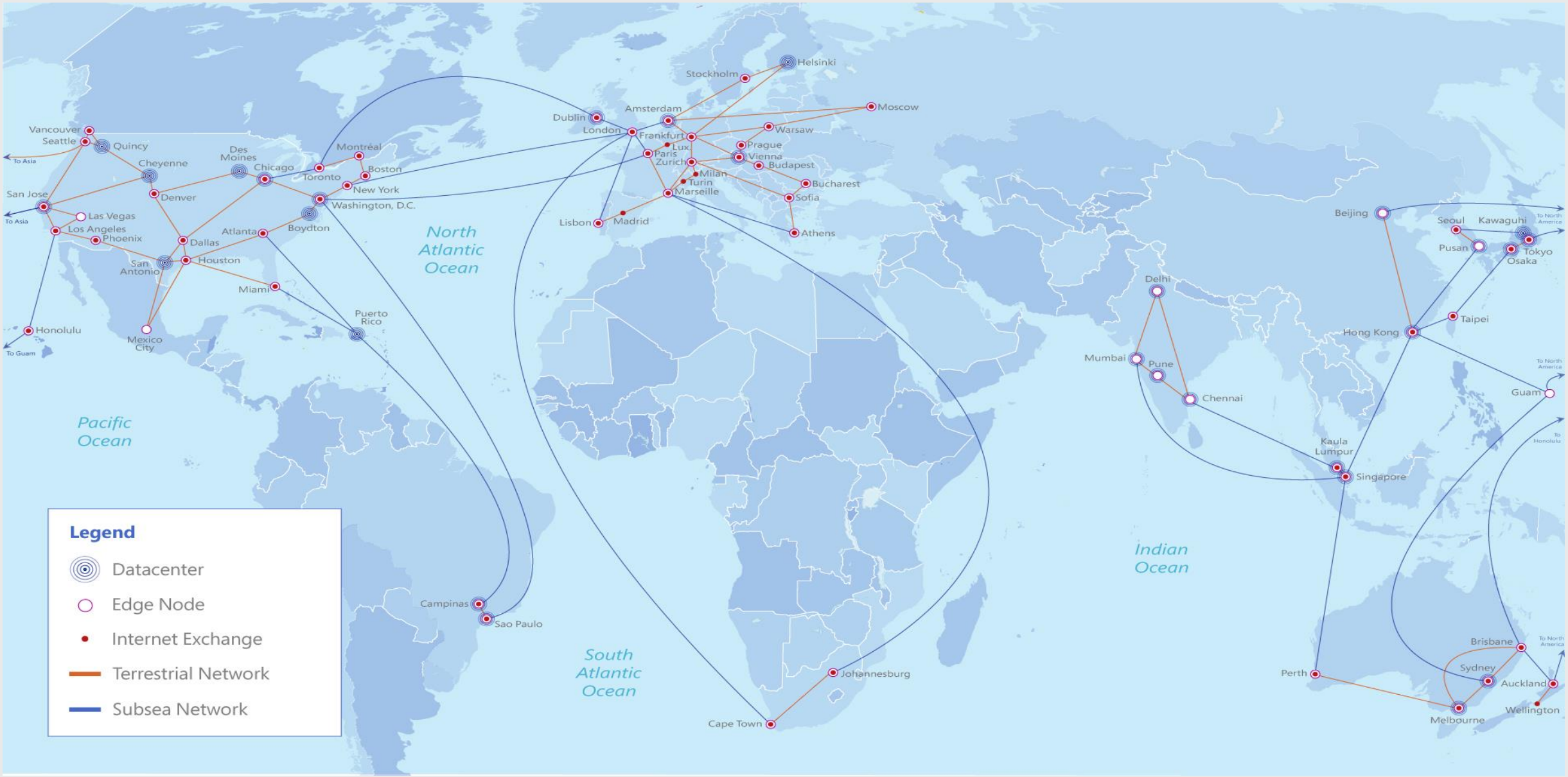
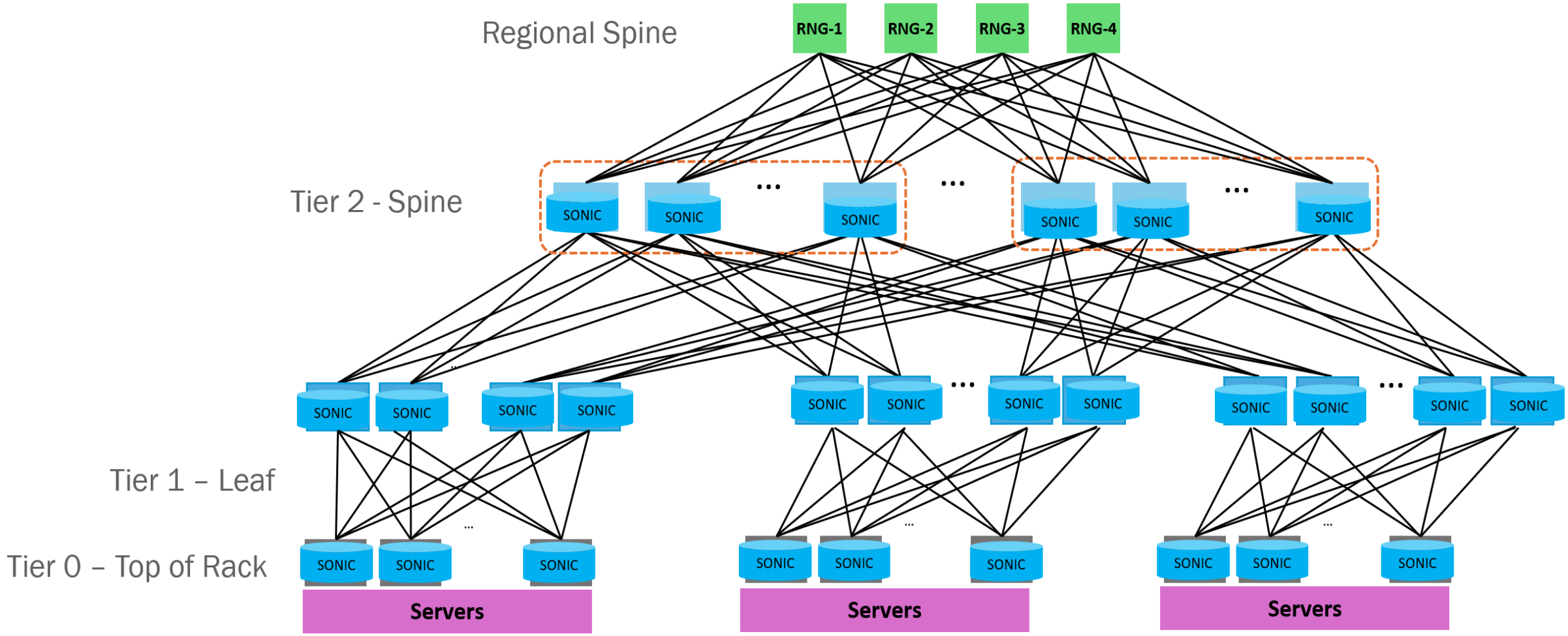
SONiC for AI with SRv6

