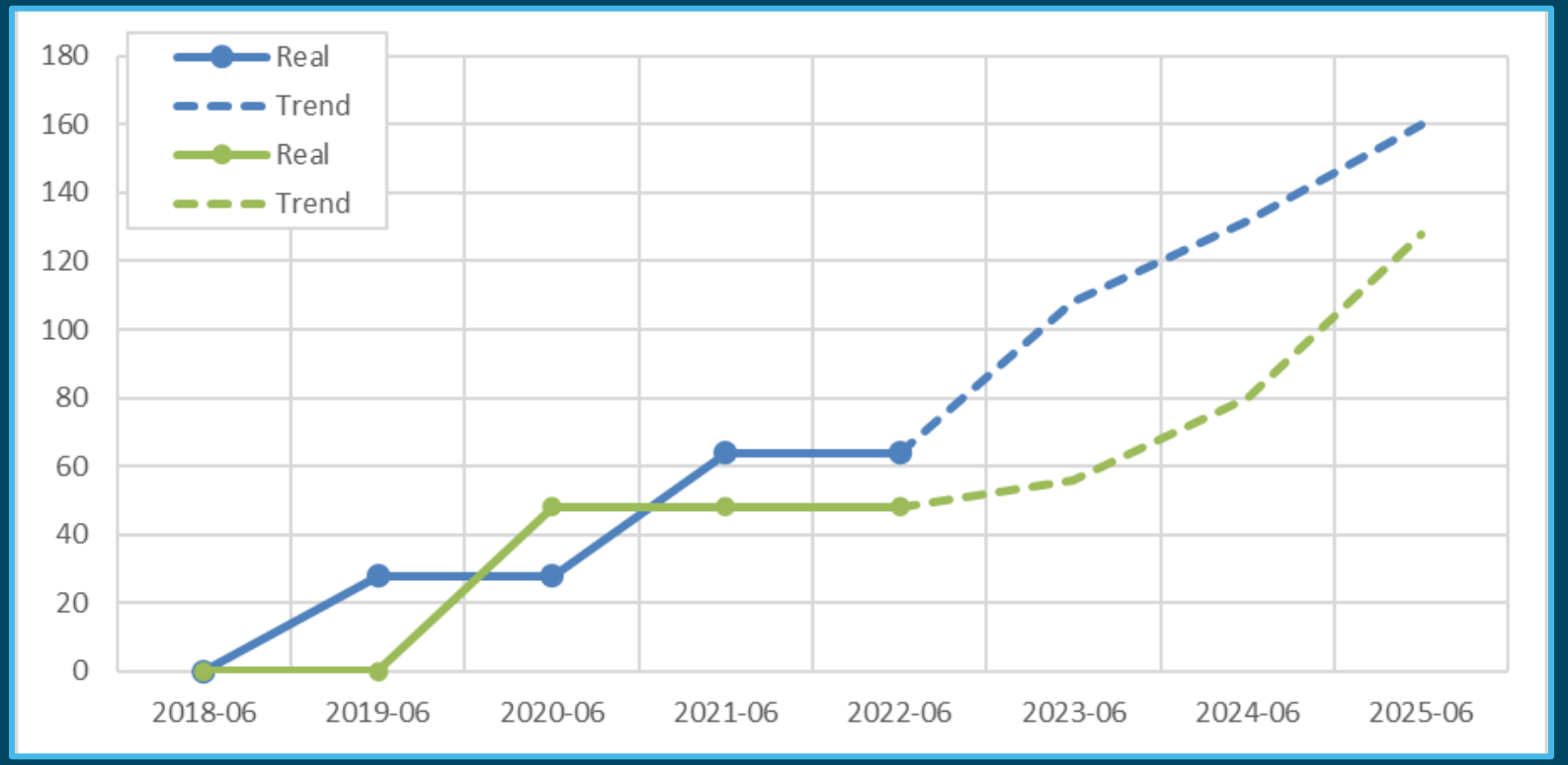


Copyright © SiPearl 2022 - Forum Teratec, 14-15 June 2022



Copyright © SiPearl 2022 - Forum Teratec, 14-15 June 2022

Core count per CPU socket for <u>new</u> machines in top10 over time



More cores in HBM-less than in HBM processors ➔ (much) **reduced memory BW per core in HBM-less processors**

