\Orchestrating a brighter world NEC

2nd Aurora Deep Dive Workshop at RWTH Aachen University
November 28, 2019

SX-Aurora TSUBASA

Passion for Sustained Performance

Shintaro MOMOSE, Ph.D.
shintaro.momose@emea.nec.com
NEC Deutschland GmbH

1. Aurora Over View
2. Vector Engine
3. VE Partitioning Mode
4. VE10E, VE20
5. Performance
6. High Density Product, A412-8
7. Aurora3

Development Philosophy

SX-Aurora TSUBASA

POINT 1 **Memory Bandwidth**

1.22TB/s / processor, 150GB/s / core

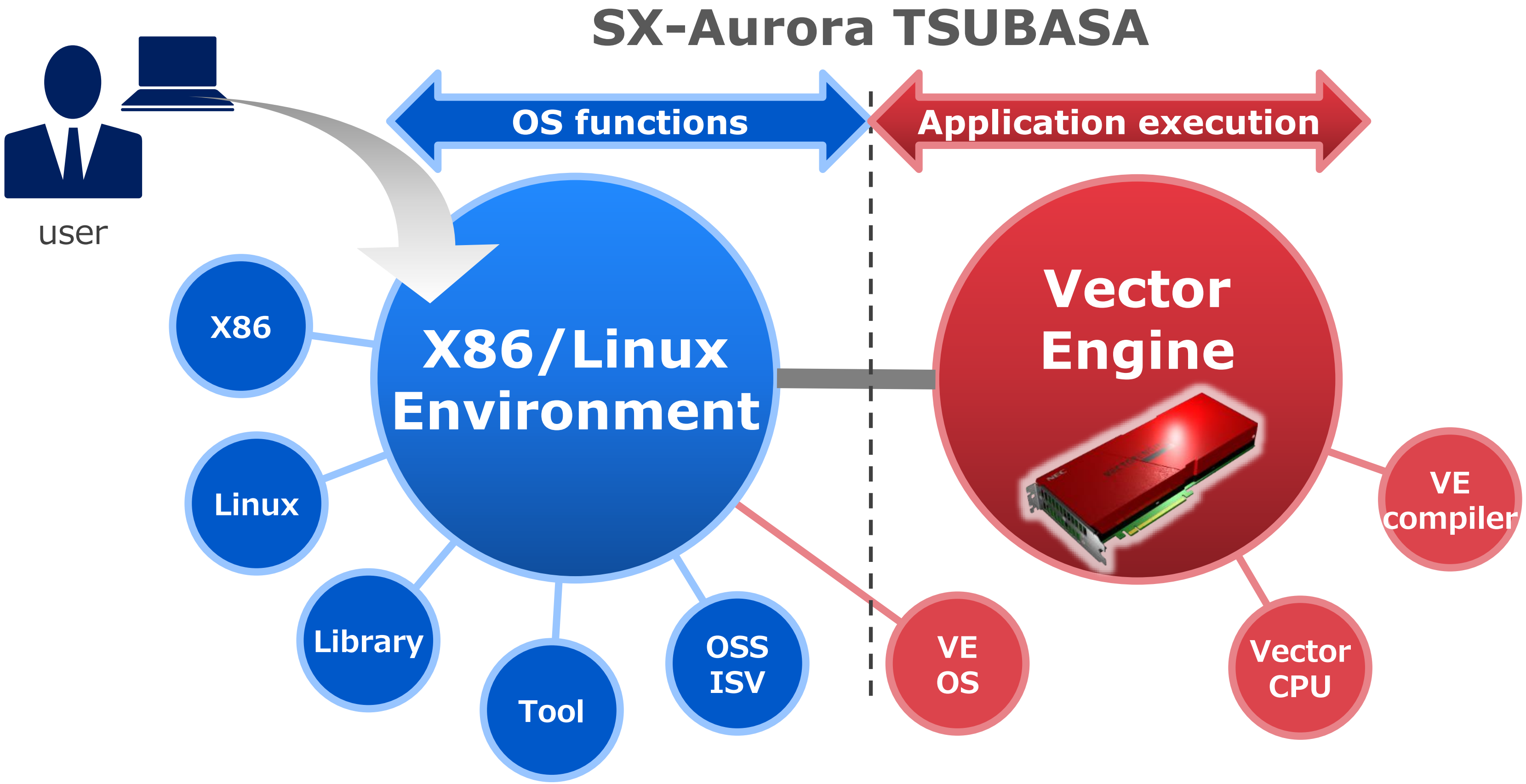
POINT 2 **Easy to Use**

Fortran/C/C++ programing, OpenMP
Automatic vectorization/parallelization

POINT 3 **x86/Linux**

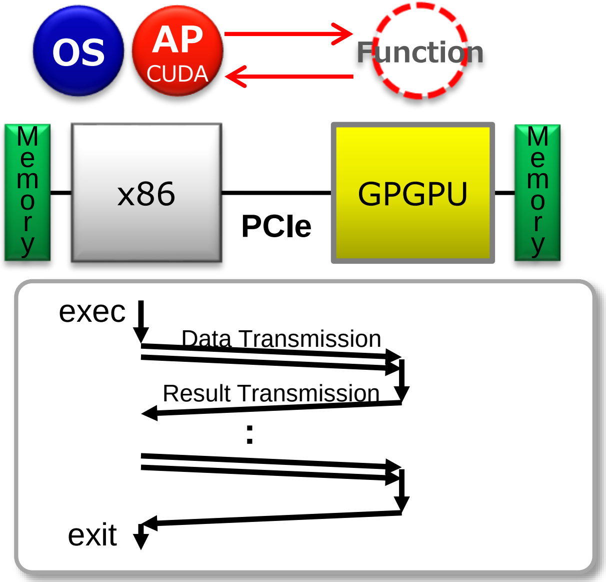
High sustained performance on
x86/Linux environment

Architecture



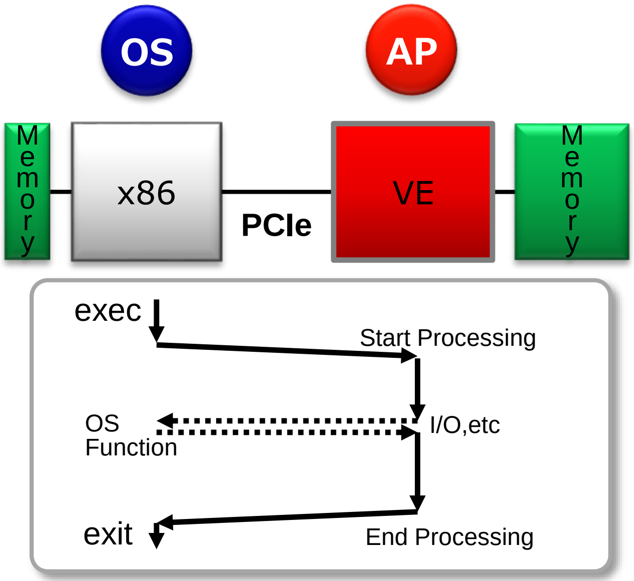
What is Different from GPGPU?

GPGPU Architecture



Frequent PCIe transmission

Aurora Architecture



Whole AP is executed on VE

disadvantage

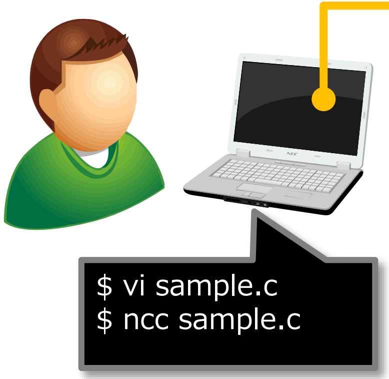
- PCIe bottleneck
- Small memory
- Programming difficulty

