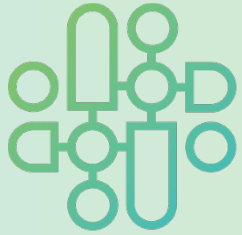


SC23
Denver, CO | i am hpc.

From Zero to Hero:
Conquering the Arm Neoverse

Brendan Bouffler (AWS)
Csaba Csoma (AWS)
John Linford (NVIDIA)
Matt Vaughn (AWS)
Conrad Hillairet (Arm Ltd)
Filippo Spiga (NVIDIA)

Acknowledgments

Join the **Arm HPC User Group** (https://a-hug.org/) Slack
Dedicated tutorial channel #sc23-tutorial-neoverse

John Linford

NVIDIA

Filippo Spiga

NVIDIA

Matt Vaughn

AWS

Brendan Bouffler

AWS

Csaba Csoma

AWS

Conrad Hillairet

ARM

iamhpc.

Tutorial objectives

WHAT IT IS ABOUT

- Show that it is just a myth: *compiling*, *executing*, *profiling* and *optimizing* on Arm-based HPC systems is hard
- Provide you an opportunity to experience Arm-based system first-hand

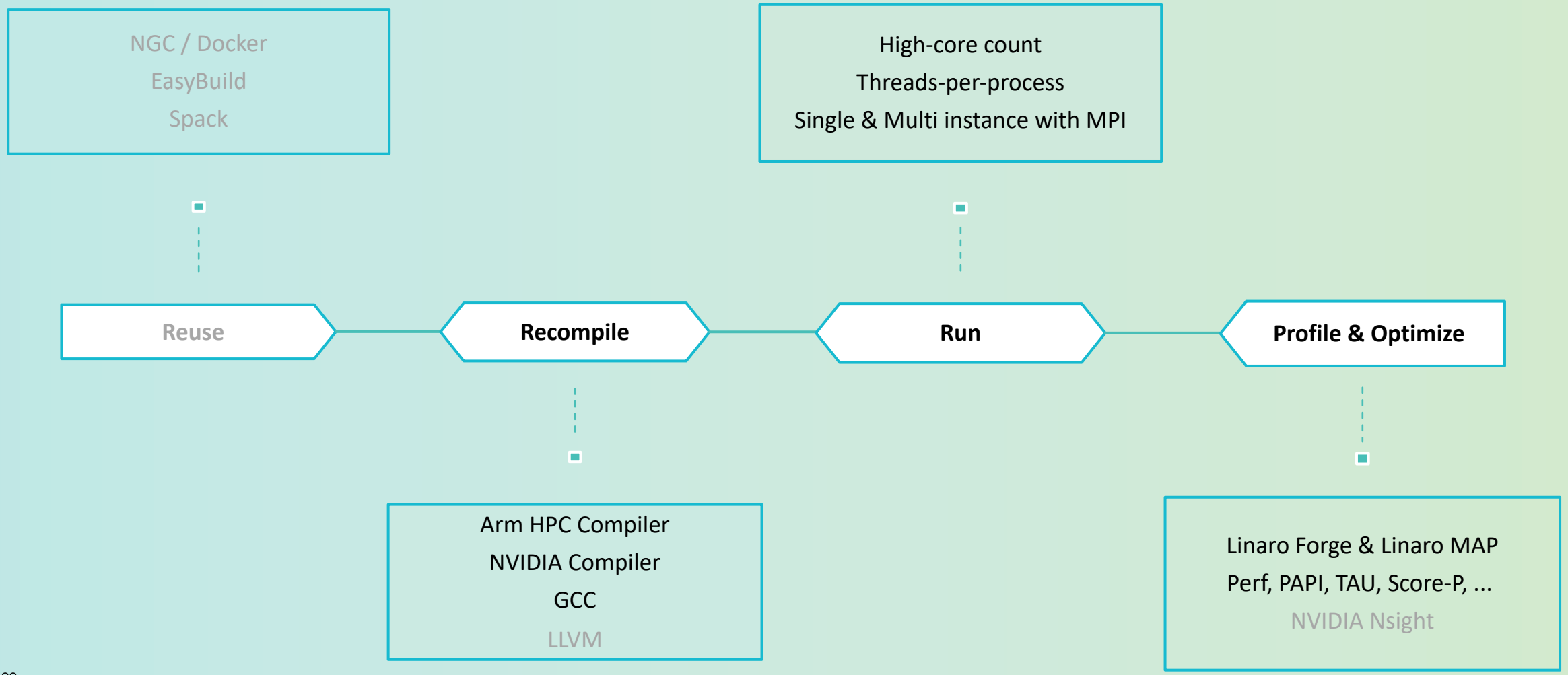
WHAT IS IT NOT ABOUT

- Dive into specific µ-arch features of Arm Neoverse IP portfolio
- Focus on accelerated CPU+GPU use-cases

