arm AGI CPU

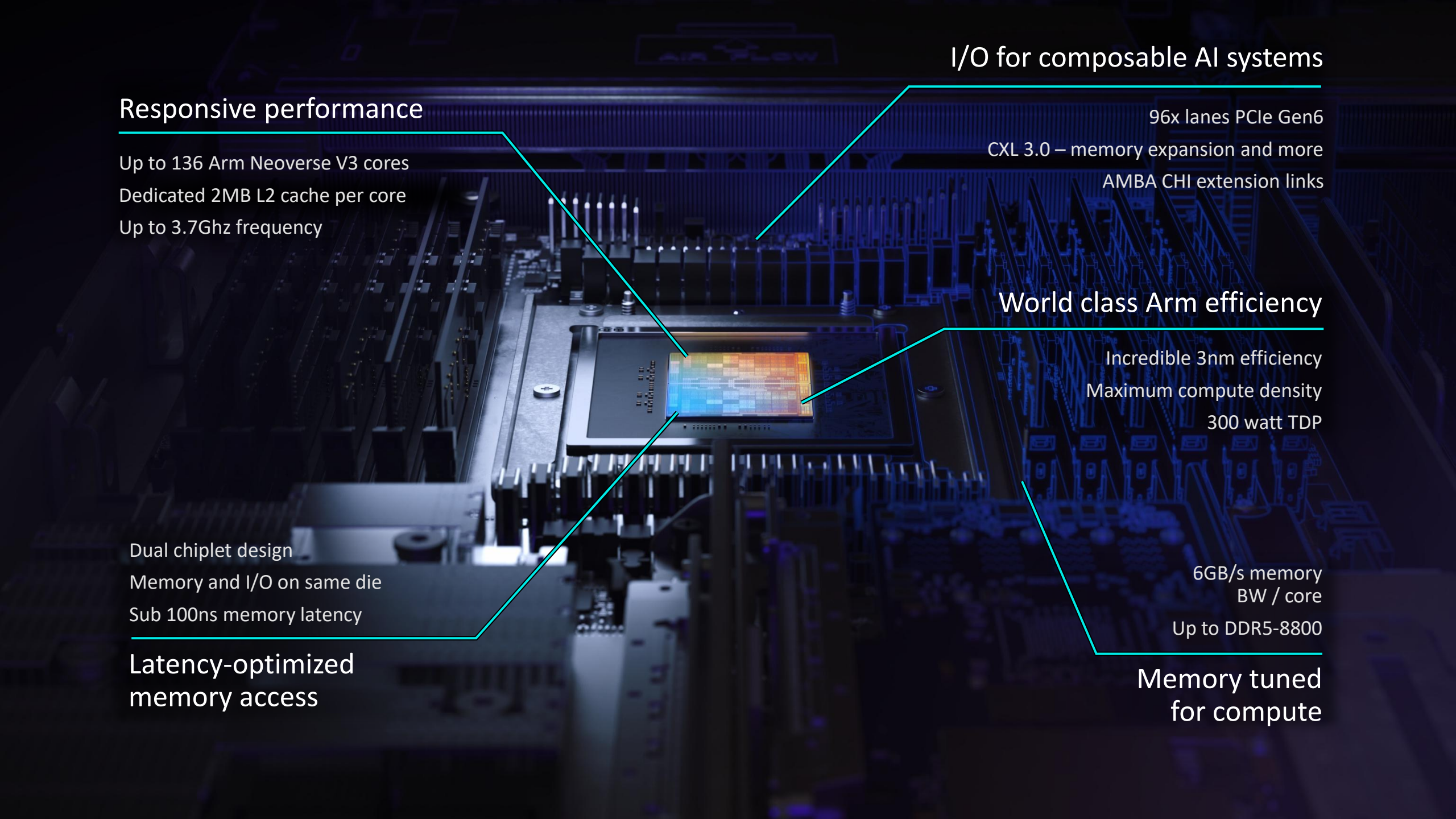
World’s most efficient agentic CPU

Performance

Scale

Efficiency

VT



Responsive performance

Up to 136 Arm Neoverse V3 cores

Dedicated 2MB L2 cache per core

Up to 3.7Ghz frequency

I/O for composable AI systems

96x lanes PCIe Gen6

CXL 3.0 – memory expansion and more

AMBA CHI extension links

World class Arm efficiency

Incredible 3nm efficiency