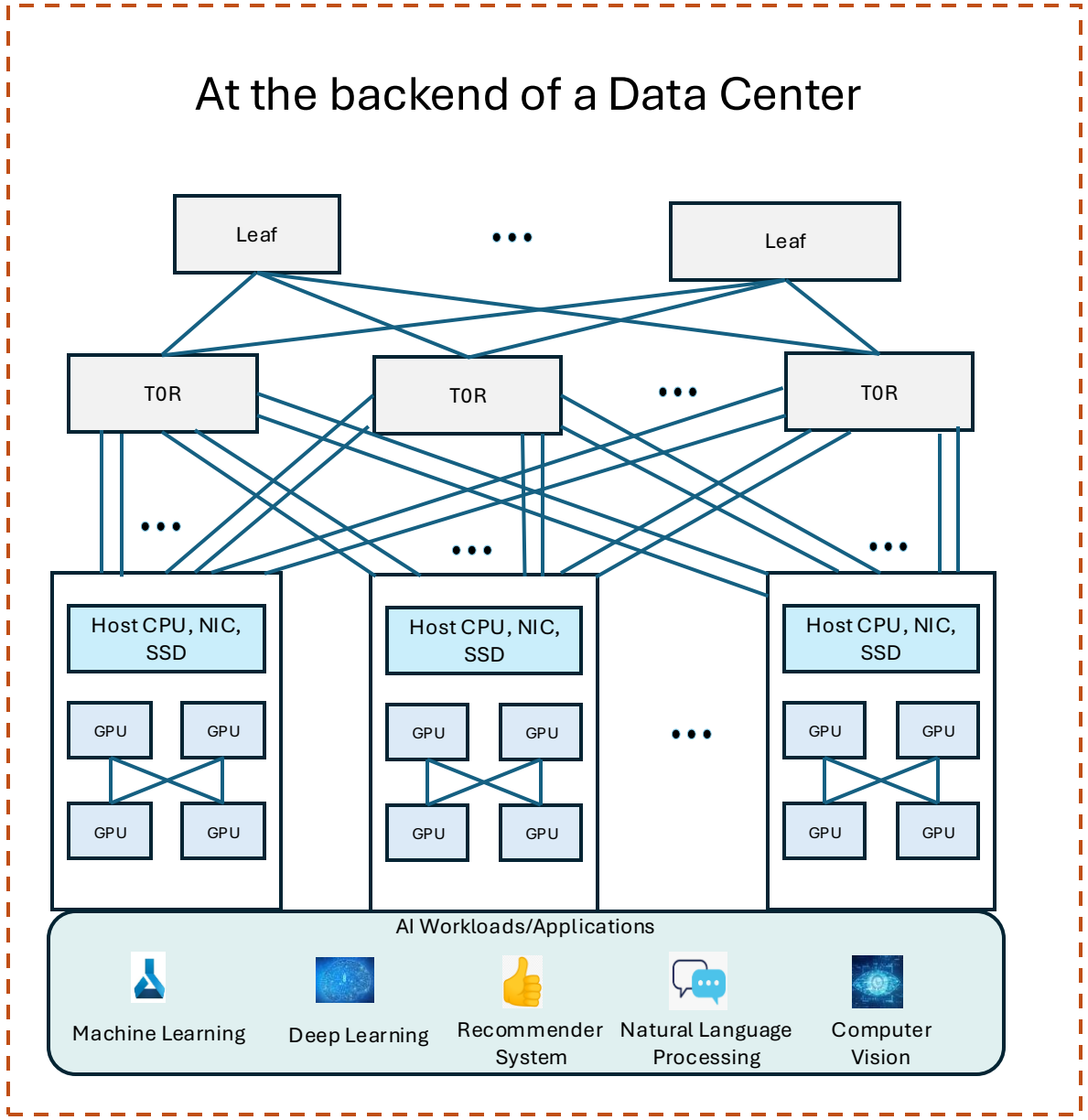
Rita Hui
Principal Software Manager, Microsoft

