

XPU Based Cloud Native Server:

Architecture, Implementation & Applications

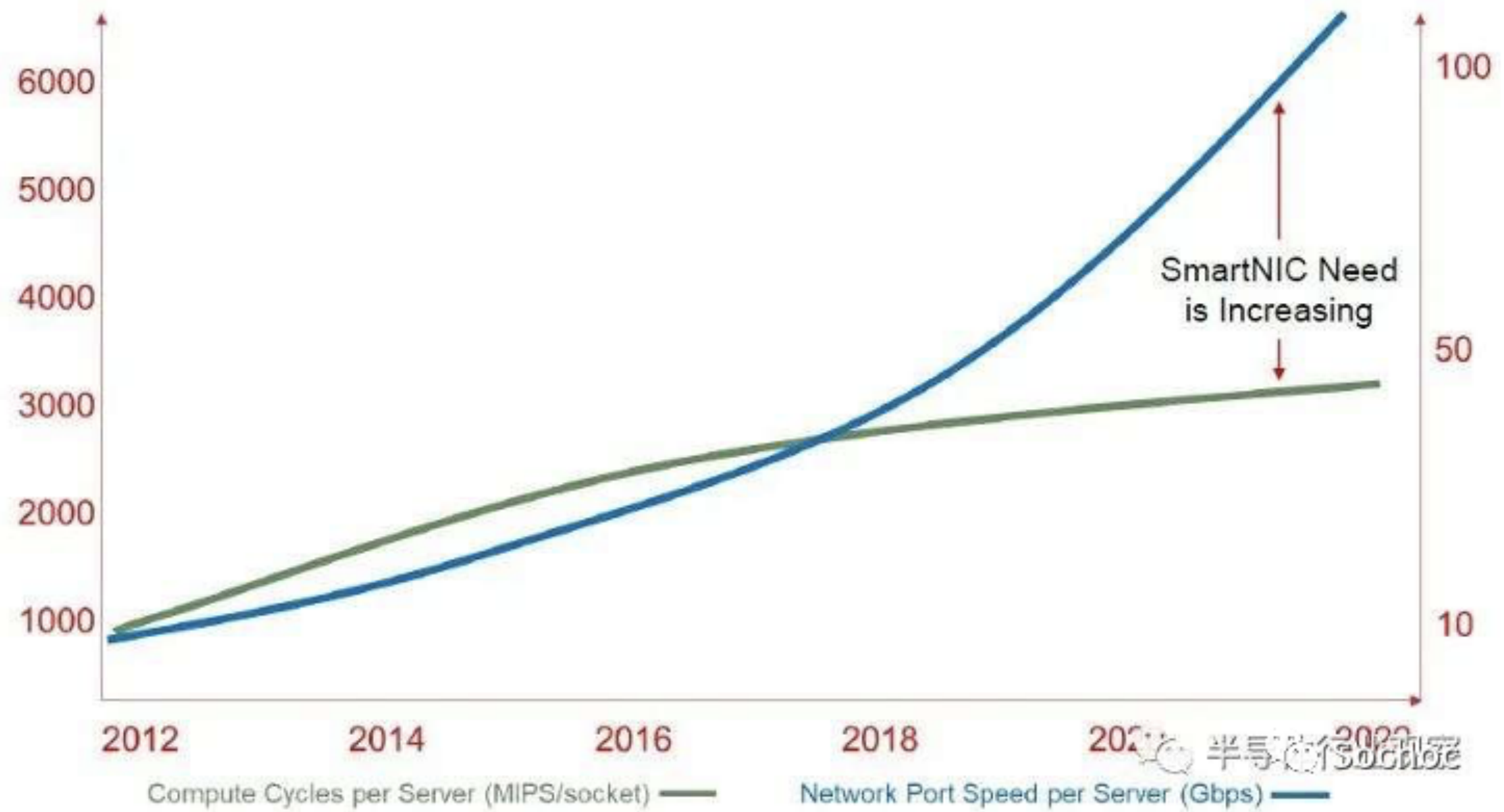
Dr. Fu Li (LEO)