Maximum compute density

300 watt TDP

Latency-optimized memory access

Dual chiplet design

Memory and I/O on same die

Sub 100ns memory latency

Memory tuned for compute

6GB/s memory BW / core

Up to DDR5-8800

Air cooled

36kW open rack V3

30x 2 node 1U servers

8,160 performant CPU cores

180TB+ low latency memory

Liquid cooled

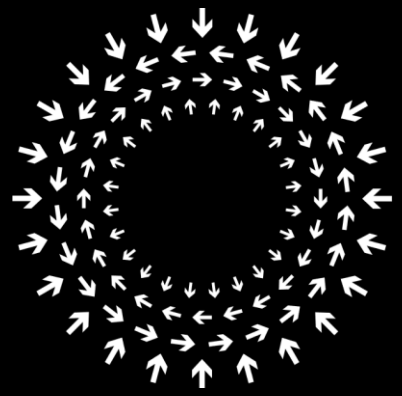
200kW open rack wide

42x 8 node 1U servers

45,696 performant CPU cores

1PB+ low latency memory

Maximum performance, density & efficiency

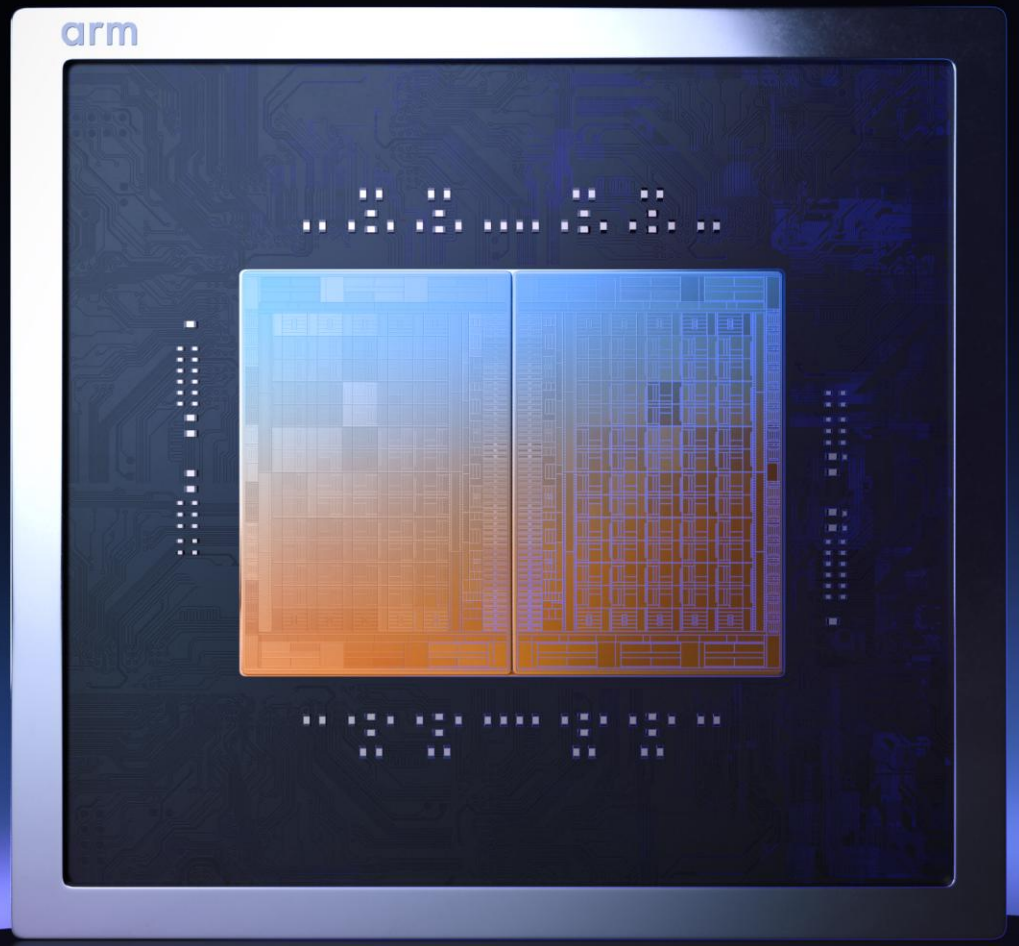


OPEN Compute Project ®

Contributions across the full stack

- Platform foundation
- Security & access
- Validation & tooling

arm AGI CPU



ASRock Rack

Lenovo

tsmc

Amkor Technology

Quanta Computer

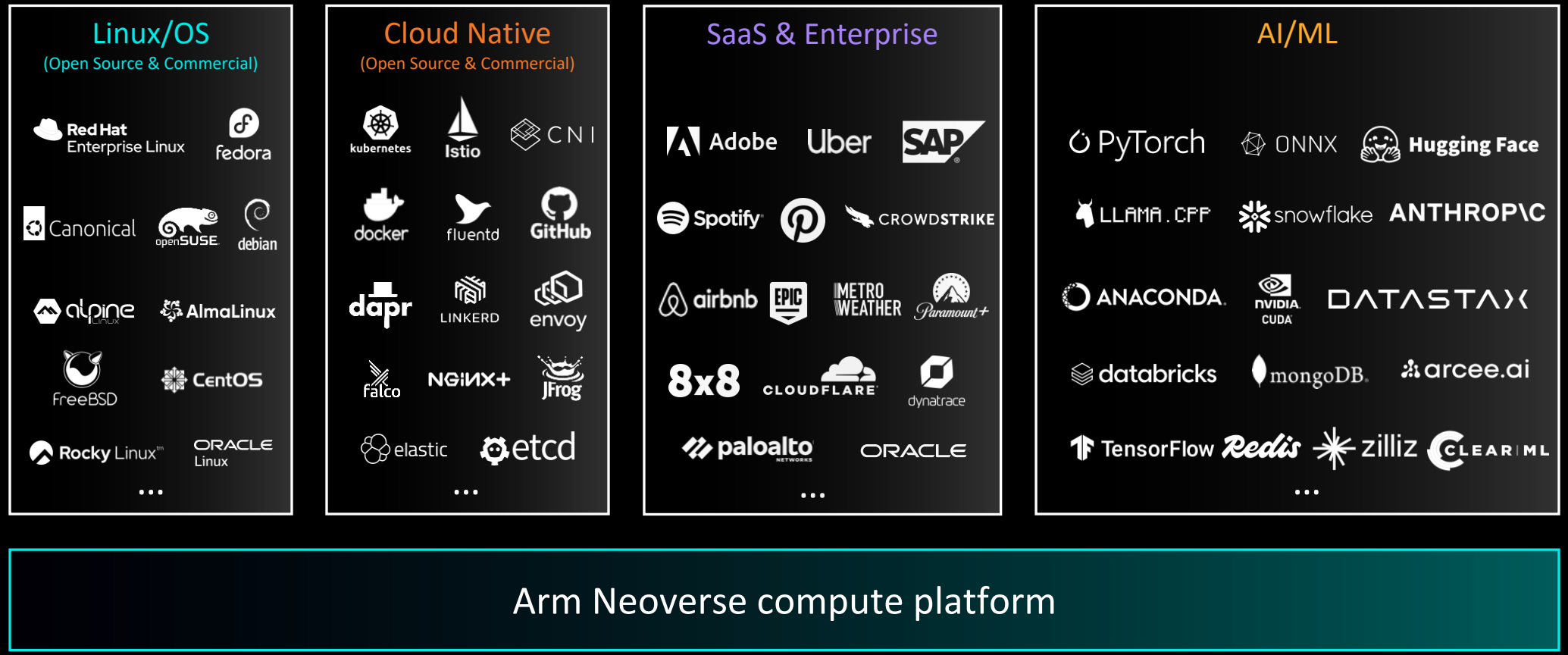