LET MAKE THIS INTERACTIVE

During the sessions please engage and ask questions

Let the journey begin…



Agenda

Time	Duration	Content	Lead Speaker
8:30 - 8:35	5 mins	Welcome	Filippo
8:35 - 8:45	10 mins	**Introduction – Arm Technologies in HPC**	Conrad
8:44 -8:50	5 mins	Logistics & System Access	Matt
8:50 - 9:10	20 mins	**Compile & Execute**	Csaba
9:10 - 10:00	50 mins	Guided hands-on session / BYOC	
10:00 - 10:30	30 mins	Coffee break	
10:30 - 11:00	30 mins	**Profile & Optimize**	John
11:00 - 12:00	60 mins	Guided hands-on session / BYOC	

Tutorial GitHub repository (refreshed slides,additional user guides)
https://github.com/arm-hpc-user-group/sc23-tutorial-neoverse

Introduction

ARM TECHNOLOGIES IN HPC

Arm Technology is Defining the Future of Computing

A semiconductor design and software platform company

250+ Billion

Arm-based chips shipped to date.

29.2 Billion

Arm-based chips shipped in FY 2021.

650+

Active licensees, growing by 50+ every year.

The global leader in the development of licensable compute technology

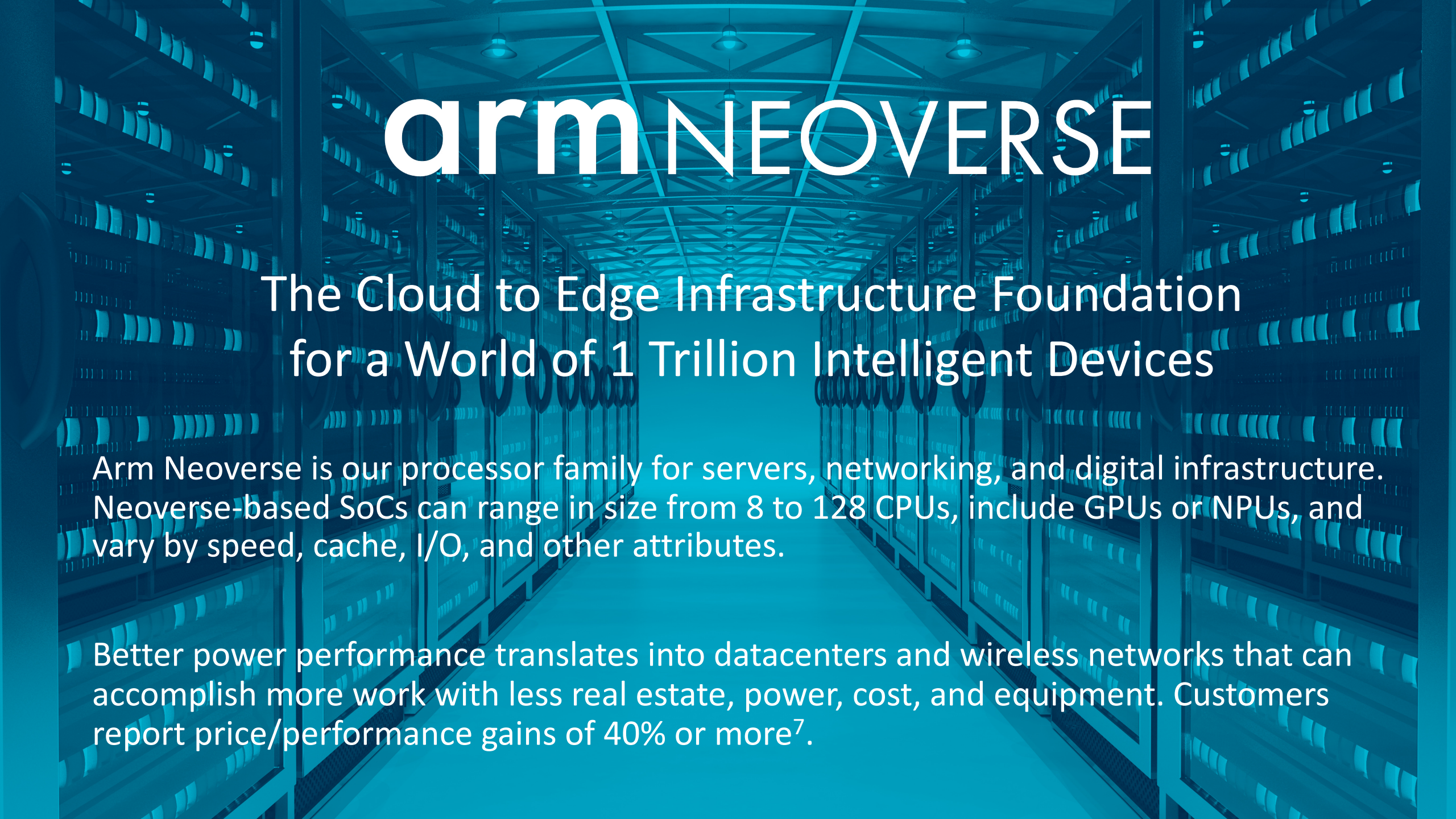
R&D excellence for semiconductor companies and large OEMs.