li@socnoc.ai

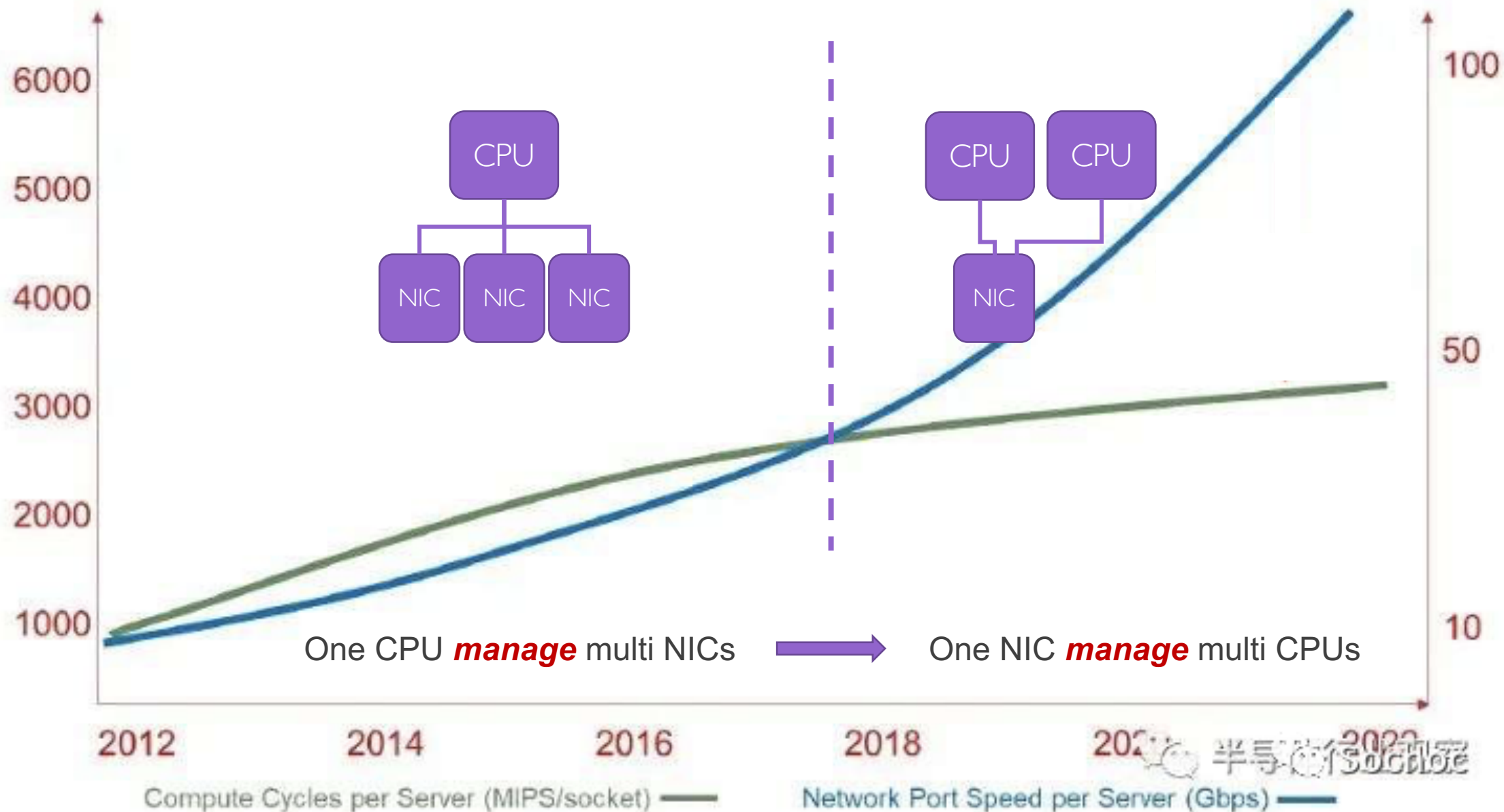
Socnoc AI Inc. 合肥边缘智芯科技有限公司



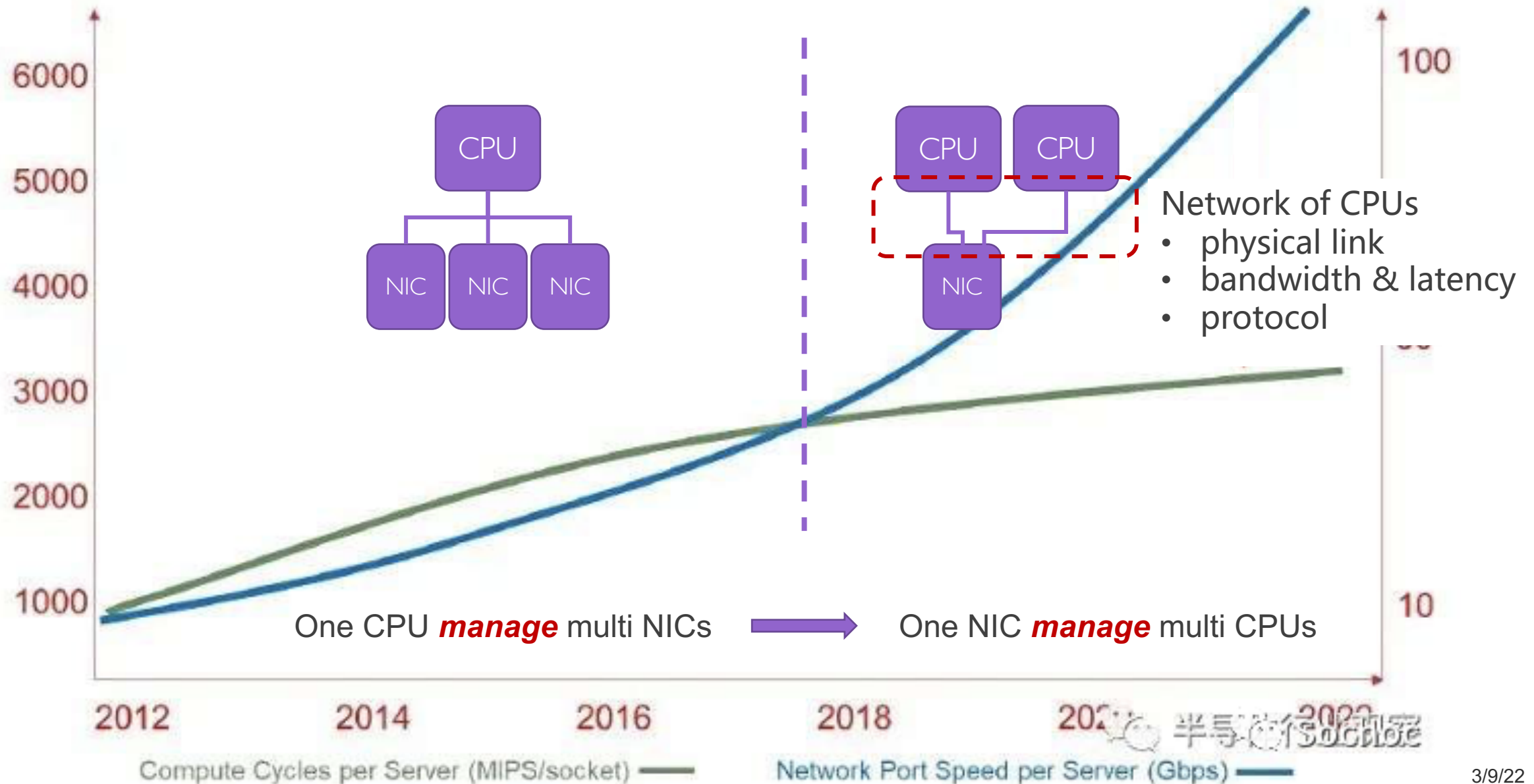
Problem to Address?



Problem Revisit: Architecture Shift

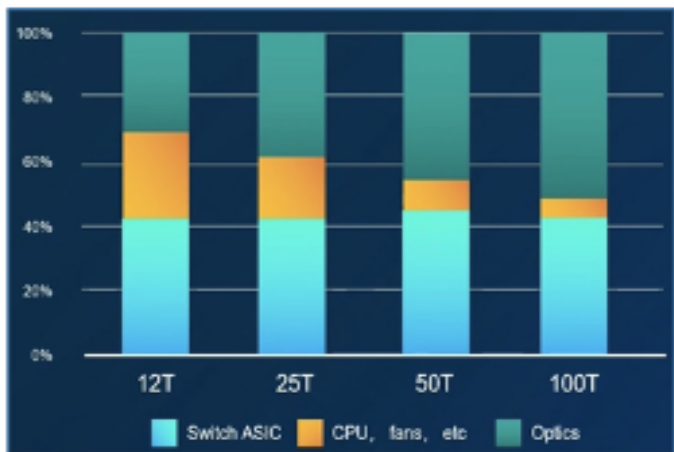


Problem Revisit: Clustered CPU on Board

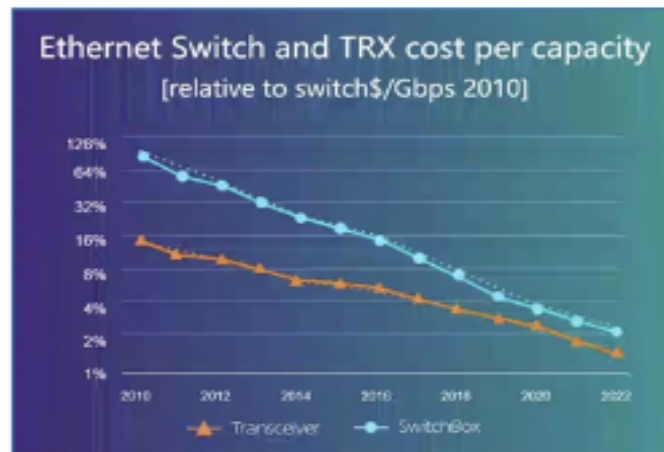


Real Problem Computing Cluster Faced!