SUPERMICRO

SAMSUNG

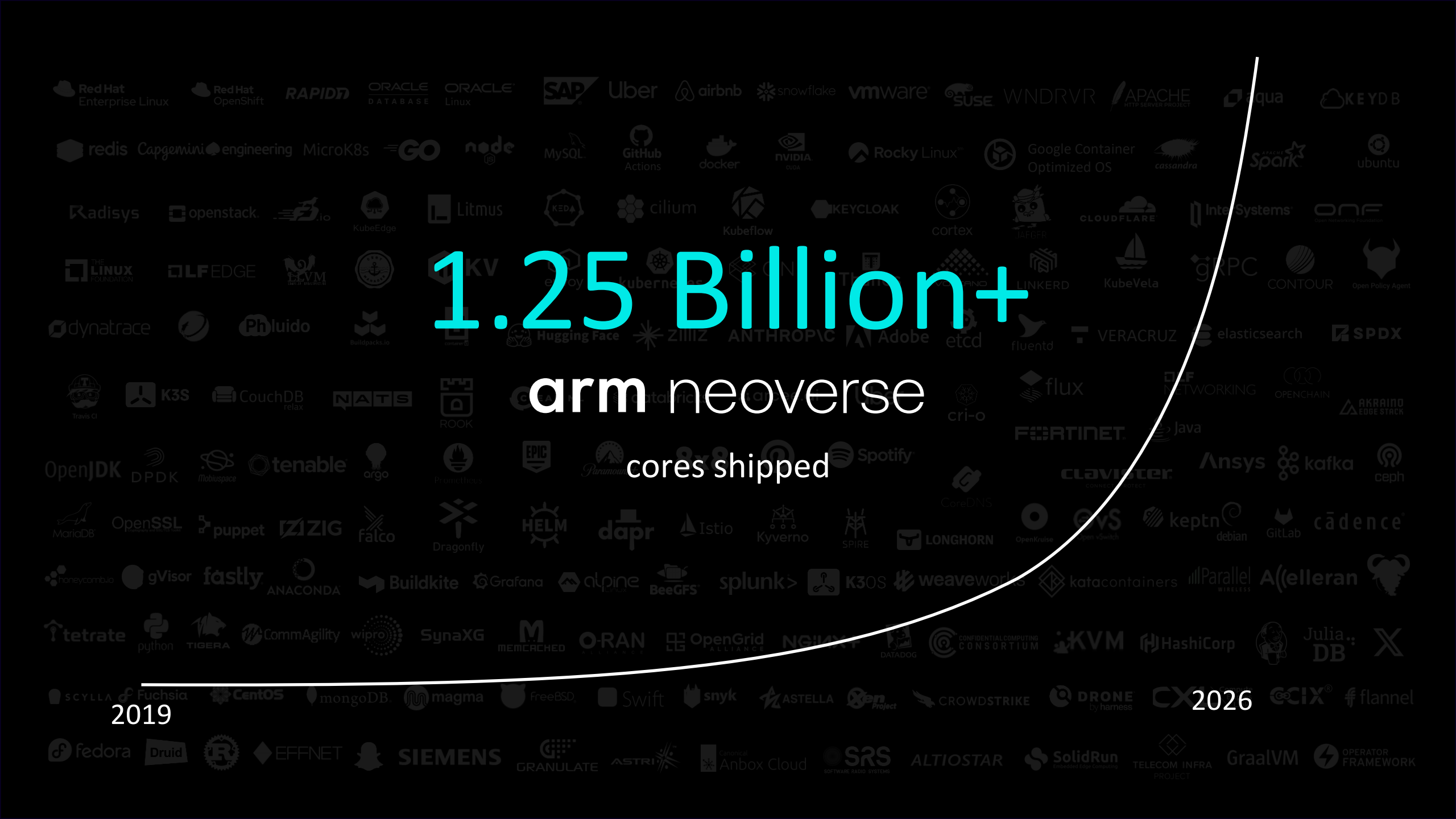
micron

SK hynix

Available now



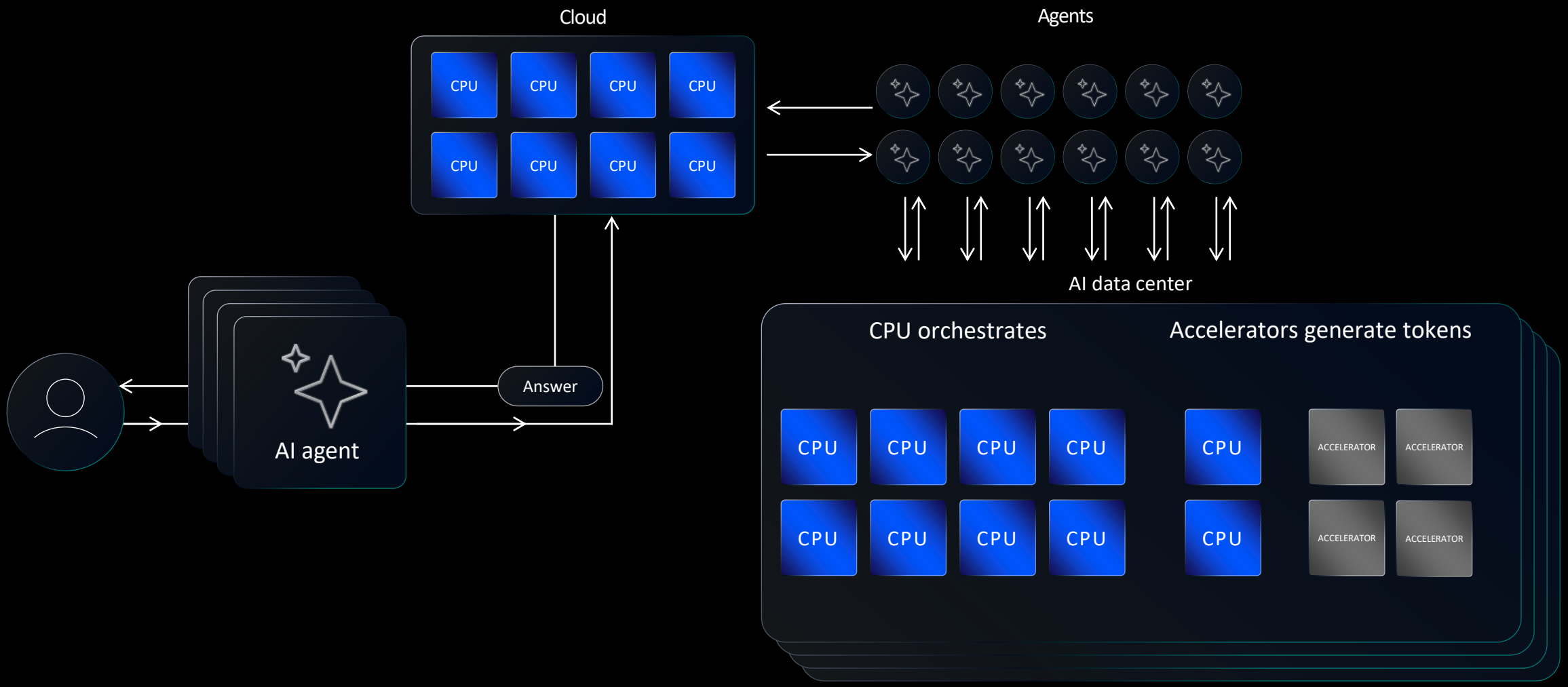
Arm Neoverse compute platform


1.25 Billion+
arm neoverse
cores shipped
2019
2026

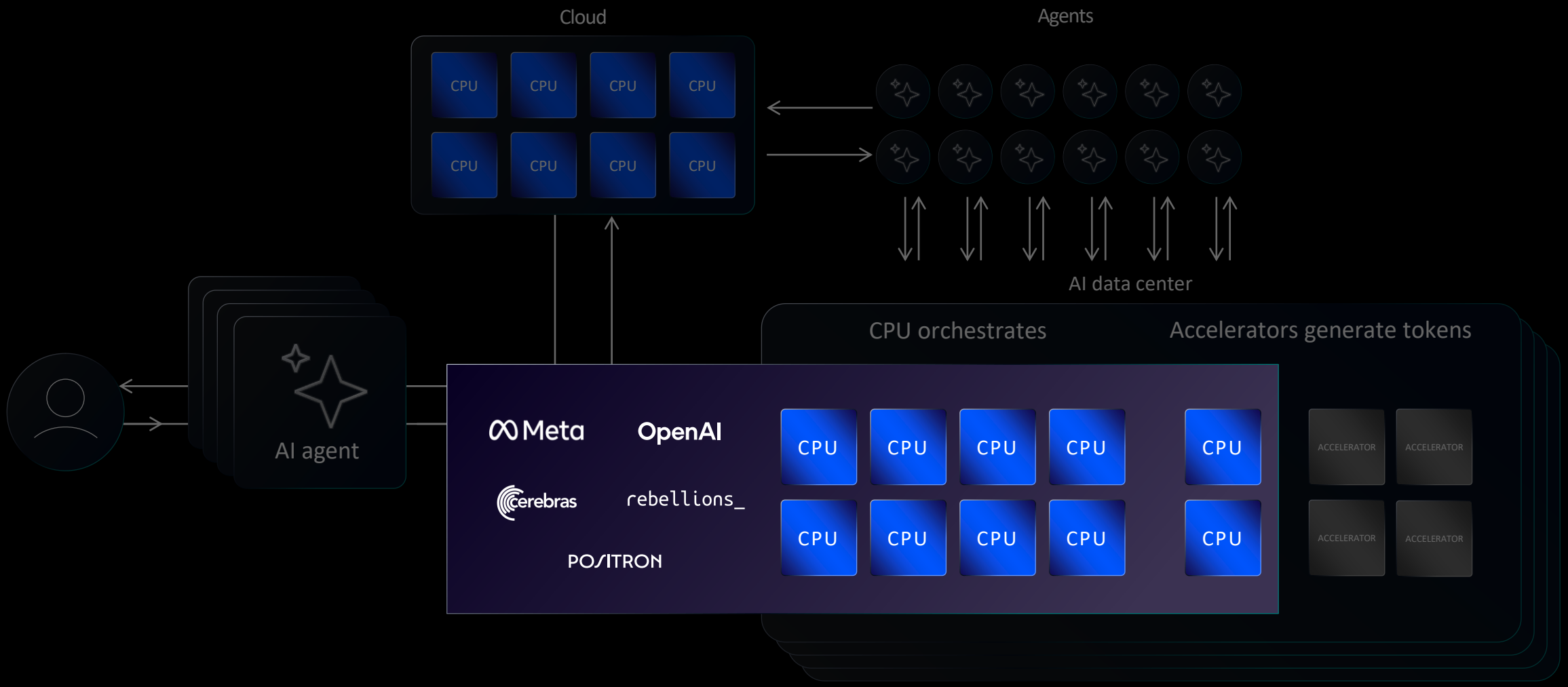
Paul Saab

Meta

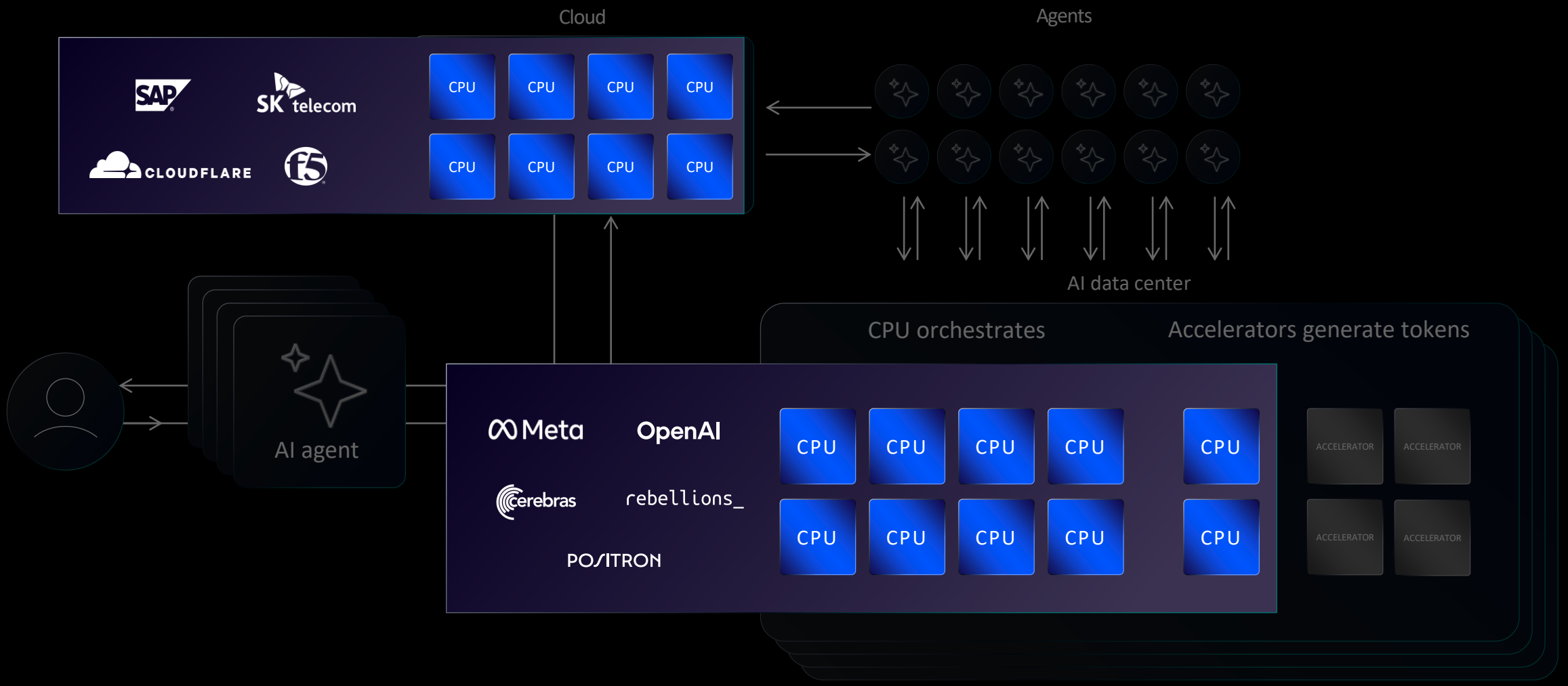
The agentic data center



The agentic data center



The agentic data center