Overtime, (2025 and beyond), HBM processors should close the gap in core count with HBM-less processors while preserving memory BW per core

Copyright © SiPearl 2022 - Confidential

Performances – Consequences

arm

HBM

CPU bound → GPUs

MEM bound → CPUs

Bytes/Flops

Max mem BW → **HBM**

Slow memory only when required

→ **DDR**

Power efficiency (perf on **real** apps per watt) really matters

→ **HPCG**

Need new specific memory and energy profiling tools

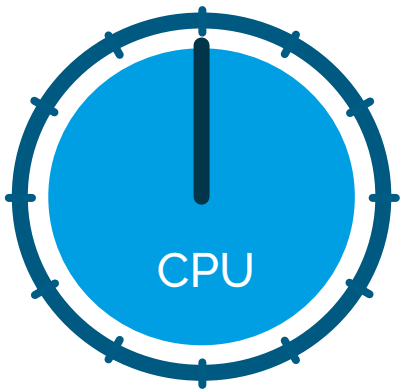
Copyright © SiPearl 2022 - Confidential

The developer/end-user perspective

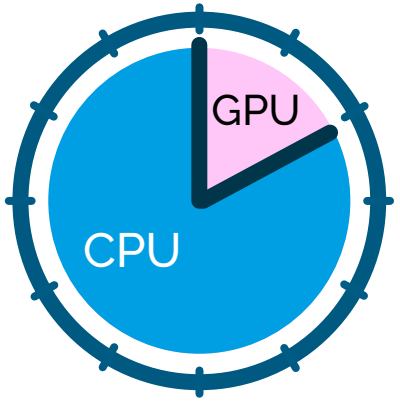
SIPEARL

Processing elements and Development time

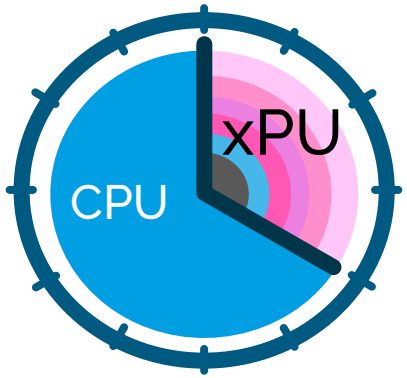
Processing elements in a system



Back in the day

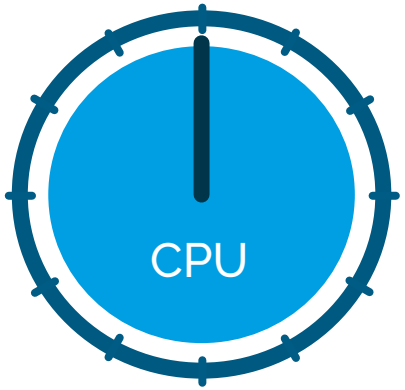


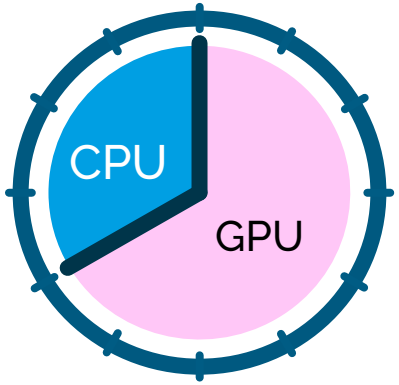
Now

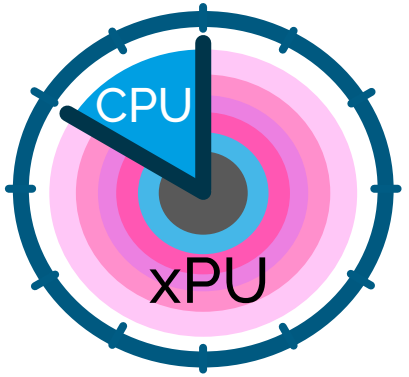


Next gen (2024+)

Development time







Developers will spend the bulk of their time programming accelerators.