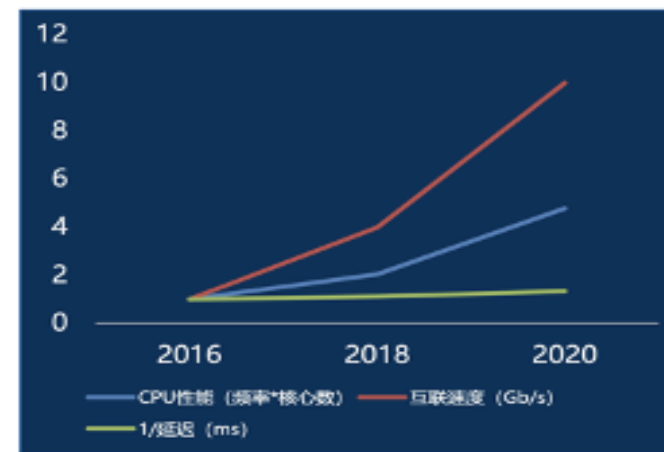
› Optics are too expensive both in **power** and **cost**



光模块功耗占比已经超过50%



光模块成本已经超过通道成本



延迟降低进展缓慢



版权所有@边缘智芯



About SOCNOC: Kill Optics in Short Distance!

Socnoc AI (<https://www.socnoc.ai>) is the startup aiming to provide low-cost, green and high-performance networking solutions to smart clusters and edge cloud computing. Socnoc employ PCIe and CXL based interface with its own protocol to build the best infrastructure for networked and composable systems.

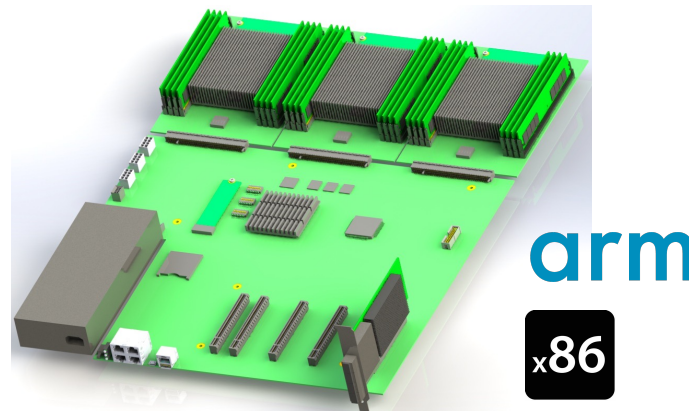
Our Goal:

- › **Reduce the network cost to \$3 per Gbps host-to-host**
- › **Eliminate Optics and moving parts in Edge Clusters**