cerebras

CLOUDFLARE

f5

Meta

arm

OpenAI

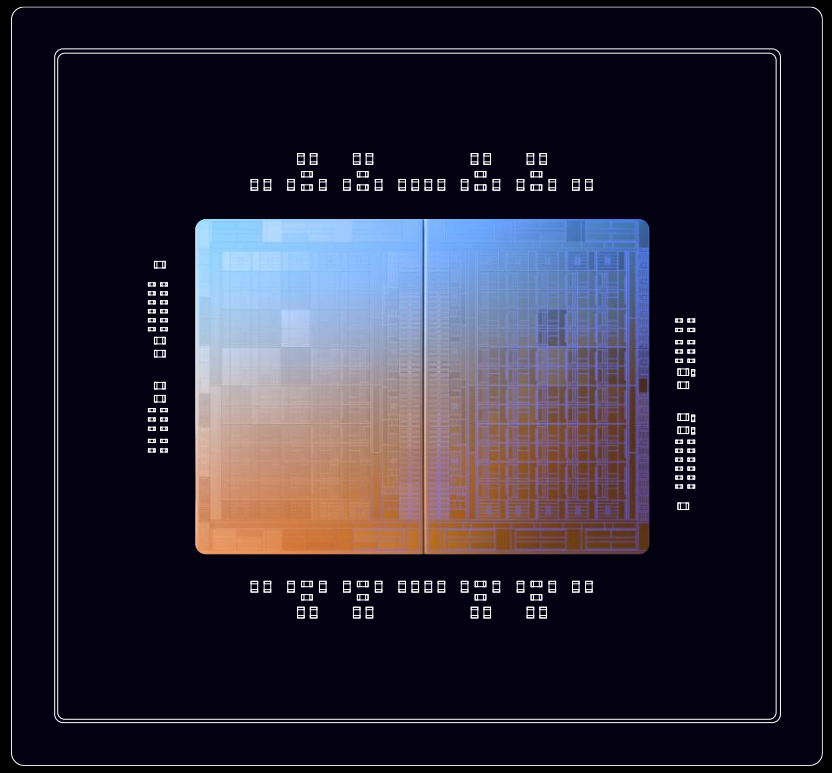
POSITRON

rebellions_

SAP

SK telecom

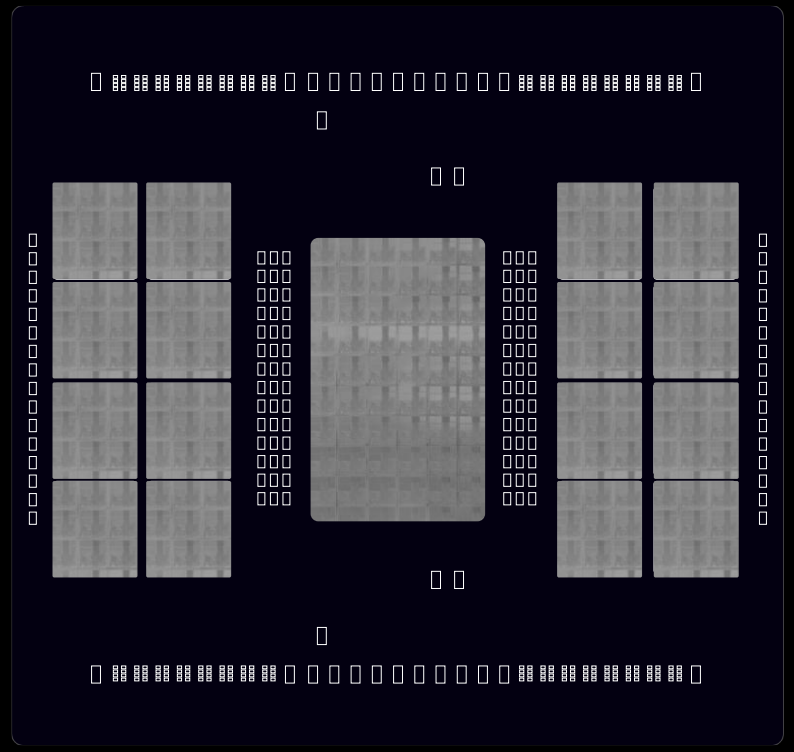
arm AGI CPU



High performance cores
World class efficiency
Low latency design

Performance scales. Power stays predictable.

x86 CPU



Execution overhead
Legacy feature support
Modularity over latency

Performance throttled. Technical debt.