CPUs that are efficient « out of the box » offer an advantage

Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

Software Ecosystem

Only SiPearl covers the complete software ecosystem

Copyright © SiPearl 2022 – Forum Teratec, 14-15 June 2022

Developer & end-user perspective – Consequences

xPU: (toward) a unified programming model

CPU: No time spent on optimization

Must Have
Global and unified memory

Must Have
Open/standard SW stacks

Develop once, run on many

Copyright © SiPearl 2022 - Confidential

Conclusion
Rhea: a Processor for the Exa era

SIPEARL

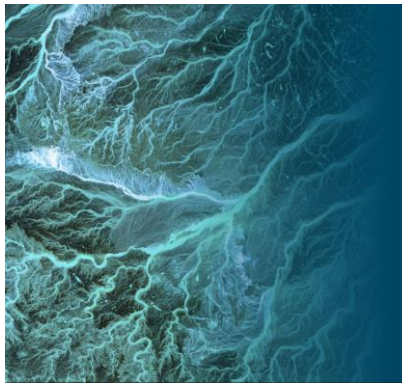
Must have
Open/Standard HW interfaces (UCIe, CXL, PCIe)

Must Have
Open/Standard SW interfaces (UCIe, CXL, PCIe)

Must have
On-permise & in the Cloud

CPUs still matter. GPUs and ACCs don't fit all needs

CPU Core count is growing (fast).
What's the trade-off?