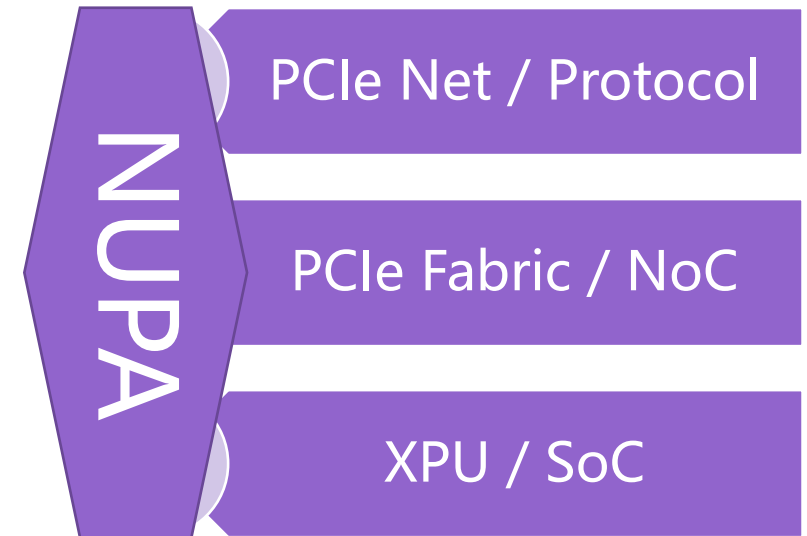
Network Unified Protocol Architecture



XPU



XPU Cloud Native Server



Data Fabric Landscape

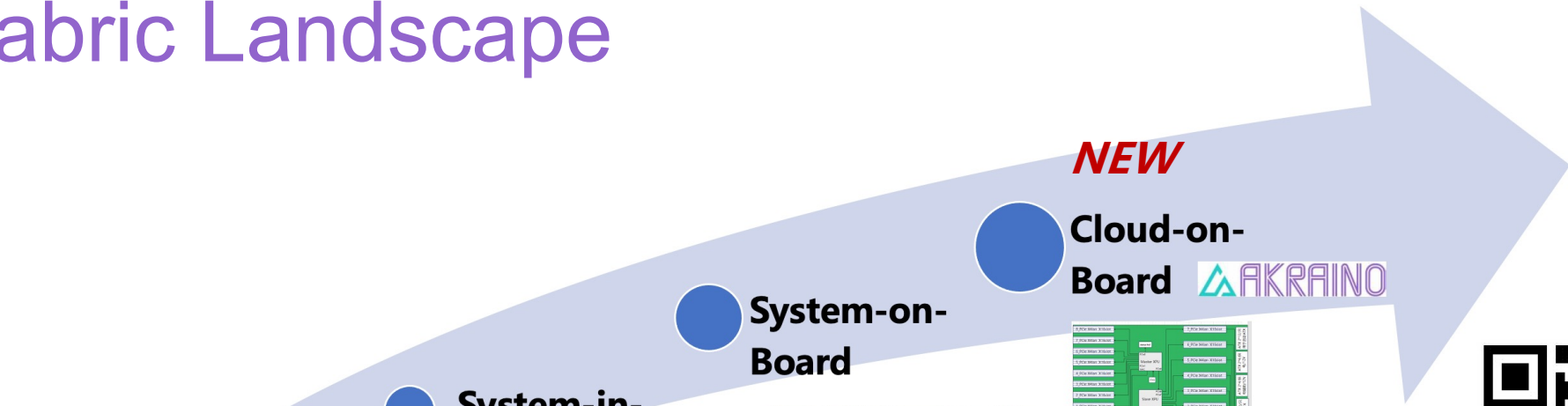
TCP/IP



PCIe/CXL
cxl.io/cxl.cache



UCIe

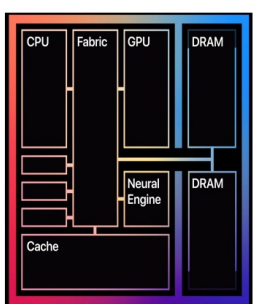


● Chiplet



I/O module

● System-in-Package



Interposer Fabric

In-Package Links

● System-on-Board

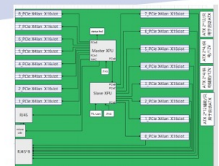


PCIe Bus based link

Off-Package Links

NEW

● Cloud-on-Board

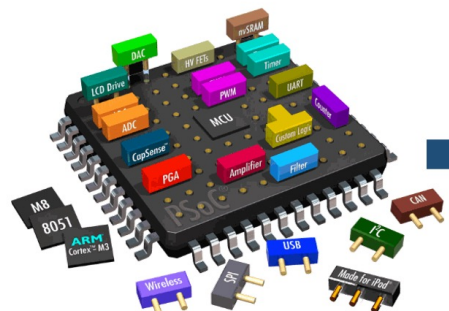


PCIe Net based Fabric

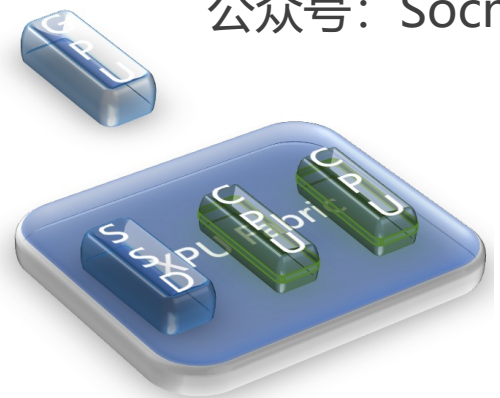


公众号: Socnoc

XPU: Center of Data Fabric

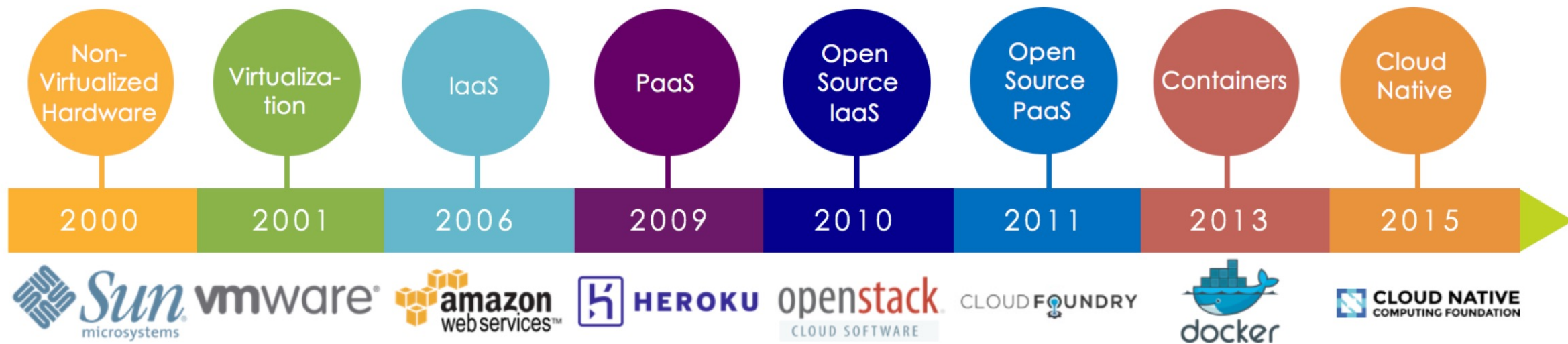


In Package Technology



Off Package Technology

Application Evolution: From Metal to API Ready!