Advantage

- Avoiding PCIe bottleneck
- Larger memory
- Standard language

Usability

Programing Environment



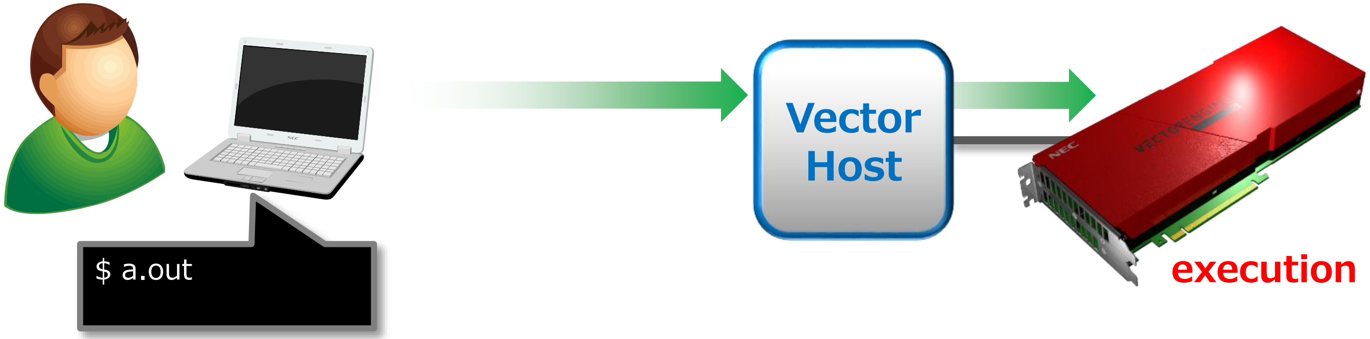
```
$ vi sample.c
$ ncc sample.c
```

Vector Cross Compiler

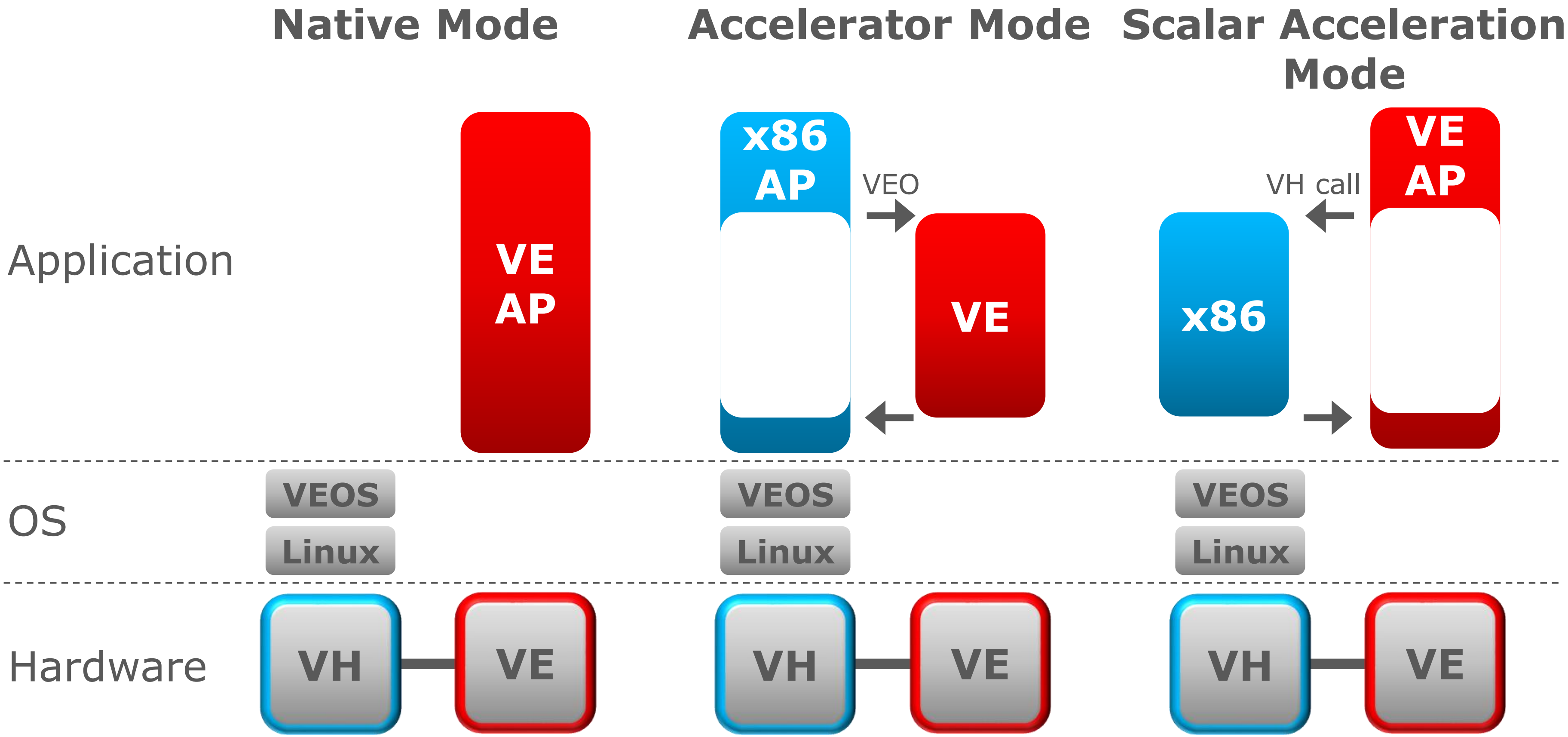
- automatic vectorization
- automatic parallelization

Fortran:	F2003, F2008
C/C++:	C11/C++14
OpenMP:	OpenMP4.5
Library:	MPI3.1, libc, BLAS, Lapack, etc
Debugger:	gdb, Eclipse parallel tools platform
Tools:	PROGINF, Ftrace Viewer

Execution Environment



```
$ a.out
```



Native Mode

Accelerator Mode

Scalar Acceleration Mode

Application

OS

Hardware

VE AP

VEOS

Linux

VH

VE

x86 AP

VEO

VE

VEOS

Linux

VH

VE

VE AP

VH call

x86

VEOS

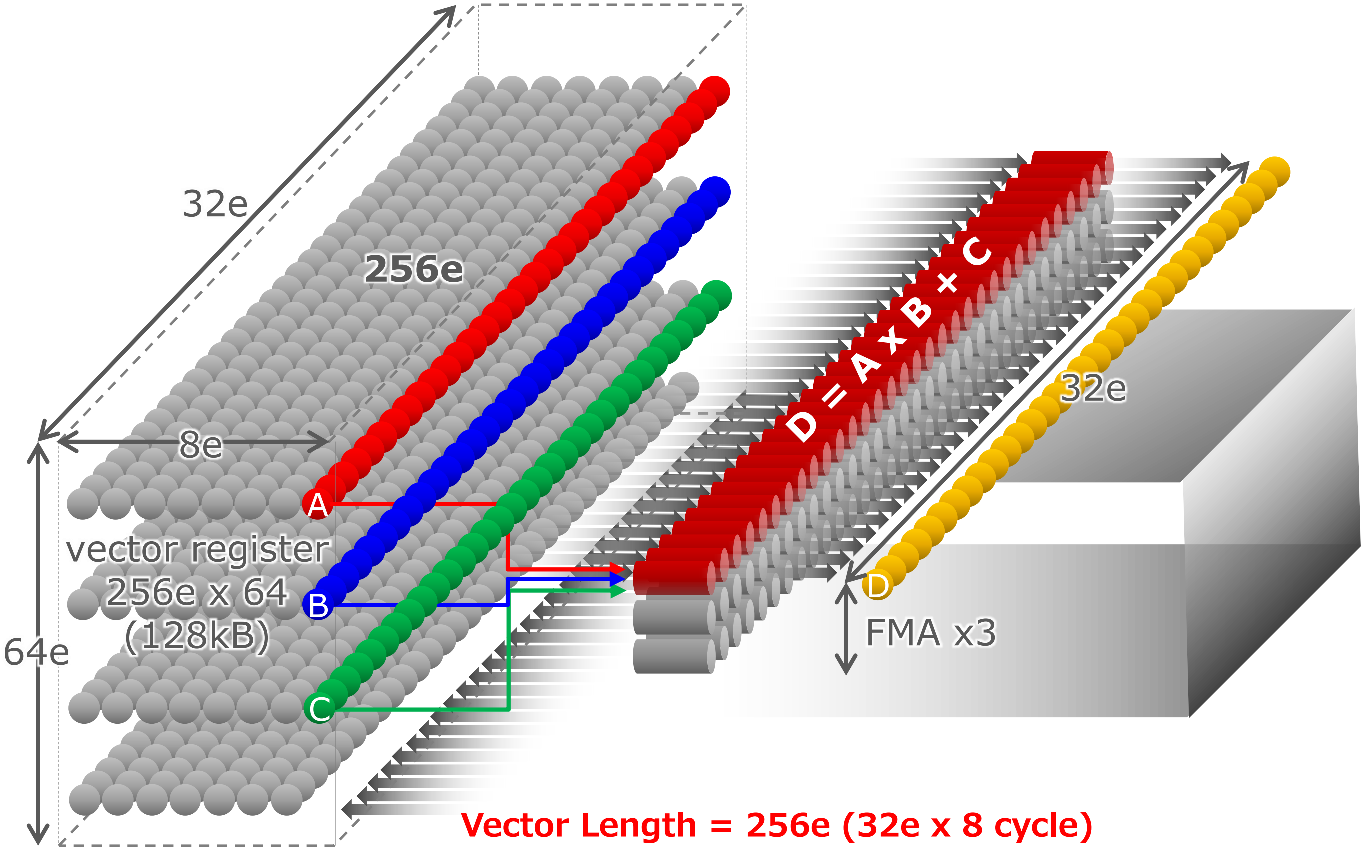
Linux

VH

VE

Vector Engine

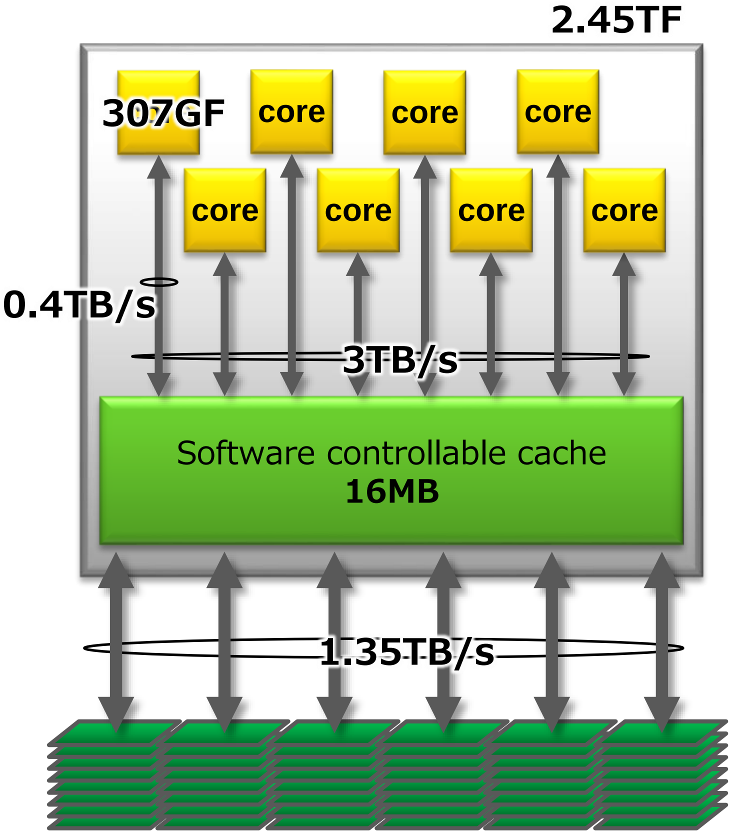
Vector Execution



Vector Length = 256e (32e x 8 cycle)
307.2GF = 32Flops/cycle x 2(FMA) x 3 x 1.6GHz