Arm's energy-efficient processor designs and software platforms enable advanced computing

Our technologies securely power products from the sensor to the smartphone and the supercomputer.

Arm delivers the foundational building blocks for trust in the digital world

Arm provides enhanced system-level security technologies such as Arm TrustZone and Arm Confidential Compute Architecture (CCA).

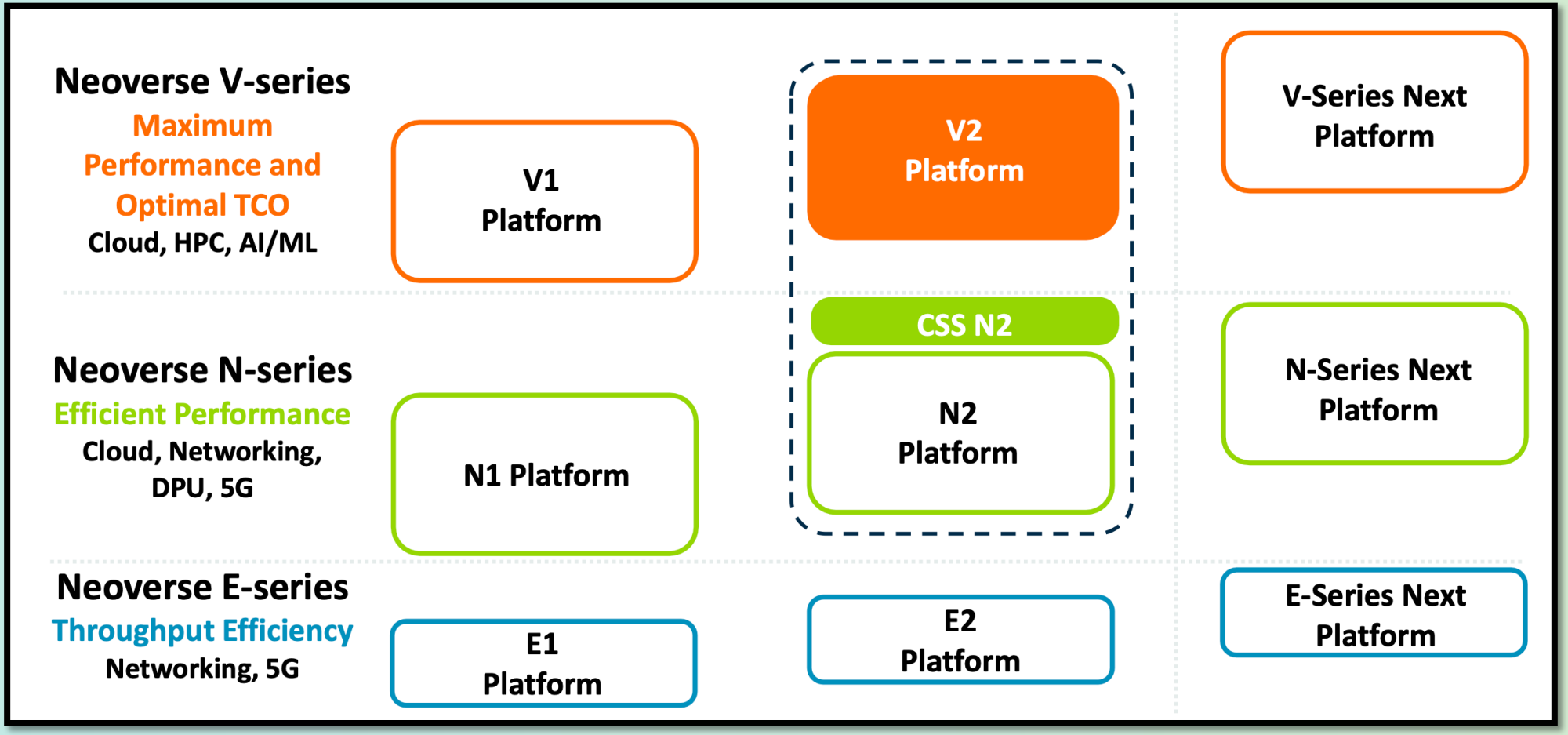


The Cloud to Edge Infrastructure Foundation for a World of 1 Trillion Intelligent Devices

Arm Neoverse is our processor family for servers, networking, and digital infrastructure. Neoverse-based SoCs can range in size from 8 to 128 CPUs, include GPUs or NPUs, and vary by speed, cache, I/O, and other attributes.

Better power performance translates into datacenters and wireless networks that can accomplish more work with less real estate, power, cost, and equipment. Customers report price/performance gains of 40% or more[7].

What is Arm Neoverse



arm NEOVERSE

Source: https://www.arm.com/products/silicon-ip-cpu/neoverse

Arm Industry Firsts in HPC

AWS

FIRST
ARM CPU FOR THE CLOUD

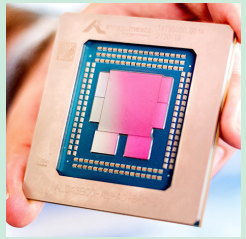
FUJITSU

FIRST
1TB/S MEM BW

AMPERE

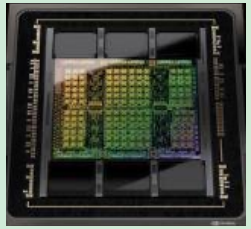
FIRST
>100 CORES PER CPU

AWS



FIRST
DDR5 PCIE GEN5.0

NVIDIA



FIRST
LPDDR5X MEM

SIPEARL

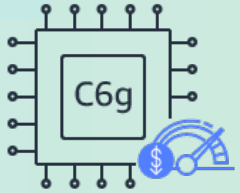
FIRST
European HPC Microprocessor