Metal Servers

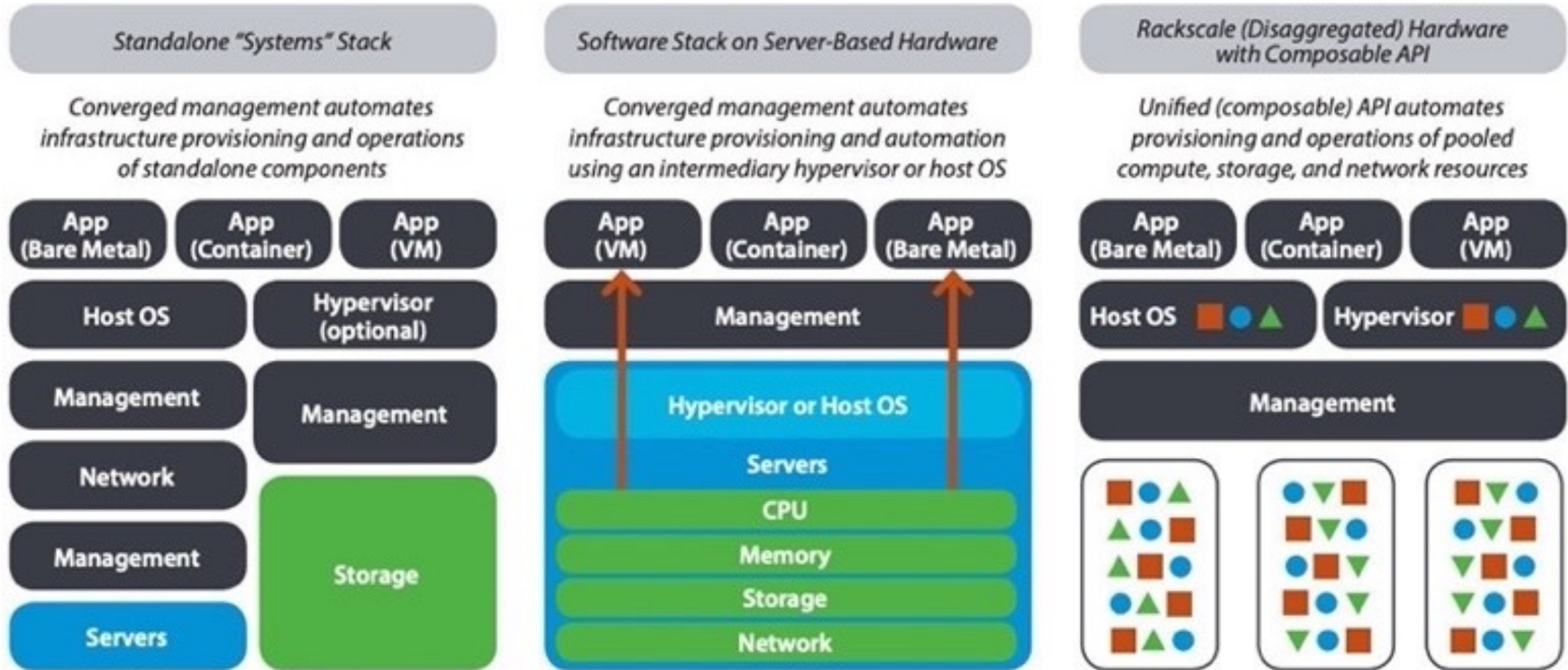
*Hardware-driven
BUS-driven*



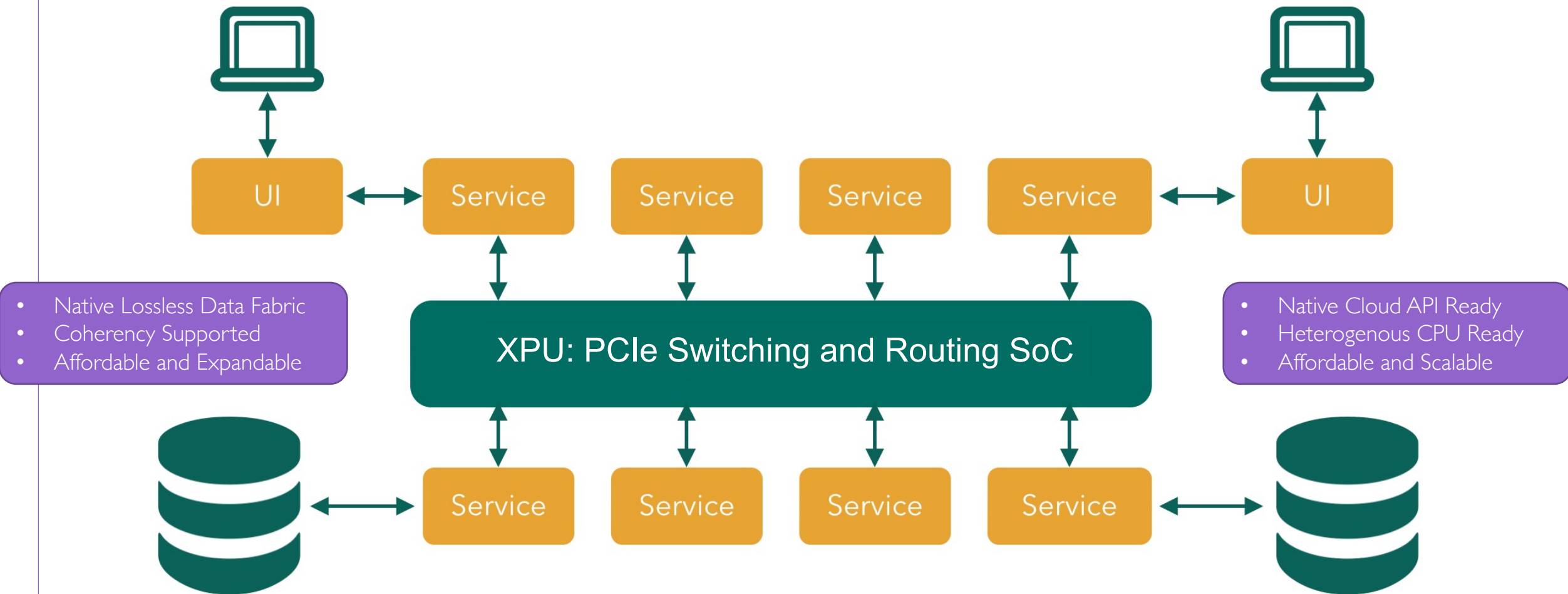
Serverless API

*Software-driven
TCP/IP-driven*

Infrastructure Trends: Turn Metal into API

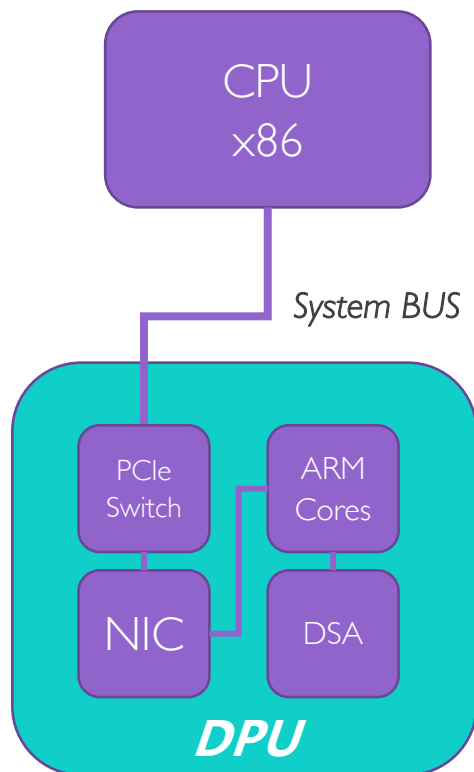


Our Solution: Bridge BUS and TCP/IP!

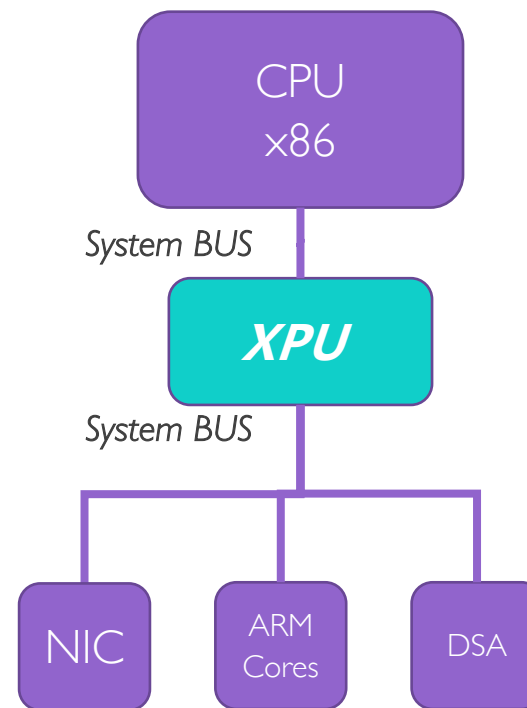


XPU vs DPU ?