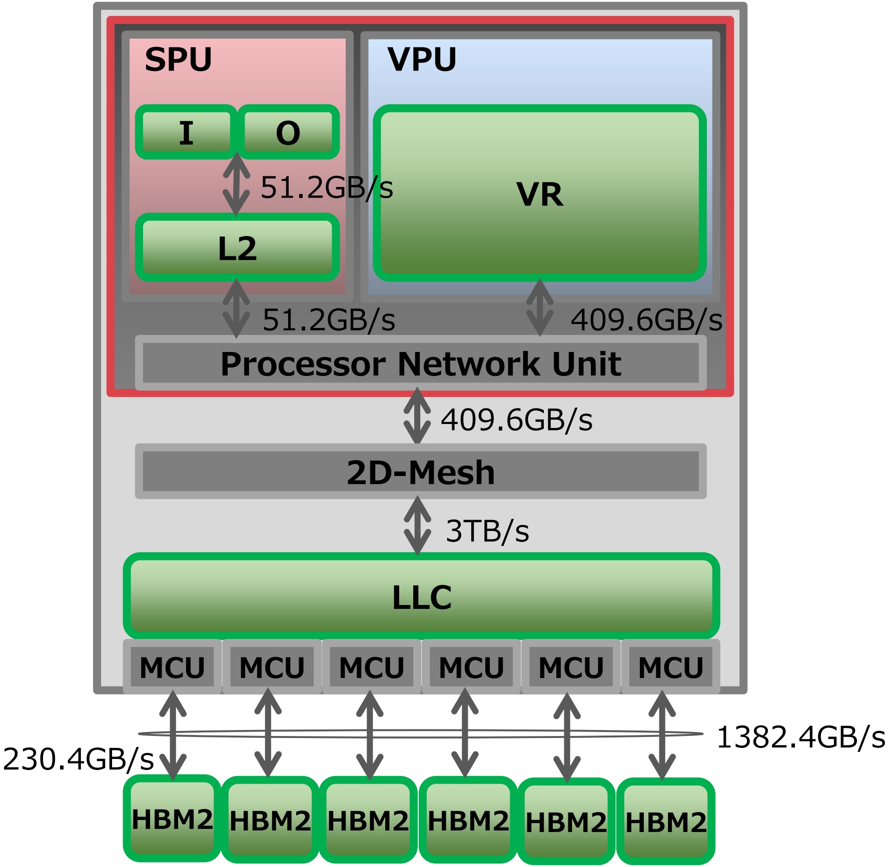
VE Processor

VE10E Specification	
cores/CPU	8
core performance	~307GF(DP) ~614GF(SP)
CPU performance	~2.45TF(DP) ~4.91TF(SP)
cache capacity	16MB shared
memory bandwidth	1.35TB/s
memory capacity	24, 48GB