Optimizing performance, scale, and efficiency

Efficiency

Purpose-built, no legacy overhead

No wasted energy or silicon

Scale

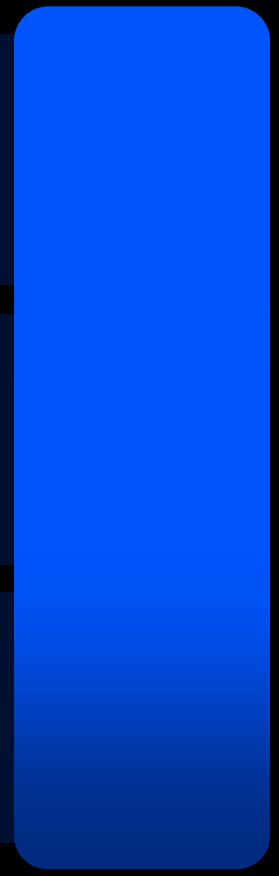
Linear scaling across cores

Memory and I/O keep pace

Performance

More work per cycle

Full performance, always

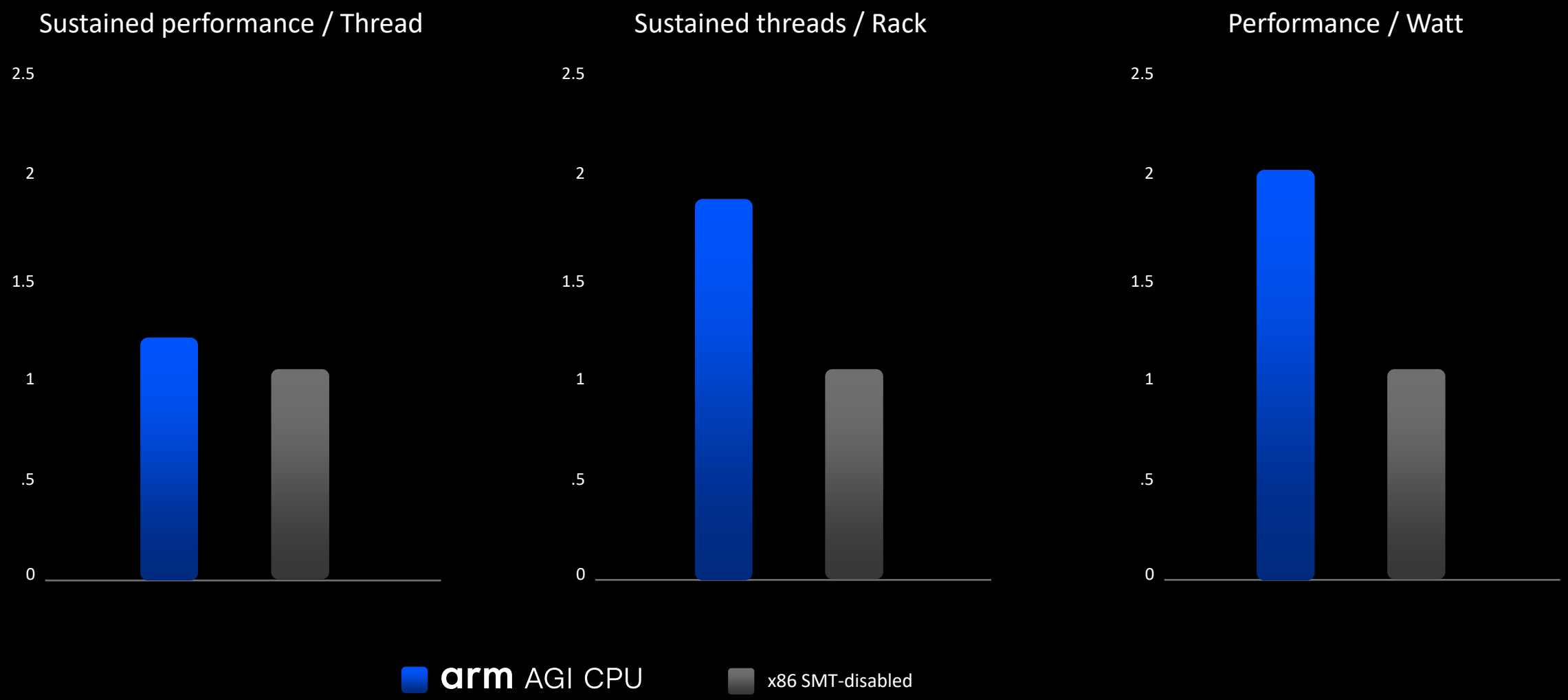


arm AGI CPU

x86 CPU

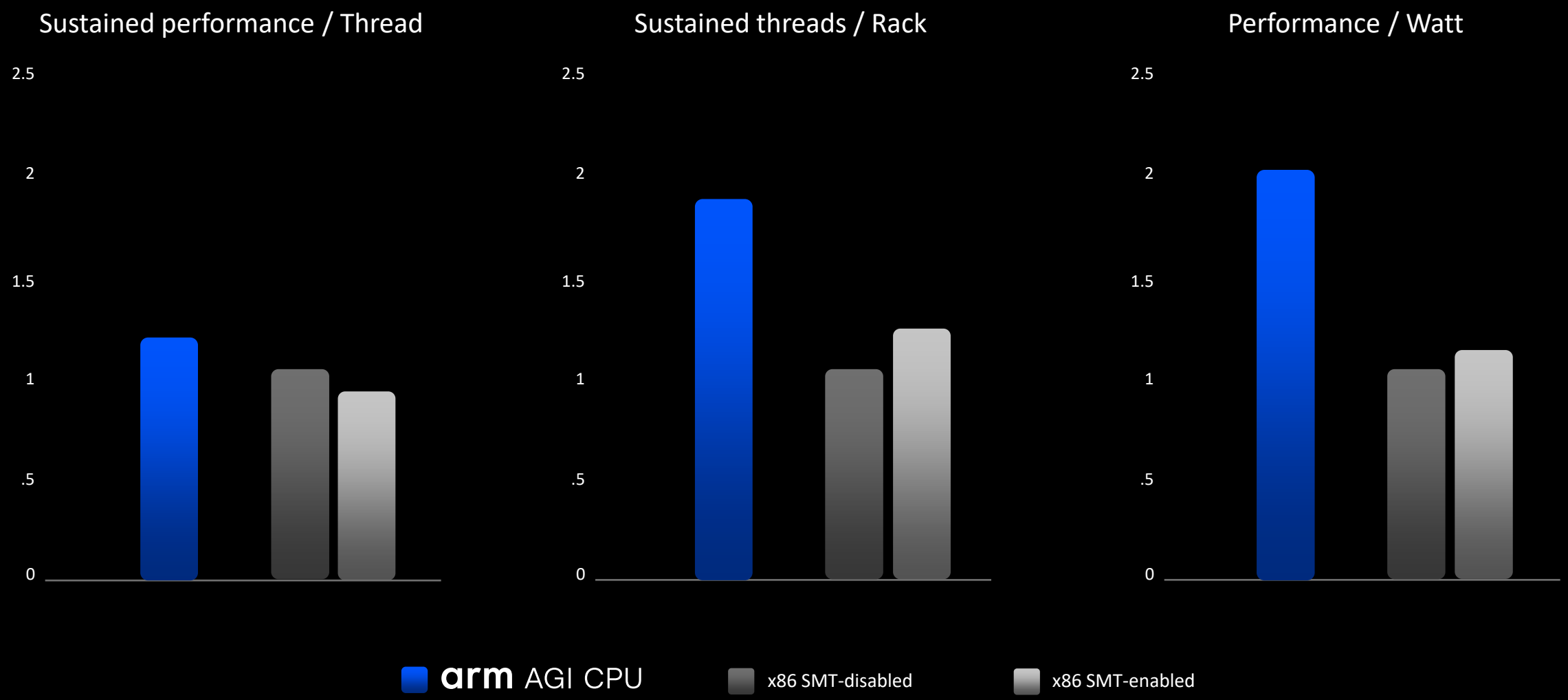
Based on estimates.

Performance. Scale. Efficiency.



Based on estimates.

Performance. Scale. Efficiency.



Based on estimates.

arm AGI CPU

World’s most efficient agentic CPU

Performance

Scale

Efficiency

More than 2x performance per rack on Arm

Arm AGI CPU

36kW rack

30x 1U servers

8,160 CPU cores

VS x86 CPU

36kW rack

17x 2U servers

4,352 CPU cores

Based on estimates.