› XPU is purely a datapath management data processor



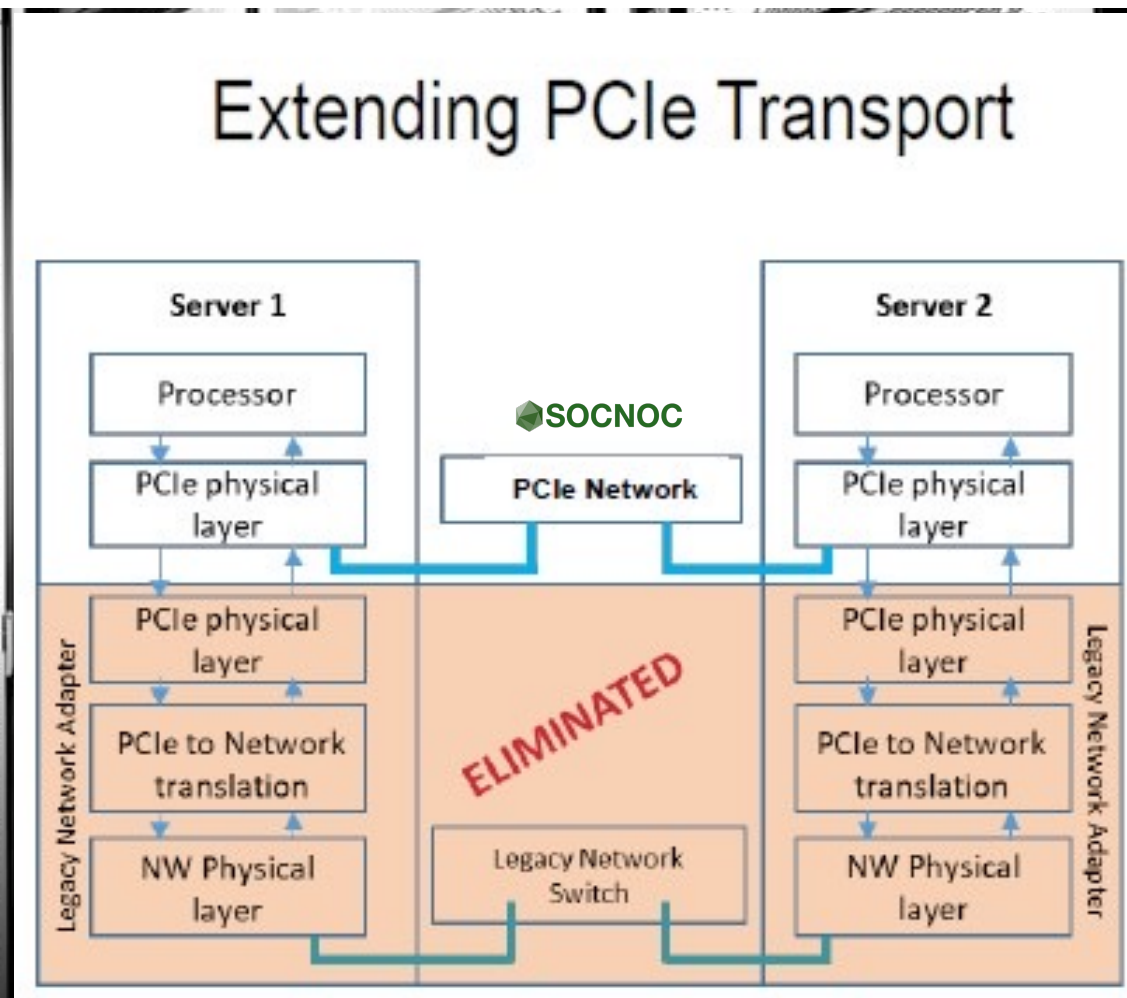
VS



Disaggregated Architecture

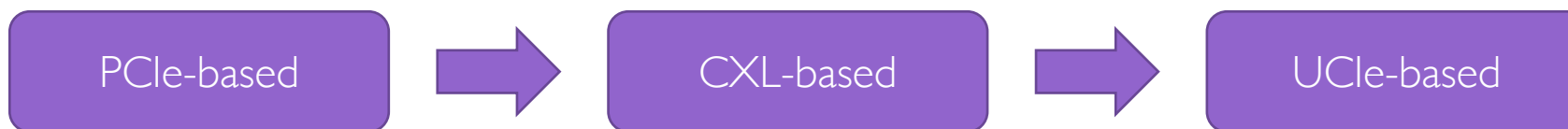
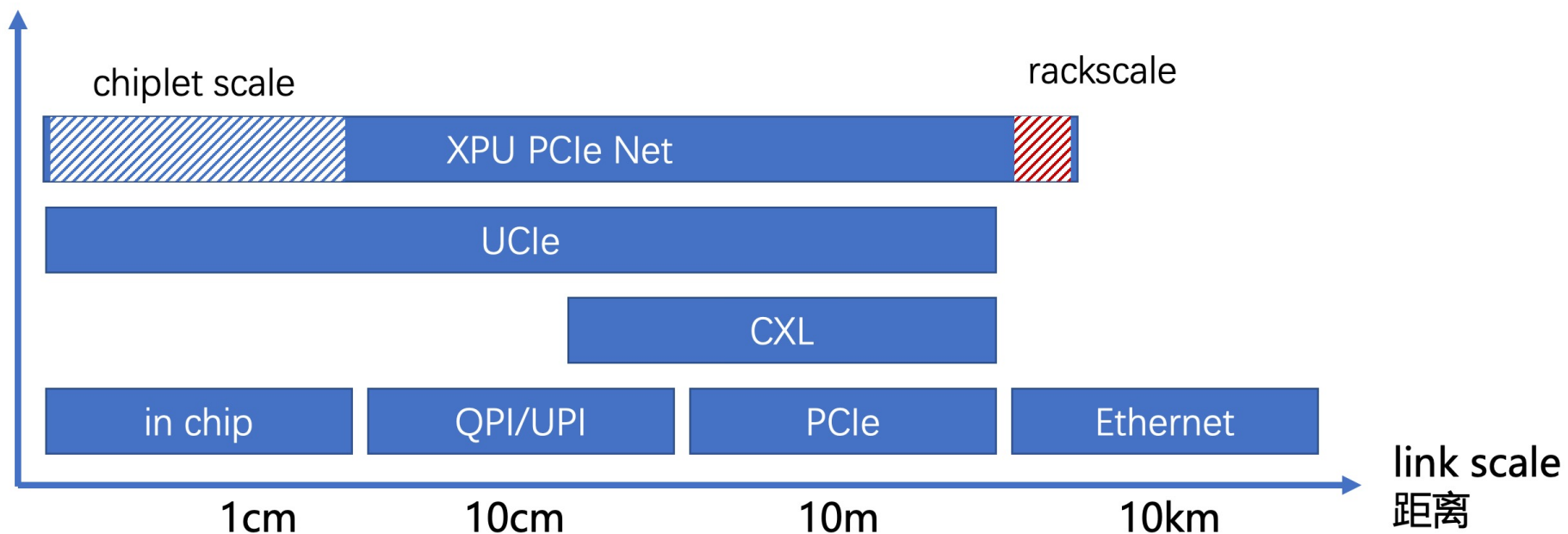
- Same BUS
- Same protocol
- Cost-effective
- Universal
- Expandable
- Scalable

Architecture Advantage: Less is More



<https://www.nextplatform.com/2019/10/02/a-new-twist-on-pci-express-switching-for-the-datacenter/>