HBM2 memory x 6

Memory Subsystem Implementation



SPU

■ I cache:	32kB
■ O cache:	32kB
■ L2 cache:	256kB

VPU

- Vector Register (VR)
  128kB = 256e x 64
  376kB physical

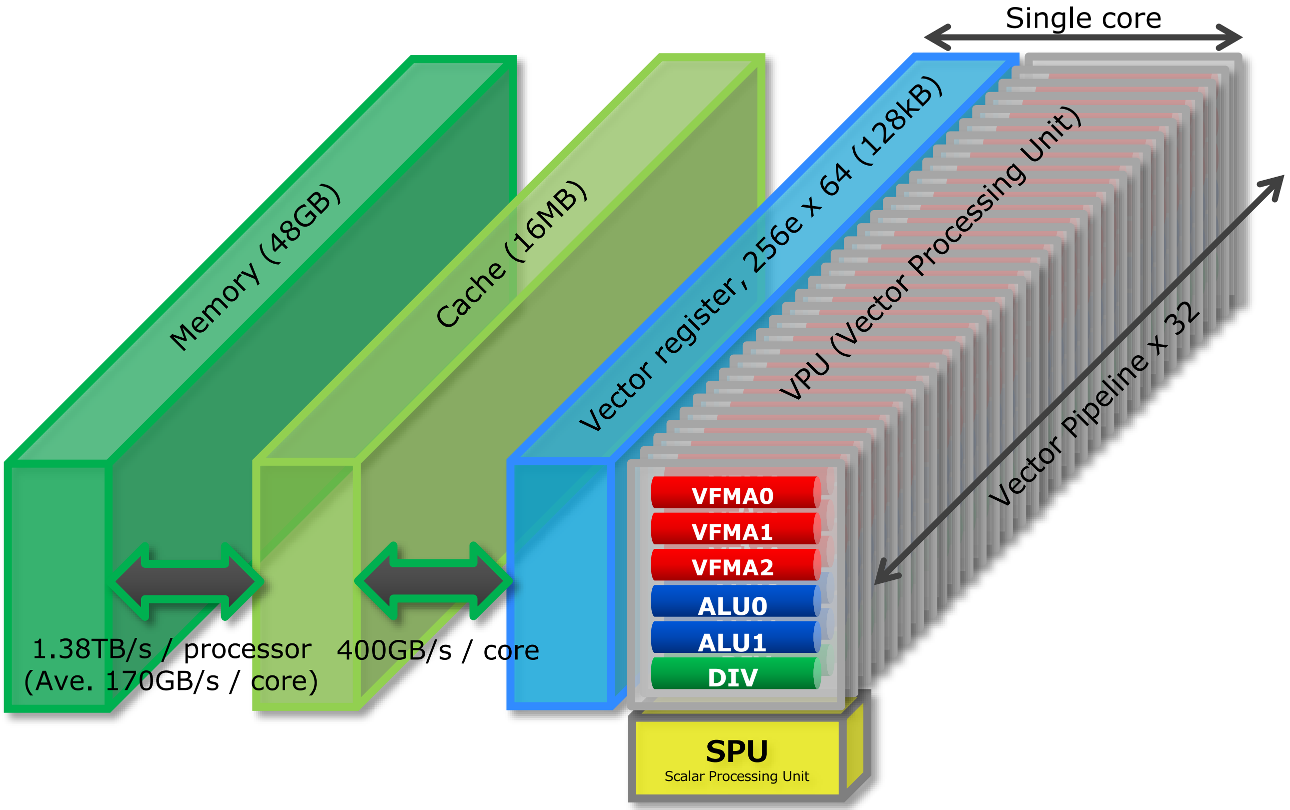
LLC

- 16MB
- Write back
- ADB/MSHR functions

Memory

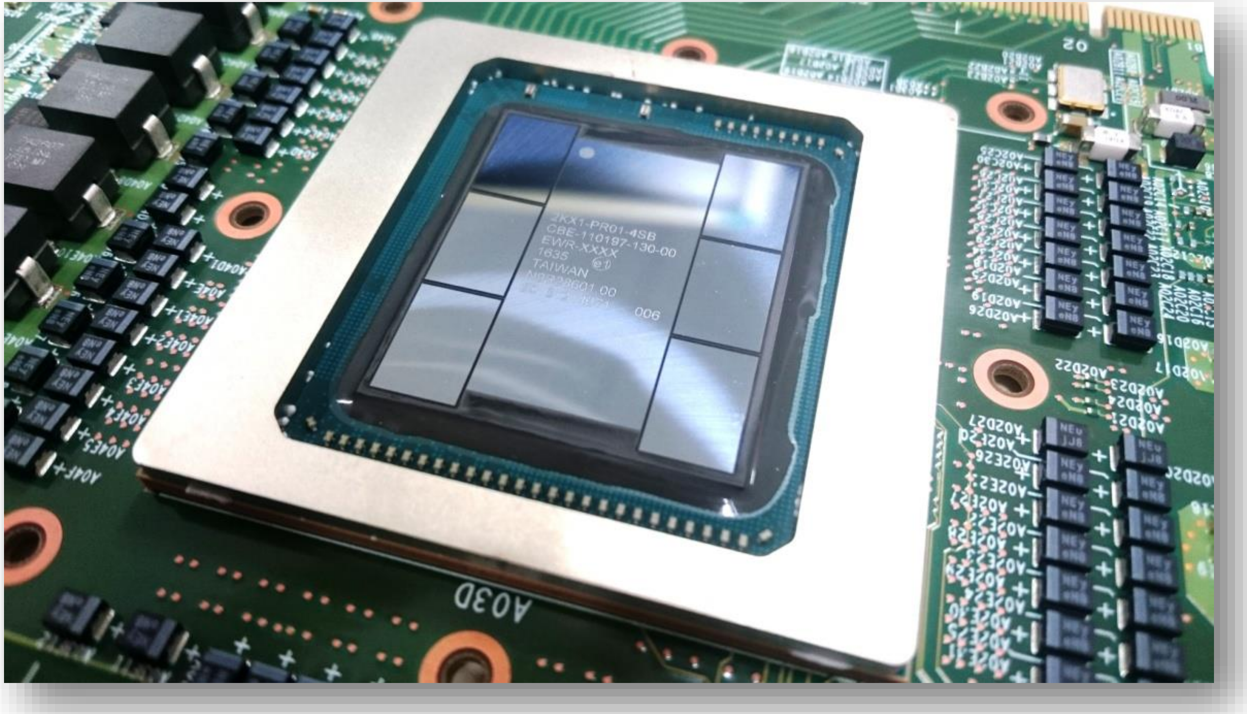
- 48GB
- HBM2 x6

Memory Architecture

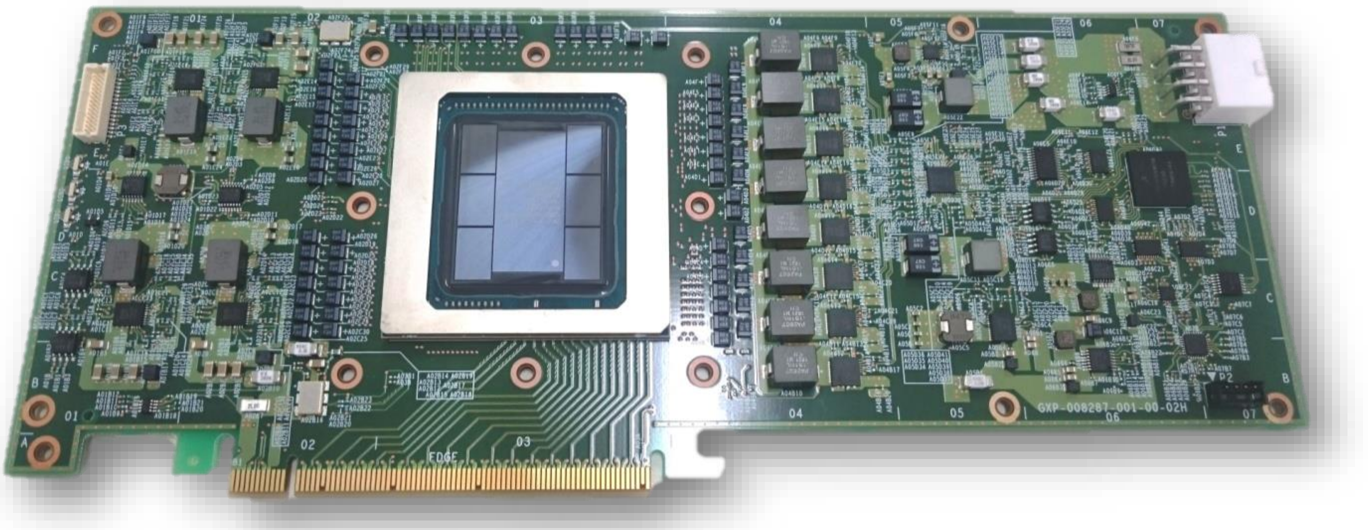


CPU Package

World’s first implementation of 6 HBM2 memories



Card Implementation



- Standard PCIe implementation
- Connector: PCIe Gen.3 x16
- Double height (same form factor as Nvidia)
- <300W (DGEMM ~210W/VE, STREAM ~200W/VE, HPCG ~215W/VE)