arm

xPU: (toward) a unified programming model

CPU: No time spent on optimization

Must Have
Global and unified memory

Must Have
Open/standard SW stacks

arm
HBM

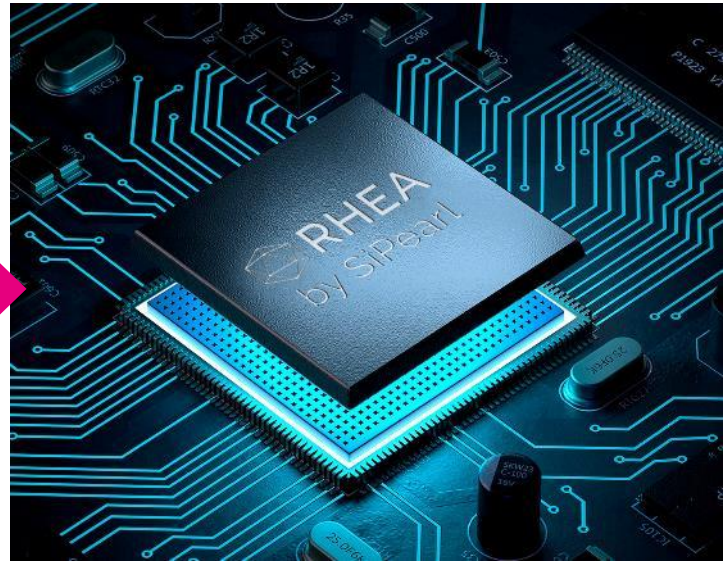
CPU bound → GPUs
MEM bound → CPUs

Bytes/Flops
Max mem BW → **HBM**

Slow memory only when required
→ **DDR**

Power efficiency (perf on <u>real</u> apps per watt) really matters
→ HPCG

Need new specific memory and energy profiling tools

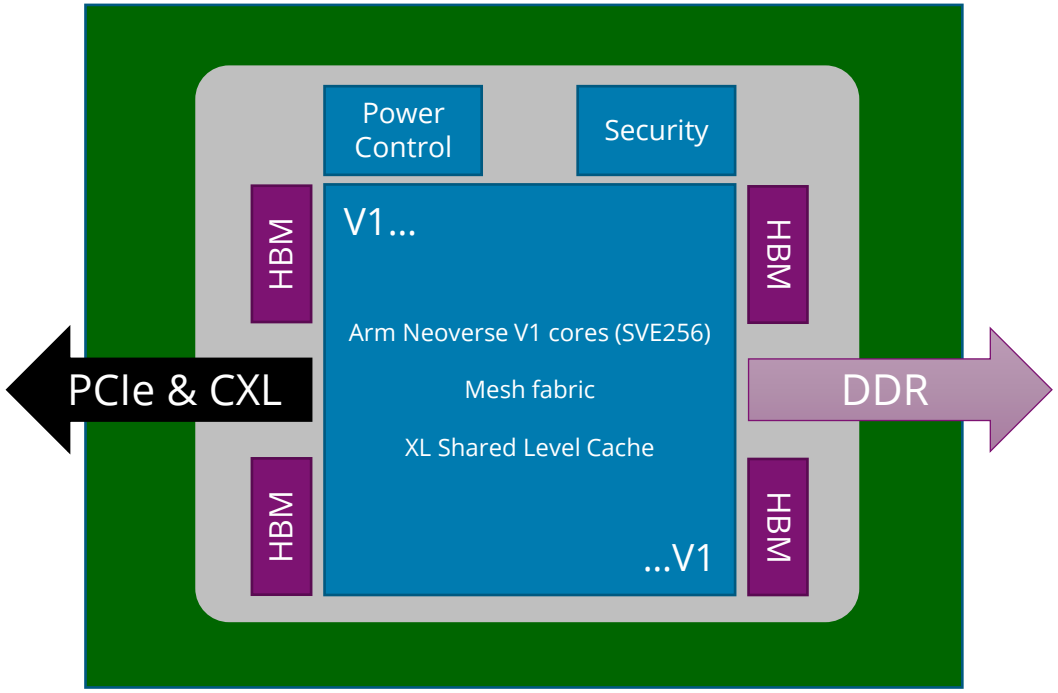


At the heart of Rhea

With its high-performance, low-power Arm Neoverse V1 architecture, Rhea will meet the needs of all supercomputing workloads.

Key features

Core	- Arm architecture - Neoverse V1 cores - SVE 256 per core supporting 64/32/BF16 and Int8 - ArmVirtualization extensions
SoC	- Arm mesh fabric - Advanced RAS support including Arm RAS extensions - Link protection for NoC & high-speed IO - ECC support for selected memory
Cache	- Large L3 (Shared Level Cache) - RAS supported for all cache levels
Memory	- HBM2e - And DDR5 - ECC for memory and link protection for controllers
High Speed I/O	- PCIe, CCIX & CXL - Root and endpoint support
Other I/O	- USB, GPIO, SPI, I^2C
Power Management	- Power management block to optimize perf/watt accross use cases and workloads.
Security Block Support	- Secure boot and secure upgrade - Crypto - True random number generation - Made in Europe



Rhea will deliver extraordinary real compute performance and efficiency with an unmatched Bytes/Flops ratio.