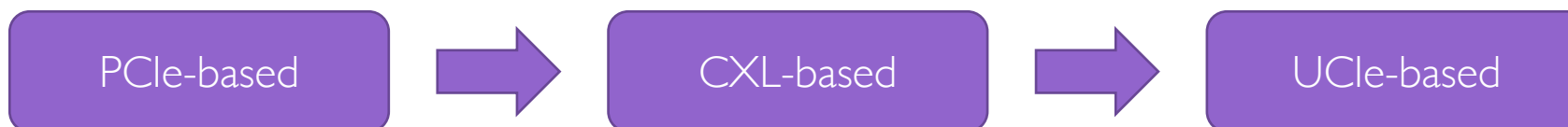
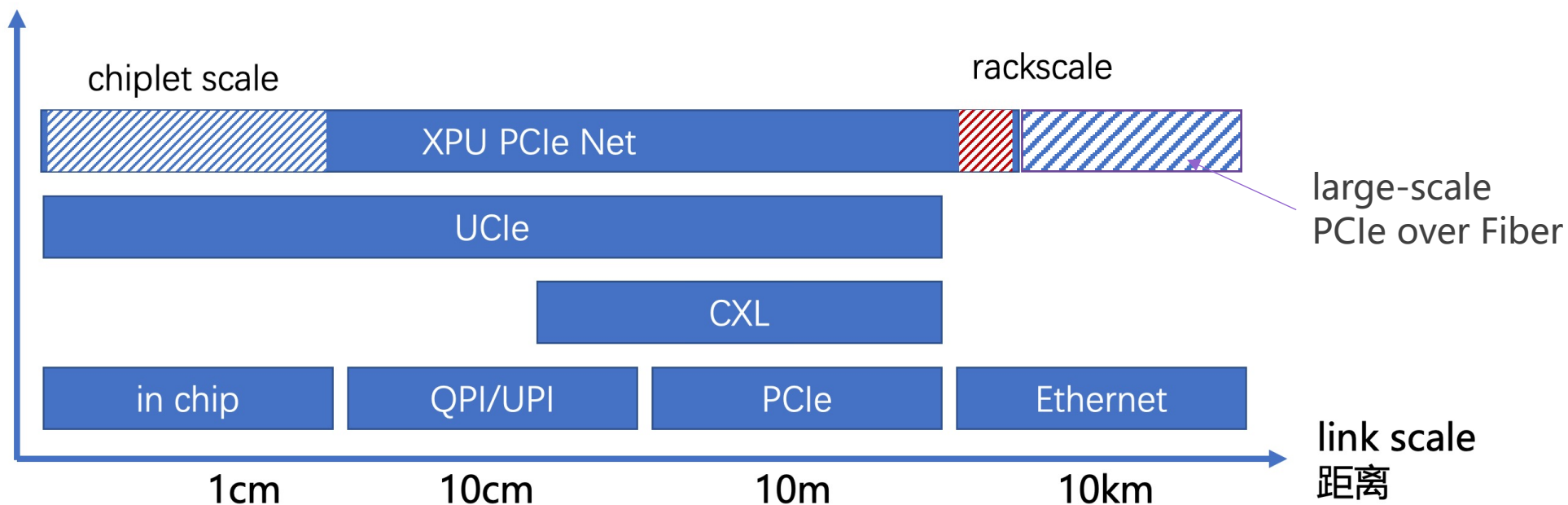
Roadmap and Ecosystem



arm x86



Roadmap and Ecosystem



arm x86

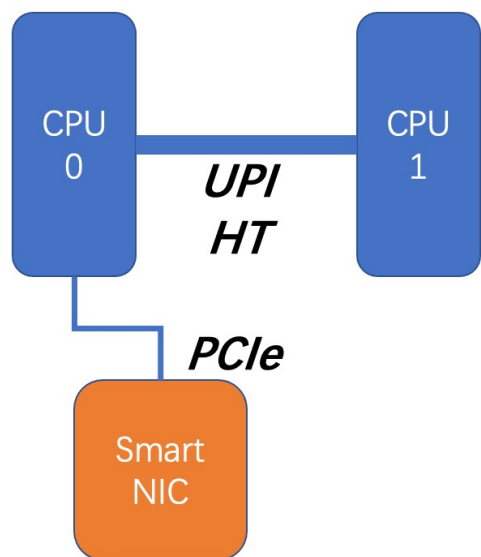


XPU Implementation & Applications



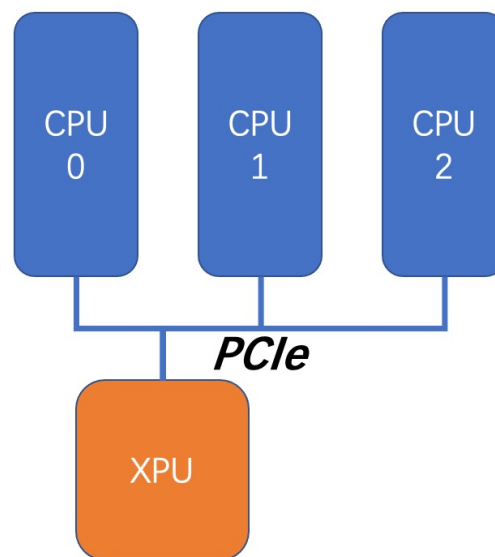
XPU Server: All in PCIe!

- Transform UPI/HT based multi-CPU Server into PCIe connected Server!