Products

A500
A400

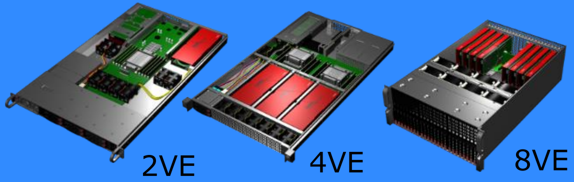
Supercomputer Model

- For large scale configuration
- DLC with 40C water

A300

Rack Mount Model

- Flexible configuration
- Air Cooled

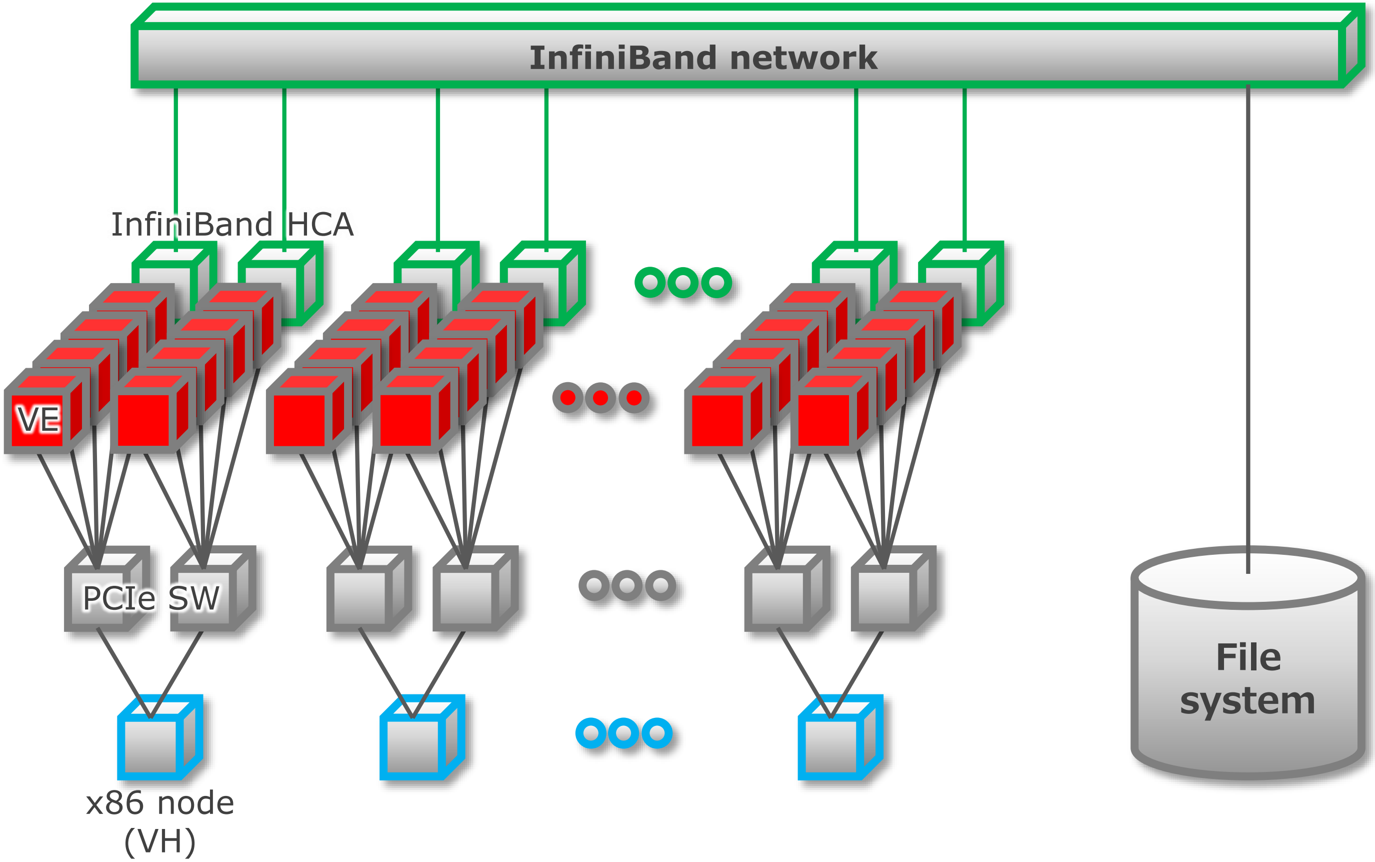


A100

Tower Model

- For developer/programmer
- Tower implementation

Large System Configuration

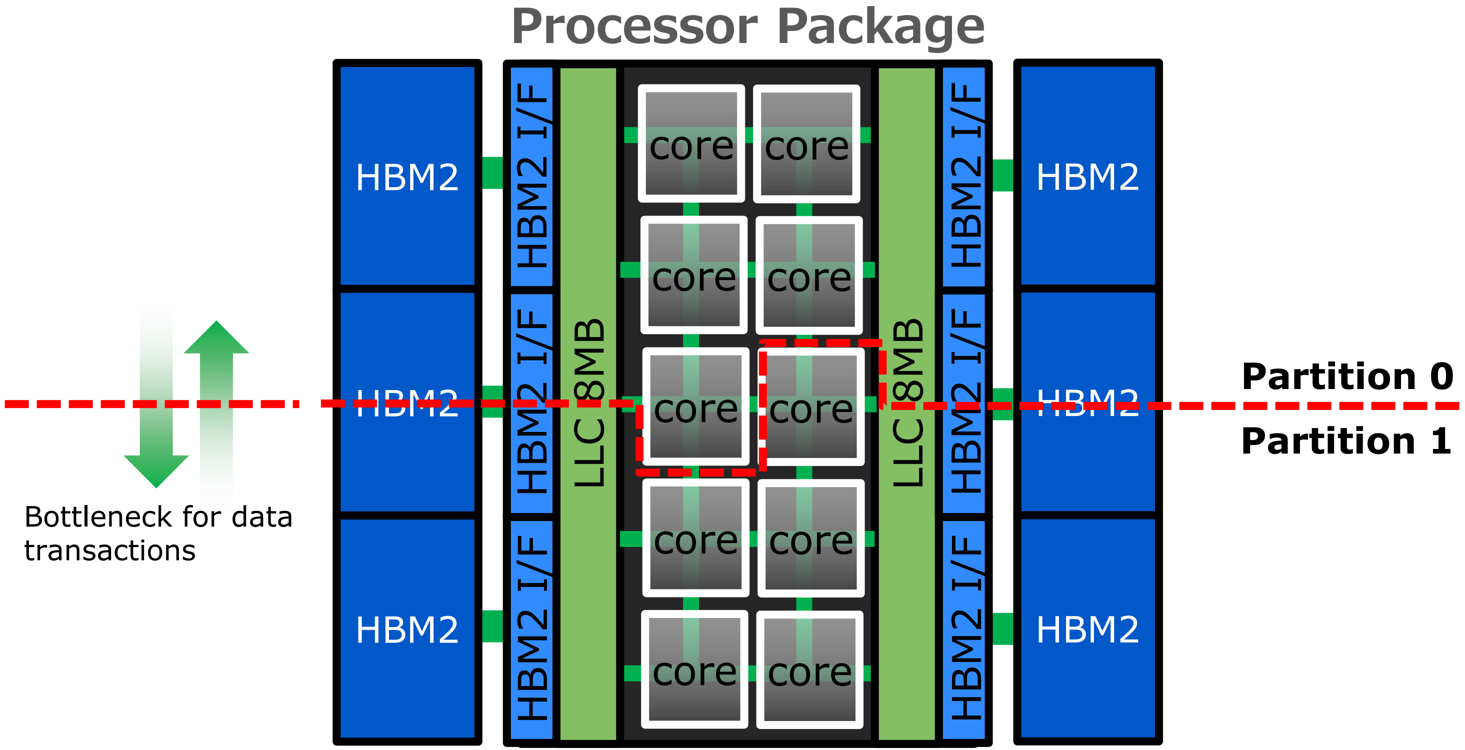


The latest version of Aurora Software

Software	Version	Date
VEOS	2.2.0	2019/10
MMM	1.2.17	2019/10
VMC Firmware	1.5.11-1	2019/10
InfiniBand for SX-Aurora TSUBASA	1.3.0	2019/10
License Management	1.3-1	2019/10
SDK; NEC C/C++ Compiler, NEC Fortran Compiler	2.5.1	2019/11
SDK; Numeric Library Collection	2.0.0-2.1	2019/10
SDK; binutils	2.26-2.3	2019/11
SDK; Tuning tool PROGINF/FTRACE (veperf)	2.1.0	2019/05
SDK; Tuning tool NEC Ftrace Viewer	1.0.0	2018/02
SDK; NEC Parallel Debugger	1.0.0	2018/02
NEC MPI	2.3.0	2019/10
NEC Network Queuing System V (NQSV) ResourceManager / JobManipulator / JobServer	1.04	2019/10
NEC Scalable Technology File System (ScaTeFS) Server	3.2	2019/10
NEC Scalable Technology File System (ScaTeFS) Client	3.0.30.5	2019/10

VE Partitioning Mode

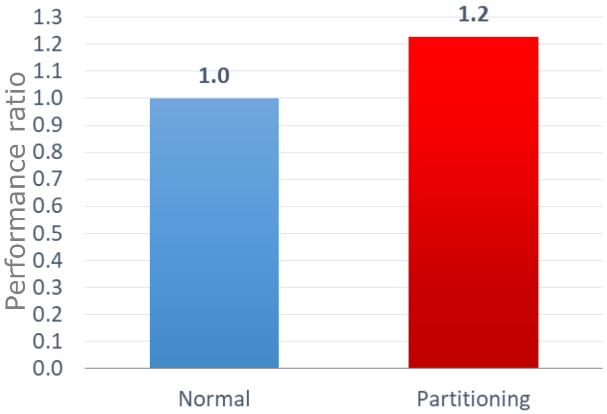
Partitioning Mode for Cache Acceleration



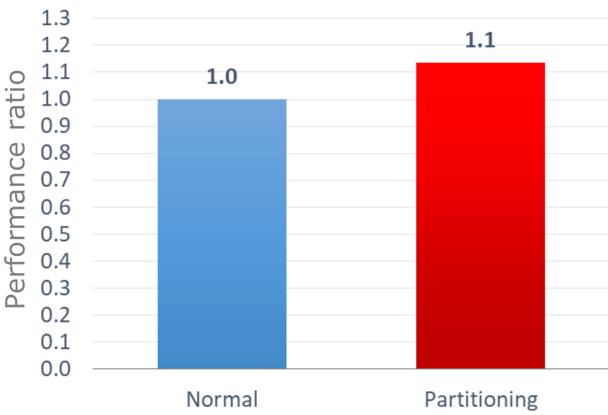
The partitioning mode:
One VE processor works as two partitions in order to avoid data transaction bottleneck. Due to this function, the sustained cache bandwidth is improved

Performance of the Partitioning Mode

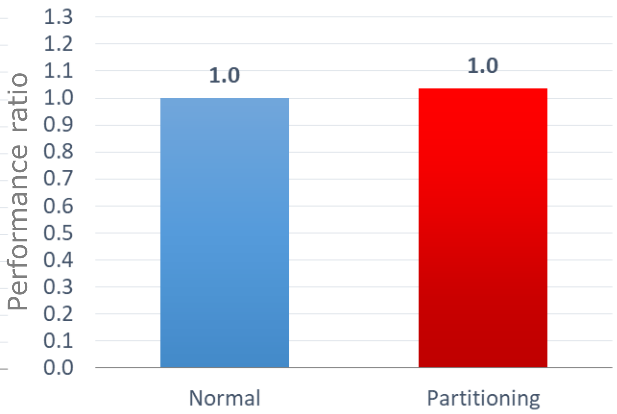
Front Flow Blue (CFD)



HPCG



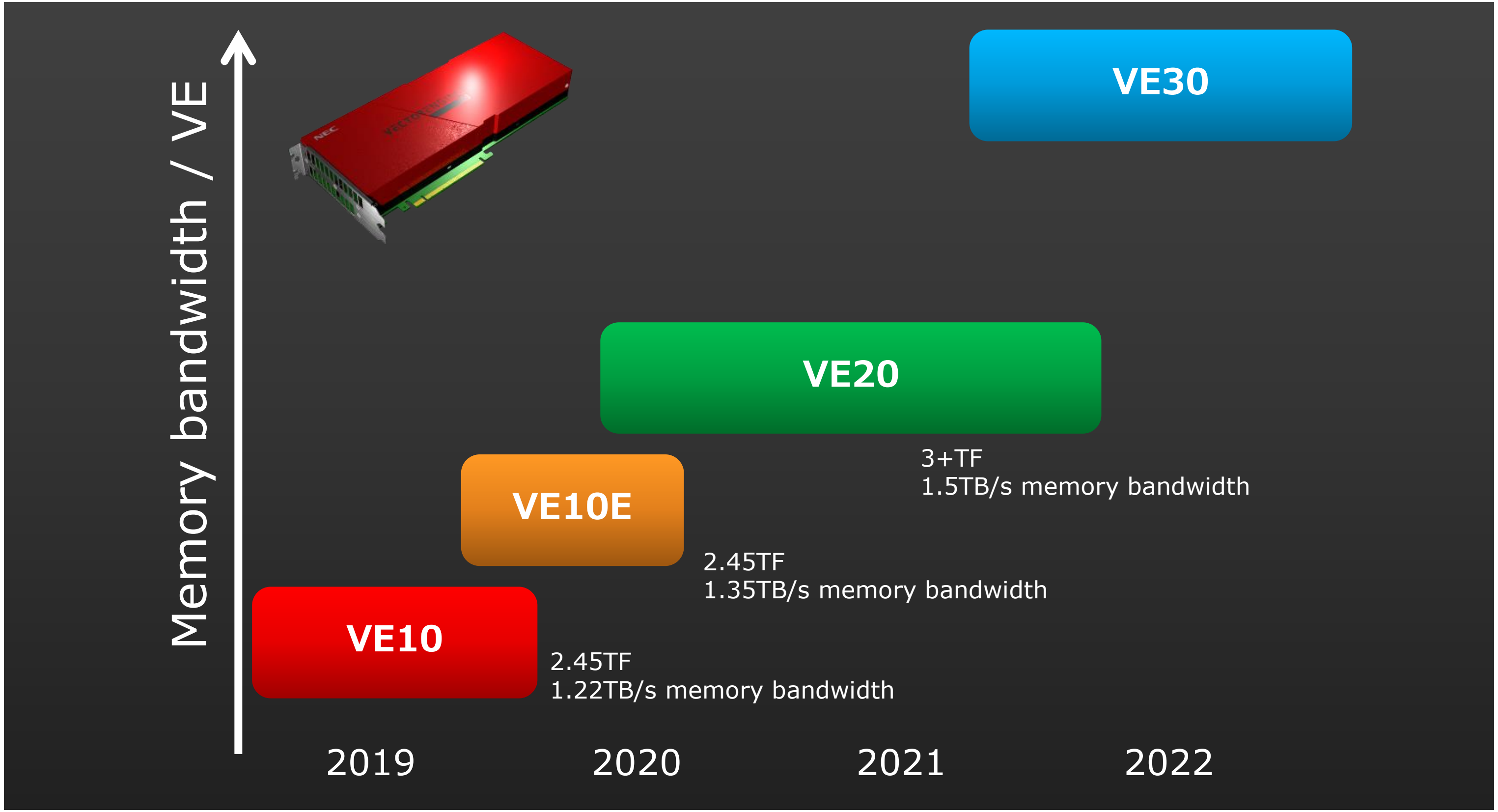
HIMENO Benchmark (CFD, Poisson Equation)