IP licensing brings **flexibility** – our partners use the flexibility to design their products for their chosen market

Although continuously differentiating and innovating – the standardization of Arm ISA and **standards** ensures compatibility with the software ecosystem

AWS Graviton

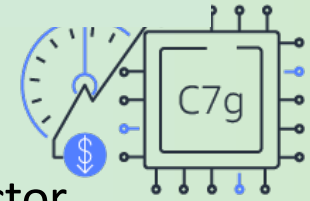
Graviton2 Processor



- 40% better price/perf vs similar instances
  72% power reduction
- ARMv8.2-a, FP16, LSE
  2x128-bit Neon
- Peak Memory B/W
  3.1 GB/s per core
  8x DDR4, encrypted
- Up to 64 cores, no NUMA
- 7 nm

Arm Neoverse N1

Graviton3 / 3E Processors



- 25% perf improvement vs Graviton2
  Graviton3E: 35% faster vector
- ARMv8.4-a, BF16, INT8, RNG
  2x256-bit SVE (4x128-bit NEON)
- Peak Memory B/W
  4.8 GB/s per core
  8x DDR5, encrypted
- Up to 64 cores, no NUMA
- 5 nm

Arm Neoverse V1

Source: https://github.com/aws/aws-graviton-getting-started
https://www.nec.com/en/press/202209/global_20220929_03.html

NVIDIA Grace Superchip

Grace CPU Superchip

- High Frequency
  3+ GHz
- Peak Flops
  7.1 TFlops*
  4x128-bit SVE2
  (4x128-bit NEON)
- Peak Memory B/W*
  up to 1 TB/s
  LPDDR5X
- 144 cores
- 2 NUMA Nodes
  900 GB/s C2C link
- NVIDIA SCF
  mesh fabric
- Distributed L3
  234 MB
- 500W TDP
  (CPU + MEM)

(*) exact numbers are SKU dependent

Arm Neoverse V2

Source : https://resources.nvidia.com/en-us-grace-cpu/nvidia-grace-cpu-superchip