More than 2x performance per rack delivers

Up to

$10B

CAPEX savings

vs x86 CPU

1GW capacity

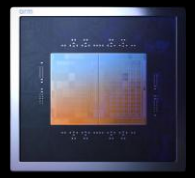
Based on estimates.

Committed to future products

Silicon

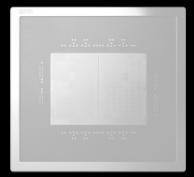
Neoverse IP + CSS

arm AGI CPU

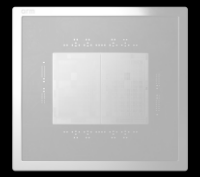


Best in class performance, scale & efficiency

arm AGI CPU 2



arm AGI CPU 3



CSS V3

CSS V4

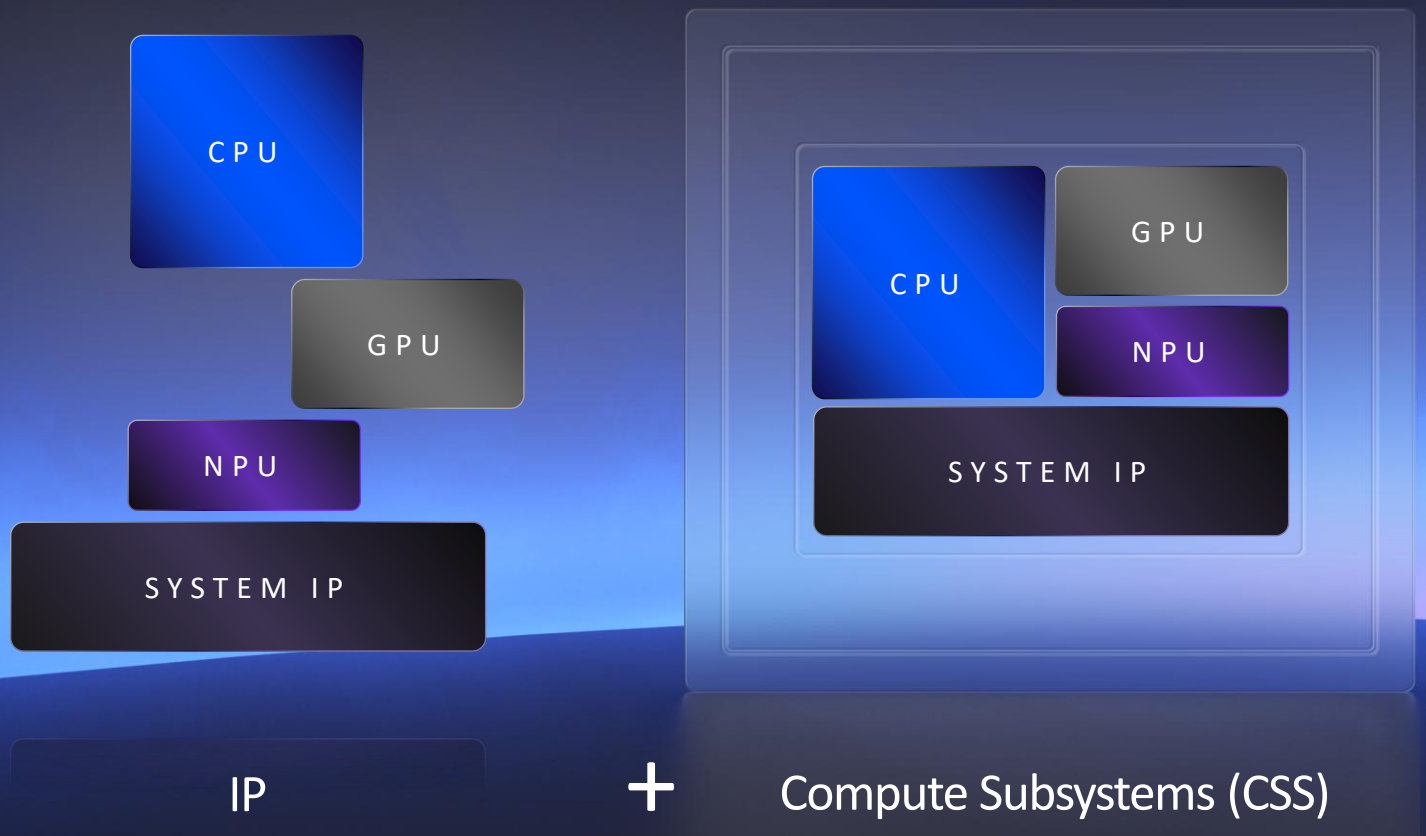
CSS V5

Now

2027

Future

Timelines depict silicon sample availability.



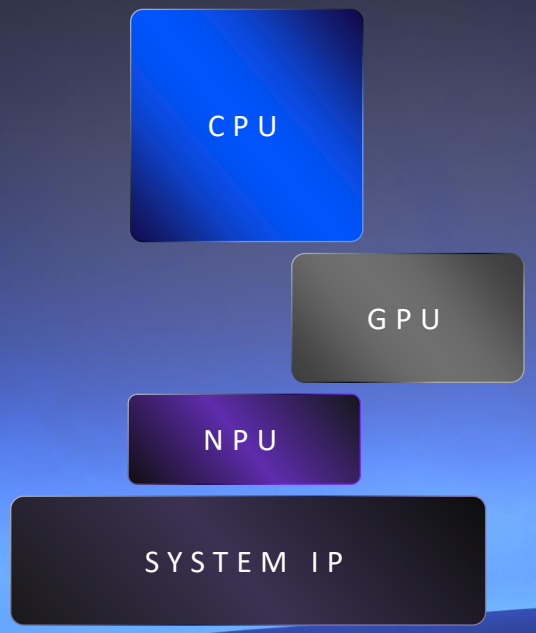
IP + Compute Subsystems (CSS)