Cache bandwidth dependency

VE10E, VE20

Roadmap



Memory bandwidth / VE

VE30

VE20

3+TF
1.5TB/s memory bandwidth

VE10E

2.45TF
1.35TB/s memory bandwidth

VE10

2.45TF
1.22TB/s memory bandwidth

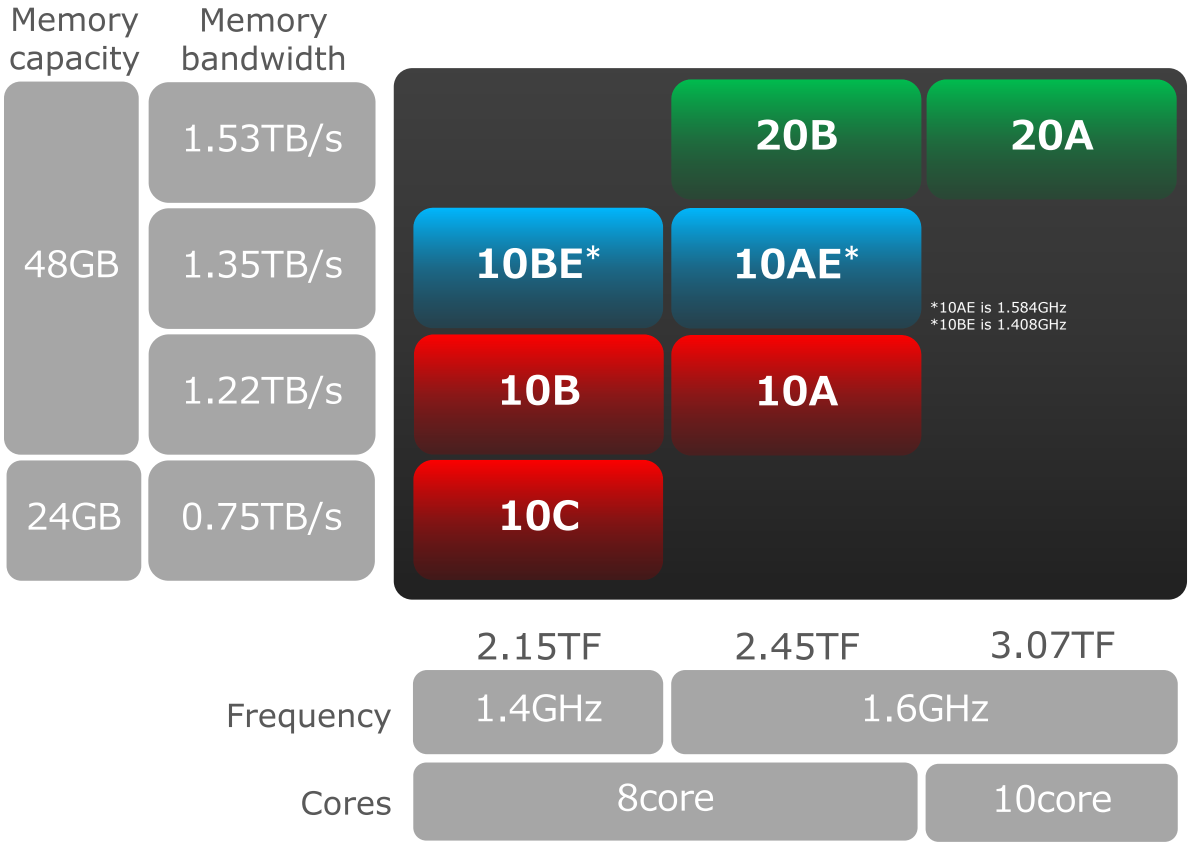
2019

2020

2021

2022

VE10/10E/20 SKU

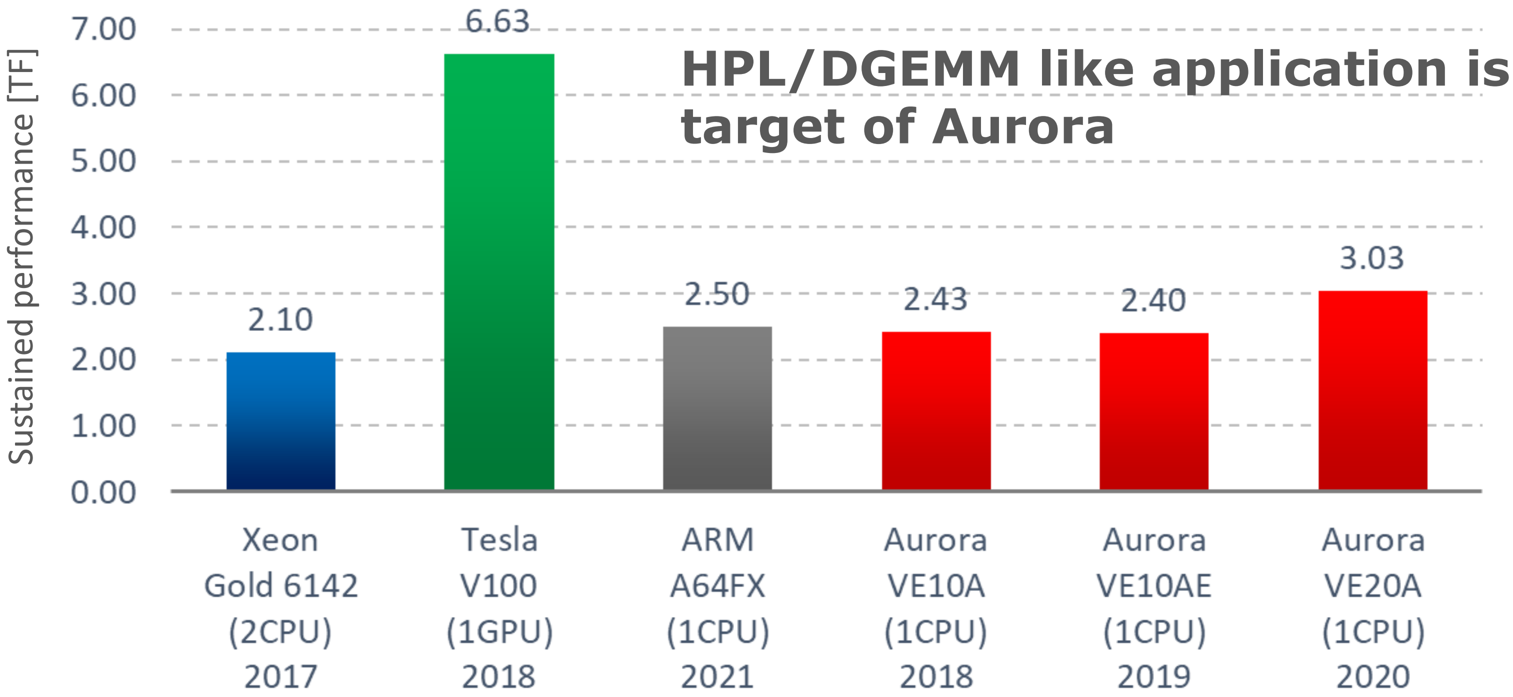


Memory capacity	Memory bandwidth			
48GB	1.53TB/s		20B	20A
	1.35TB/s	10BE*	10AE*	
	1.22TB/s	10B	10A	
24GB	0.75TB/s	10C		
		2.15TF	2.45TF	3.07TF
	Frequency	1.4GHz	1.6GHz	
	Cores	8core		10core

*10AE is 1.584GHz
*10BE is 1.408GHz

Performance

DGEMM (Calculation Capability Evaluation)



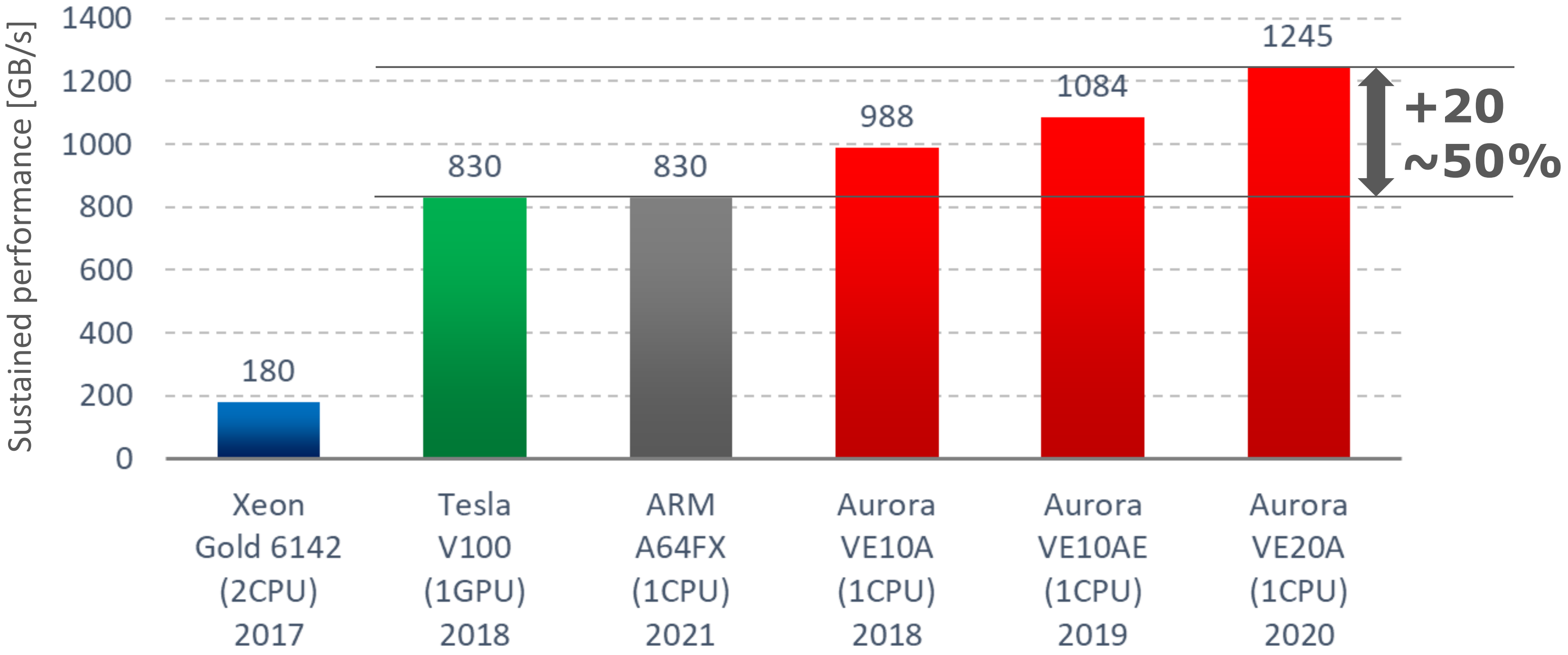
HPL/DGEMM like application is not target of Aurora