Copyright © SiPearl 2022 - Forum Teratec, 14-15 June 2022

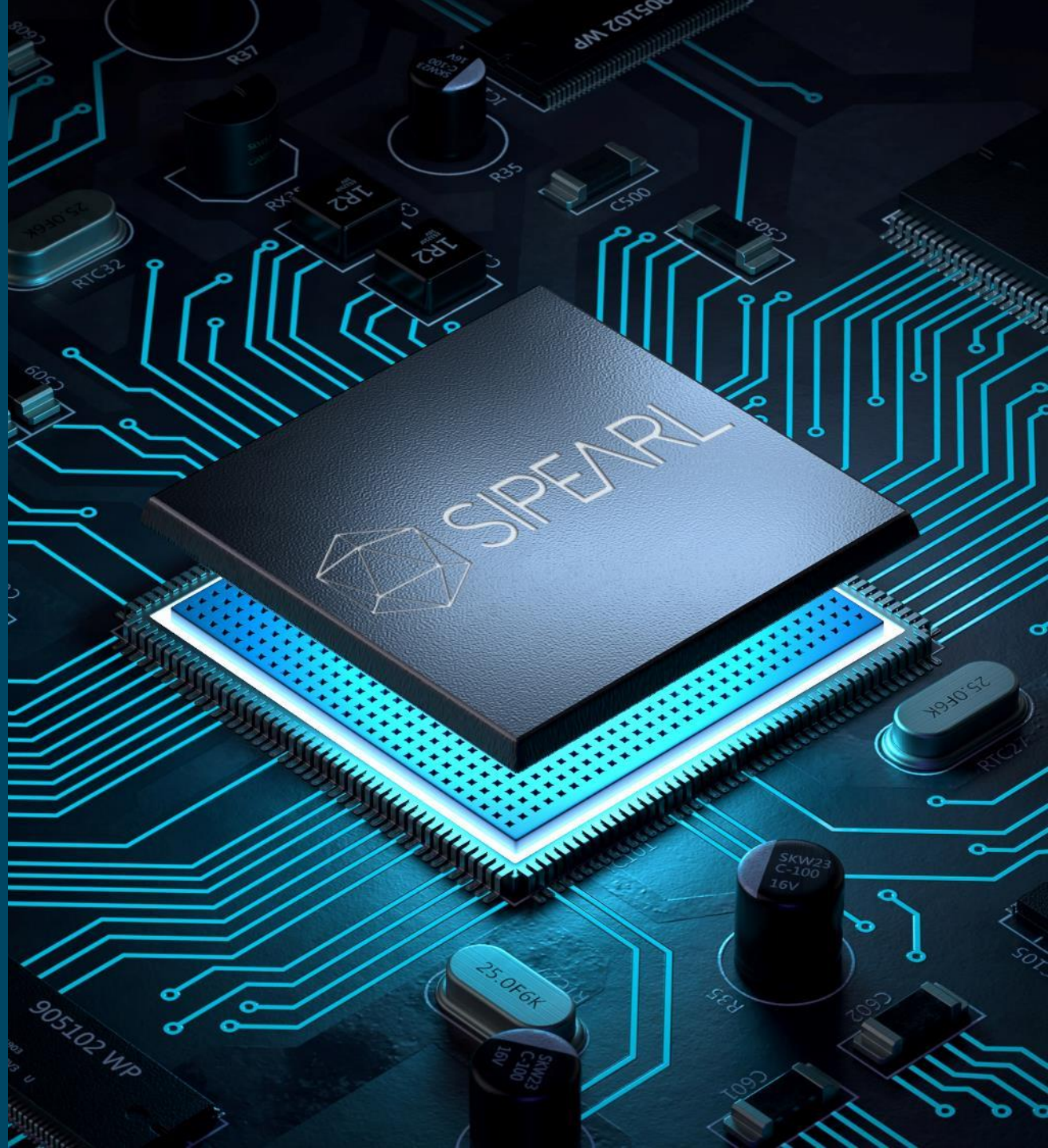
About SiPearl

Created by Philippe Notton, SiPearl is designing the high-performance, low-power microprocessor for European exascale supercomputers. This new generation of microprocessors will enable Europe to set out its technological sovereignty in strategic high performance computing markets such as artificial intelligence, medical research or climate modelling.

SiPearl is working in close collaboration with its 27 partners from the European Processor Initiative (EPI) consortium - leading names from the scientific community, supercomputing centres and industry - which are its stakeholders, future clients and end-users.

SiPearl employs 109* people in France, Germany and Spain. Its first range of microprocessors, Rhea, will be launched at the end of the year.

The company is supported by the European Union (funding from the European Union's Horizon 2020 research and innovation program under specific grant agreement no.826647).

* as of June 15th 2022

Contact
Jean-Marc Denis
Chief Strategy Officer
jean-marc.denis@sipearl.com