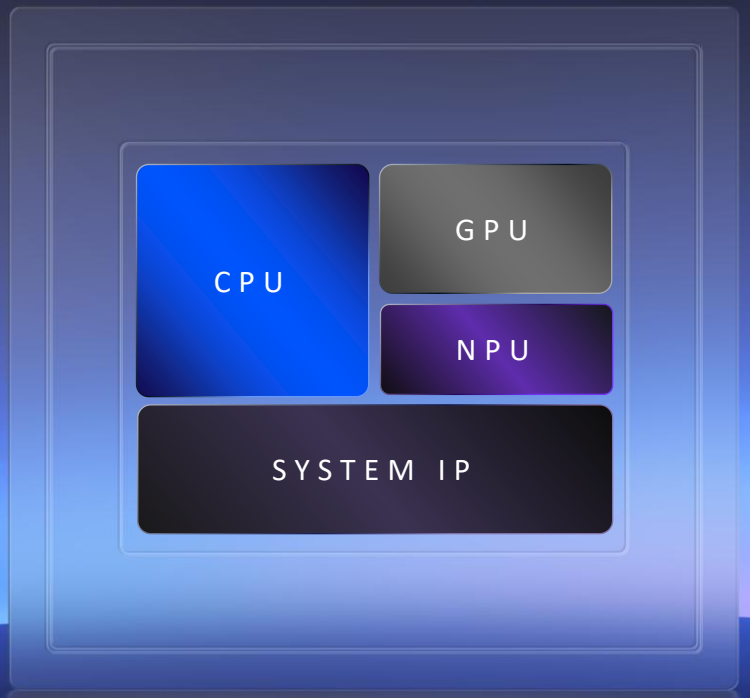
~$3B TAM

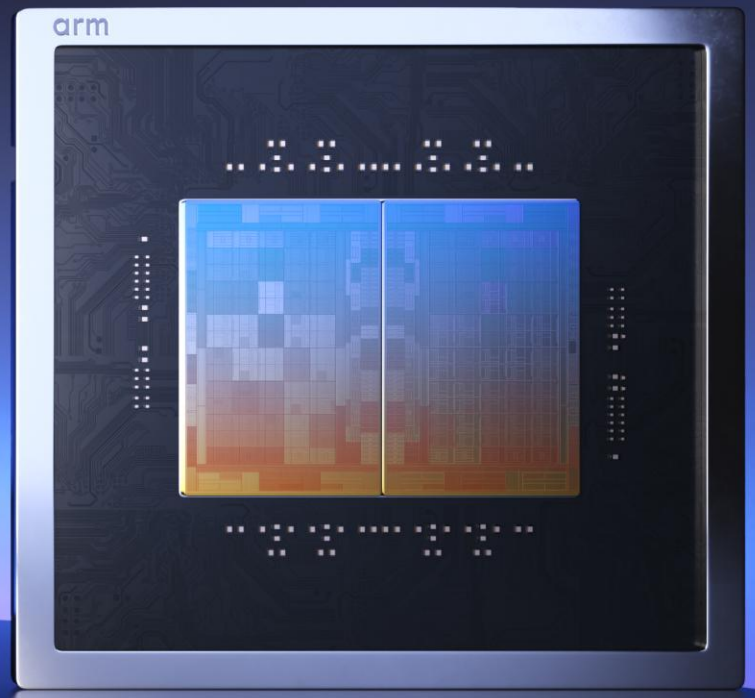
IP + Compute Subsystems (CSS)

Based on estimates.

arm AGI CPU







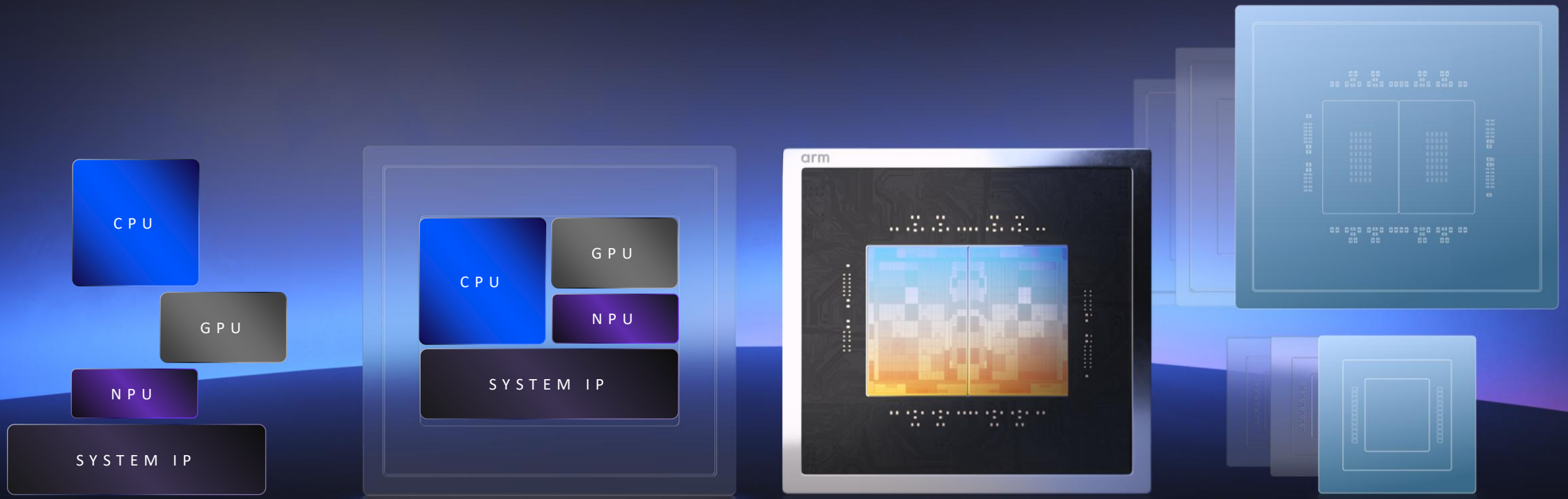
IP + Compute Subsystems (CSS) + First production silicon chip

>$100B TAM

By 2030

IP + Compute Subsystems (CSS) + First production silicon chip

Based on estimates.



IP + Compute Subsystems (CSS) + First production silicon chip

We are not stopping here

>$1T TAM

IP + Compute Subsystems (CSS) + First production silicon chip

By 2030

Based on estimates.

ADVANTEST
altera
ami
Amkor Technology
ARISTA
ASE HOLDINGS
ASRock
aws
BROADCOM
cadence
Canonical
cerebras
CISCO
CLOUDFLARE
databricks
f5
FURIOSA
GitHub
Google
Hugging Face
intel foundry services
Lenovo
MARVELL
MEDIATEK
Meta
micron
Microsoft
MongoDB
NVIDIA
NXP
OpenAI
OPEN Compute Project
ORACLE
POSITRON
Quanta Computer
rebellions_
Red Hat
Redis
SAMSUNG
SAP
SIEMENS
SK telecom
SK hynix
snowflake
socionext
STATSChipPAC
SUPERMICRO
SUSE
SYNOPSYS
tsmc
University of BRISTOL

The Arm trademarks featured in this presentation are registered trademarks or trademarks of Arm Limited (or its subsidiaries) in the US and/or elsewhere. All rights reserved. All other marks featured may be trademarks of their respective owners.

www.arm.com/company/policies/trademarks