V100 result: AMD NEXT HORIZON
http://ir.amd.com/static-files/ef99f84b-e1ad-4e12-8058-f3488f4c47b7

ARM A64FX result: The post-K project and Fujitsu ARM-SVE enabled A64FX processor
https://indico.math.cnrs.fr/event/4705/attachments/2362/2942/CEA-RIKEN-school-19013.pdf

VE10A: 210W/card

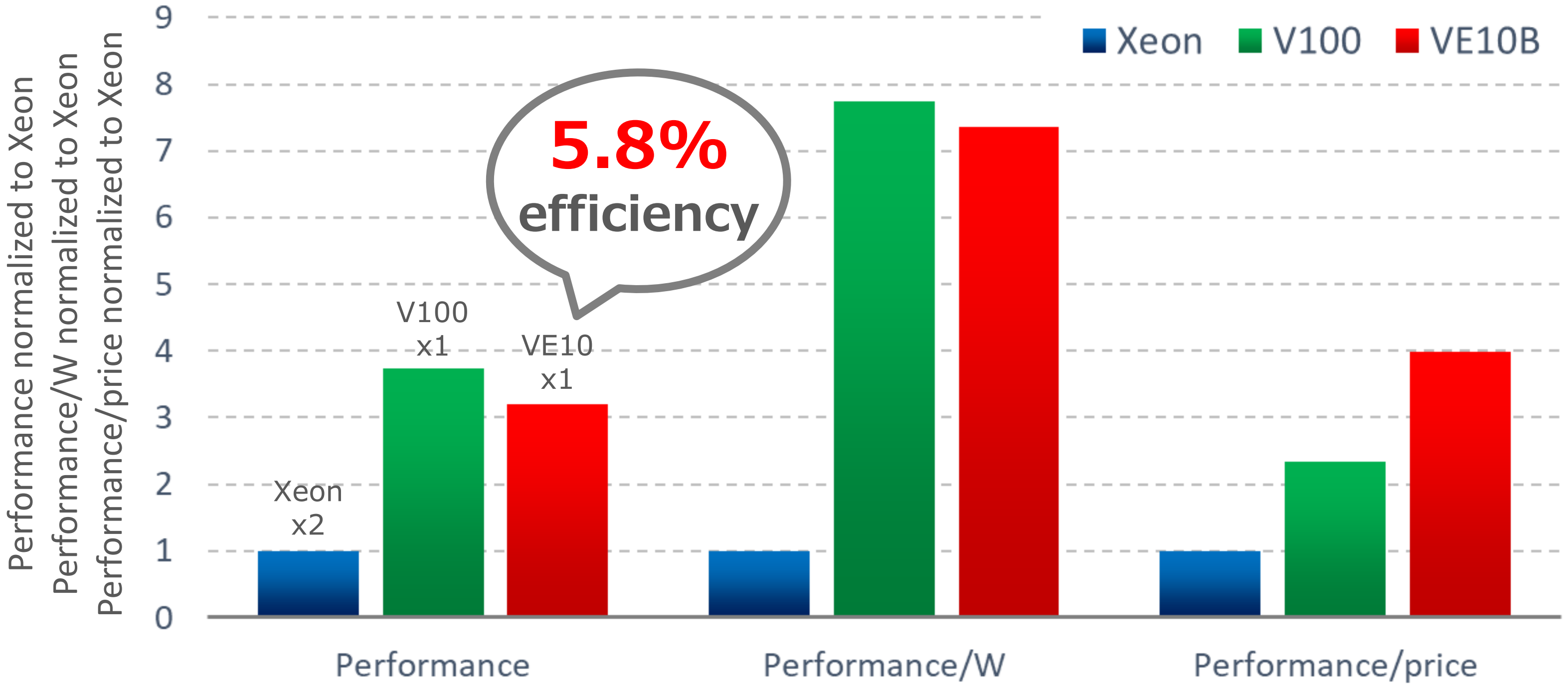
STREAM TRIAD (Memory Bandwidth Evaluation)



VE10A: 195W/card

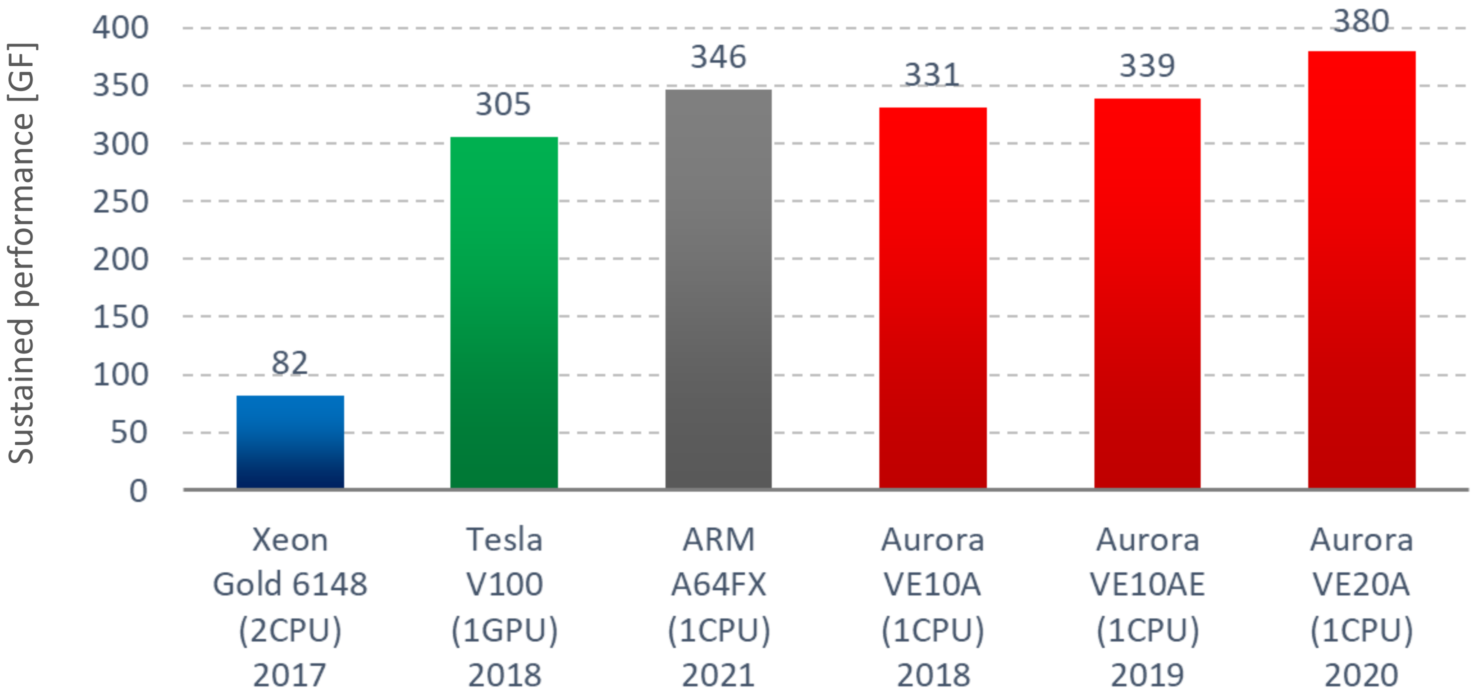
ARM A64FX result: The post-K project and Fujitsu ARM-SVE enabled A64FX processor
https://indico.math.cnrs.fr/event/4705/attachments/2362/2942/CEA-RIKEN-school-19013.pdf

HPCG



VE10A: 200W/card

HIMENO Benchmark (CFD, Poisson Equation)