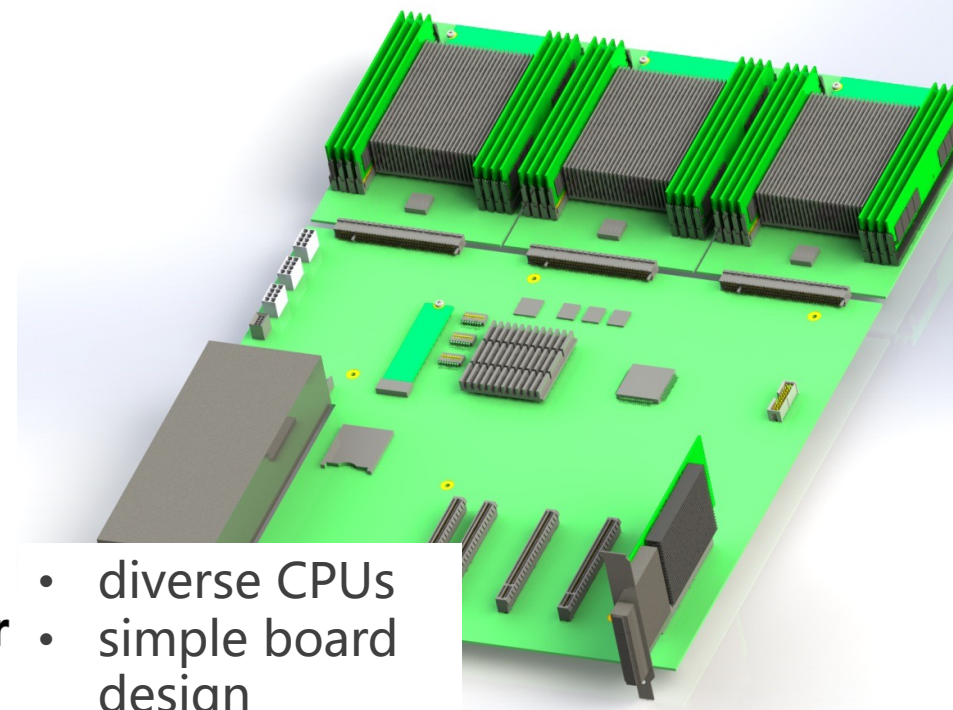


Dual CPU Server

VS

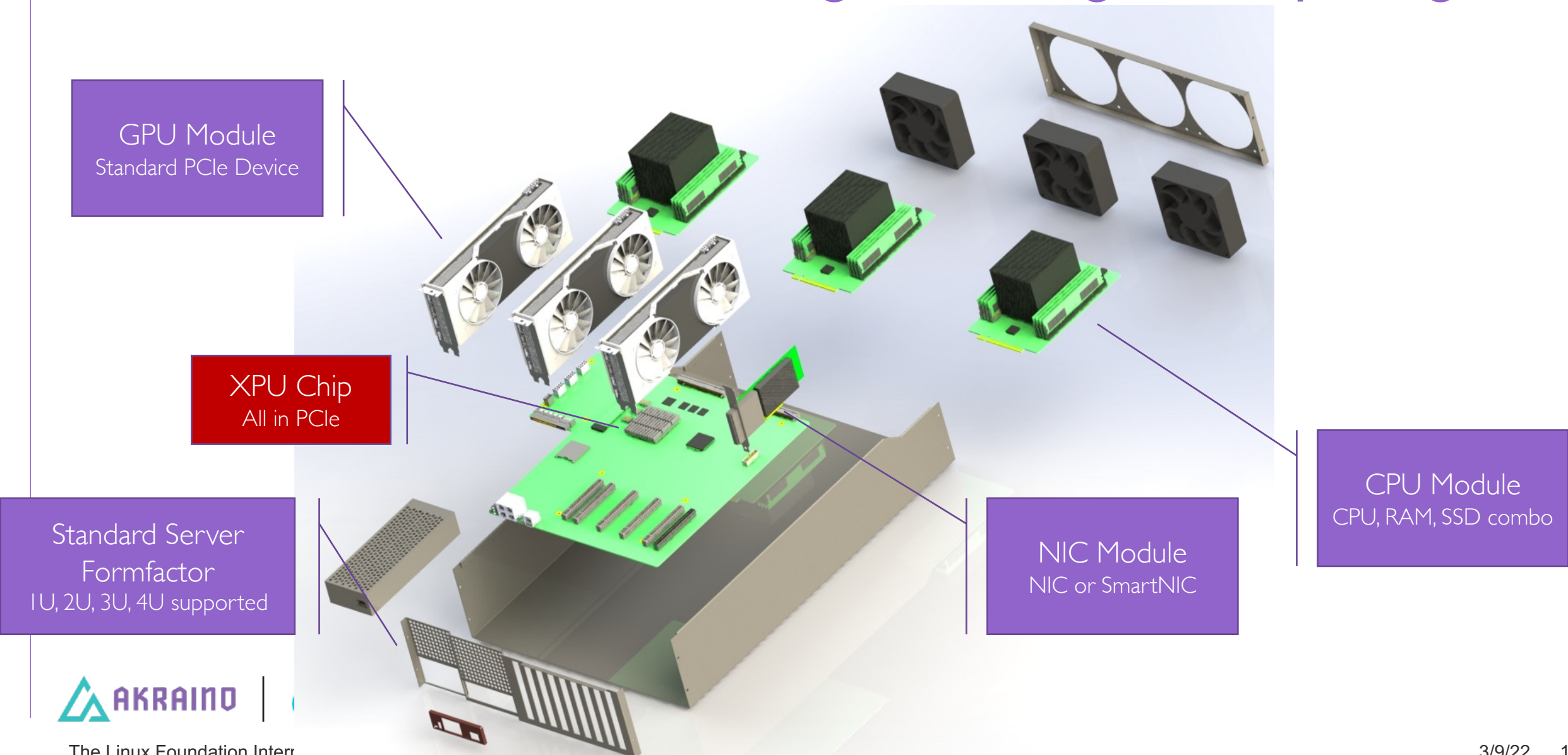


XPU-centered Server

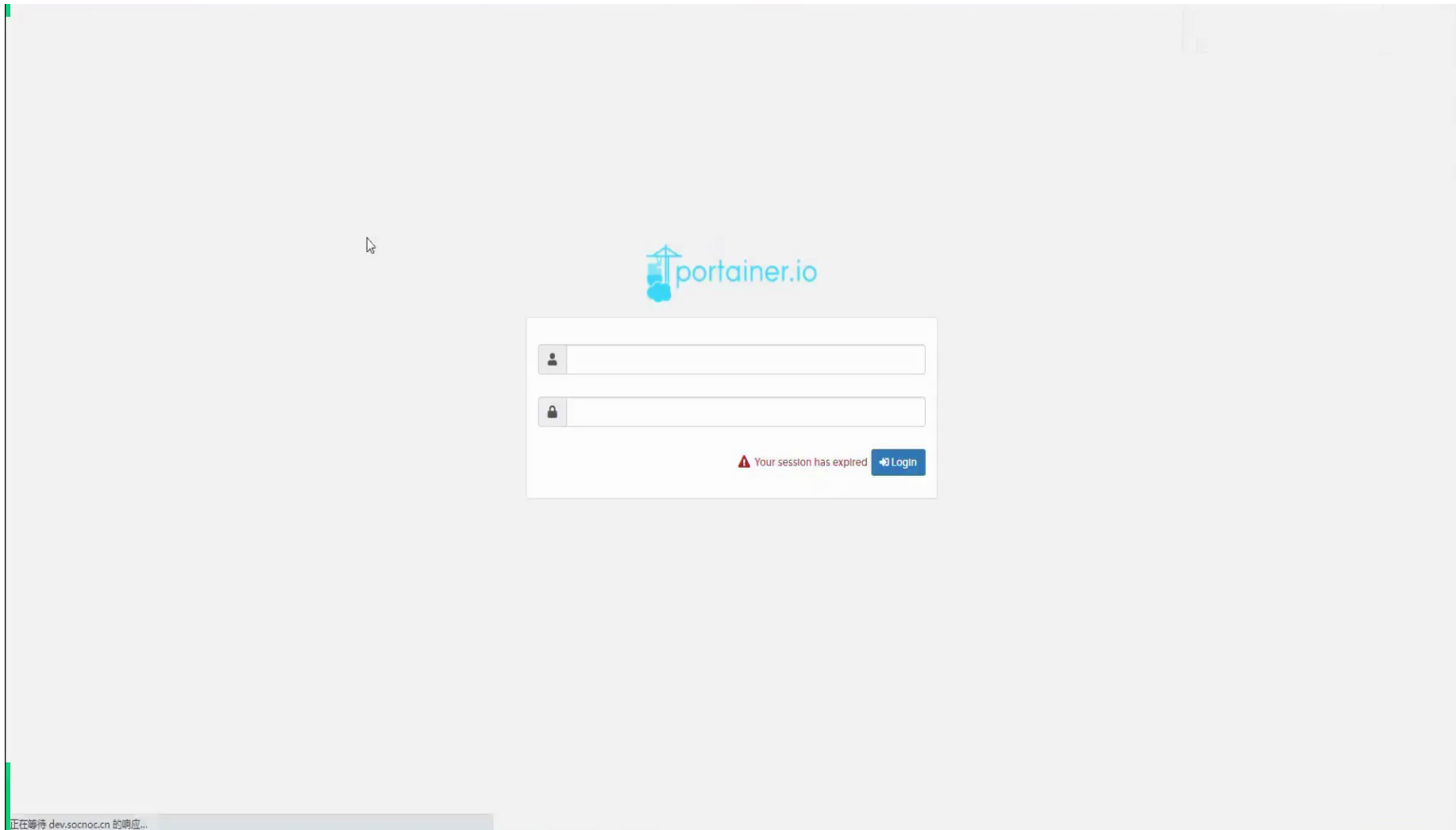


- diverse CPUs
- simple board design
- expendable
- less expensive
- HA-available

Cloud Native Server for Integrated Edge Computing



Online Demo



Thank you !



www.socnoc.ai