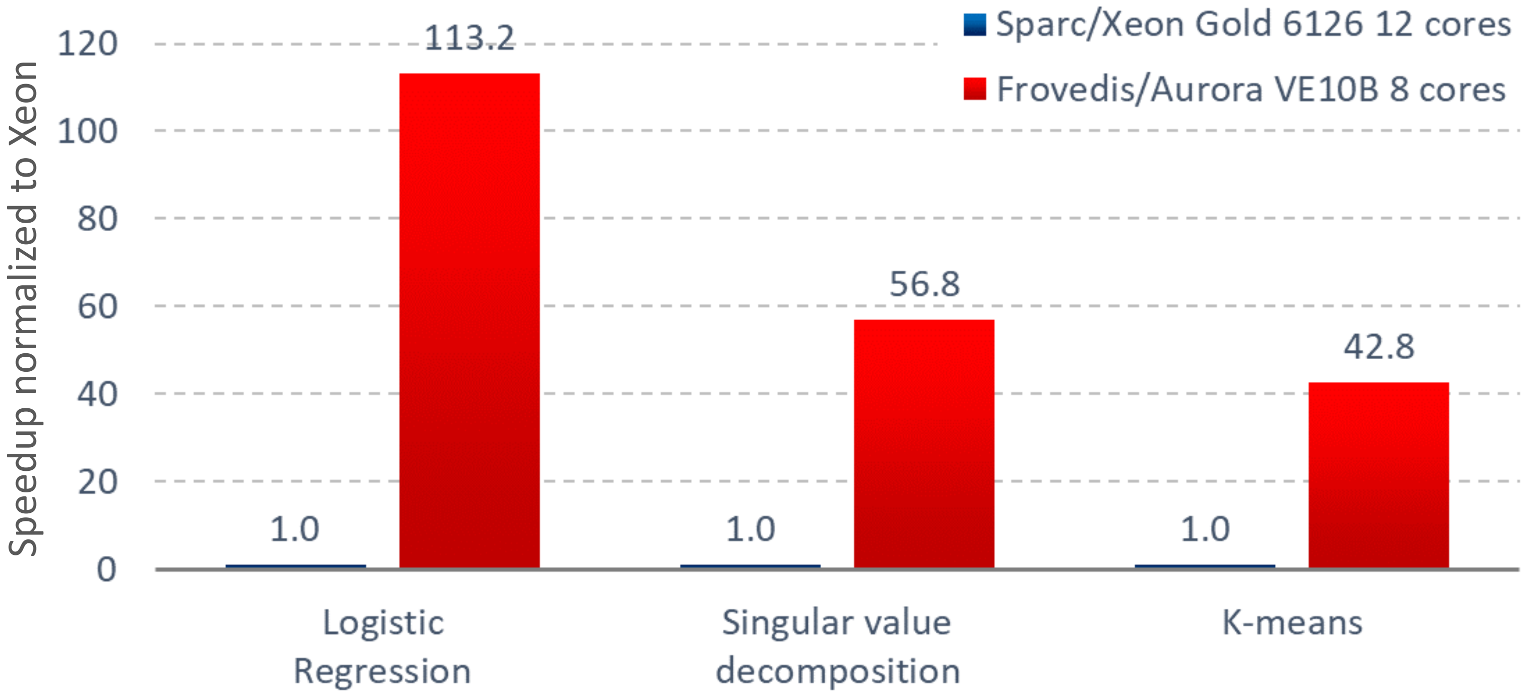
V100 result: Performance evaluation of a vector supercomputer SX-aurora TSUBASA
https://dl.acm.org/citation.cfm?id=3291728

ARM A64FX result: Supercomputer ”Fugaku” Formerly known as Post-K
https://www.fujitsu.com/global/Images/supercomputer-fugaku.pdf

VE10A: 210W/card

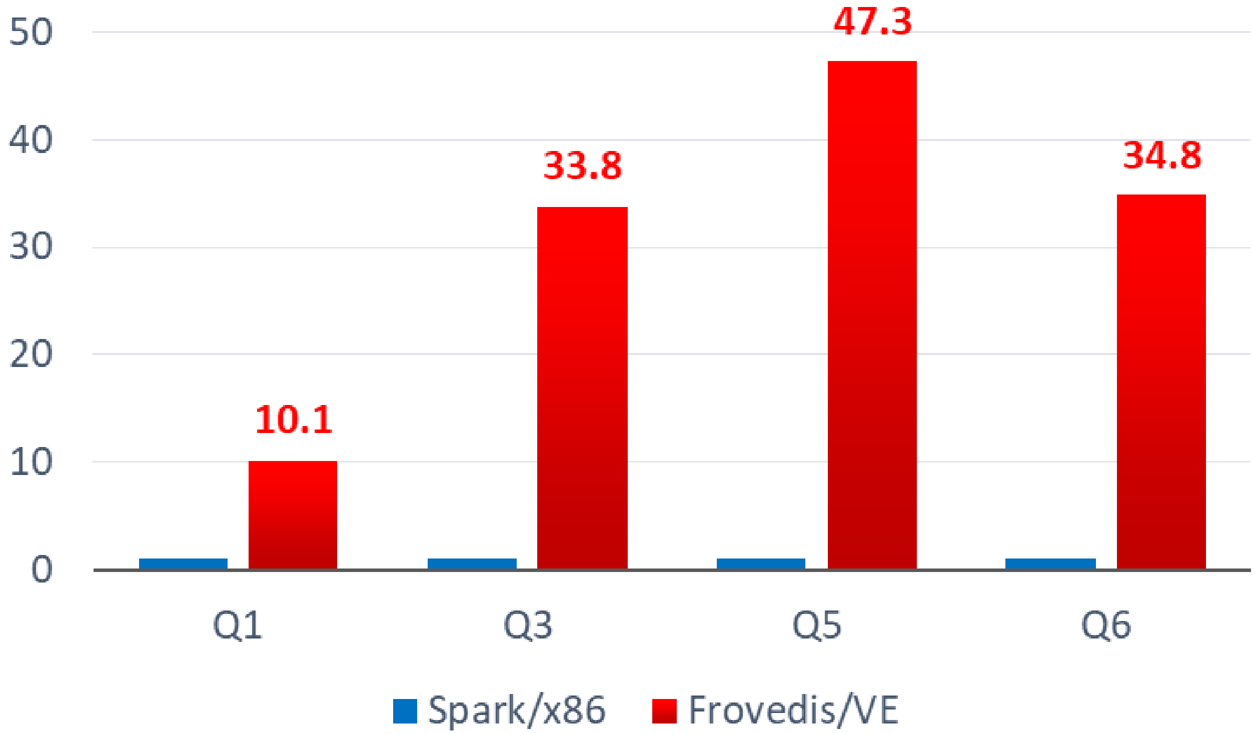
Machine Learning

Frovedis: NEC's Sparc library fully optimized for Aurora



DataFrame

Frovedis: NEC’s Sparc library fully optimized for Aurora

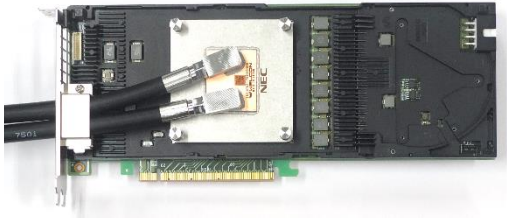


Q1: group by/aggregate
Q3: filter, join, group by/aggregate
Q5: filter, join, group by/aggregate (larger join)
Q6: filter, group by aggregate

High Density Product A412-8

A4:	A400 Series
1:	VE10E Generation
2:	Rome processor
-8:	Maximum number of VEs

High Density VH with DLC



Direct Liquid Cooling

VE DLC

To provide higher density
Hot water cooling

High density VH server

8VE/2U

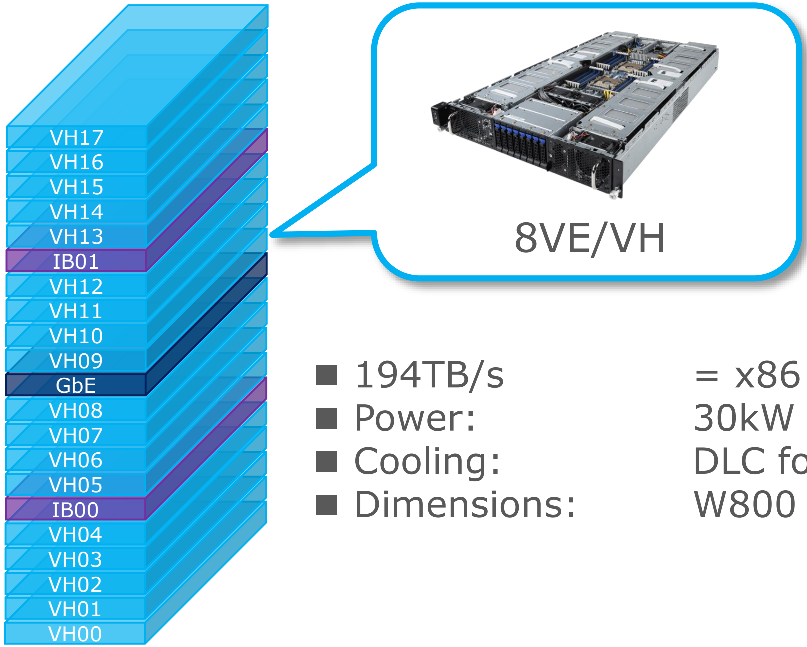
One VH consists of:
VE x8
AMD Rome processor x1
IB HCA x2

Performance:	19.6TF /VH
Memory bandwidth:	10.8TB/s /VH

High Density Rack w/ DLC

144VE/rack can be operated with 35C water

- High density and hot water cooling model
- VE10AE/10BE are supported



x18

144VE/rack
Up to 352TF/rack
194TB/s/rack

■ 194TB/s	= x86 processor x 1000
■ Power:	30kW for application run
■ Cooling:	DLC for VE and X86
■ Dimensions:	W800 x D1400 x H2200 (47U)

\Orchestrating a brighter world

NEC