Microsoft Global Cloud Network



SONiC Is Powering Cloud At Scale



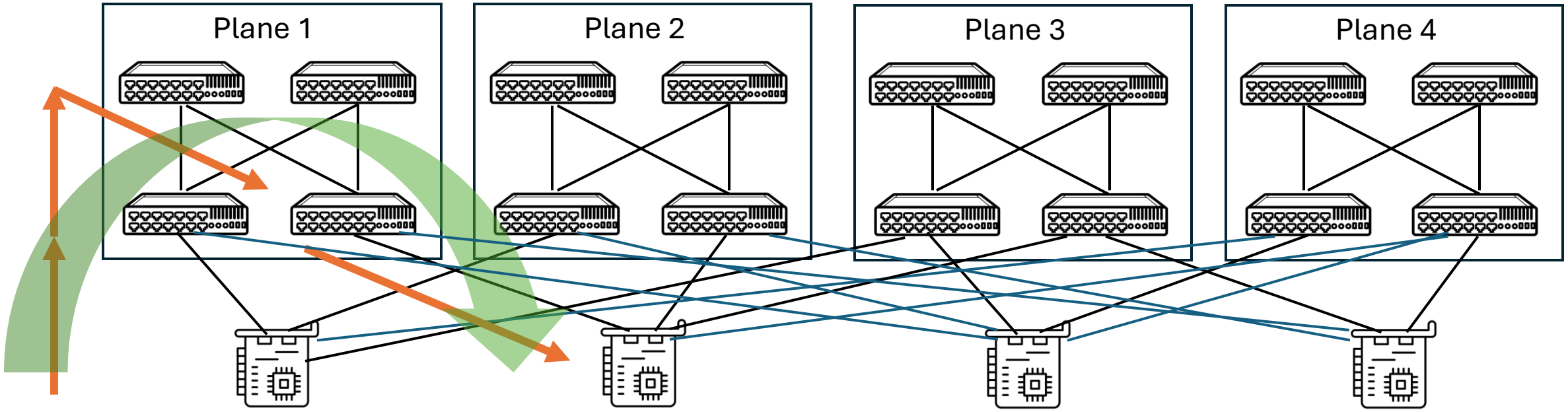


Artificial Intelligence in the Cloud

Raising the Bar for Hyperscale Datacenter Networks

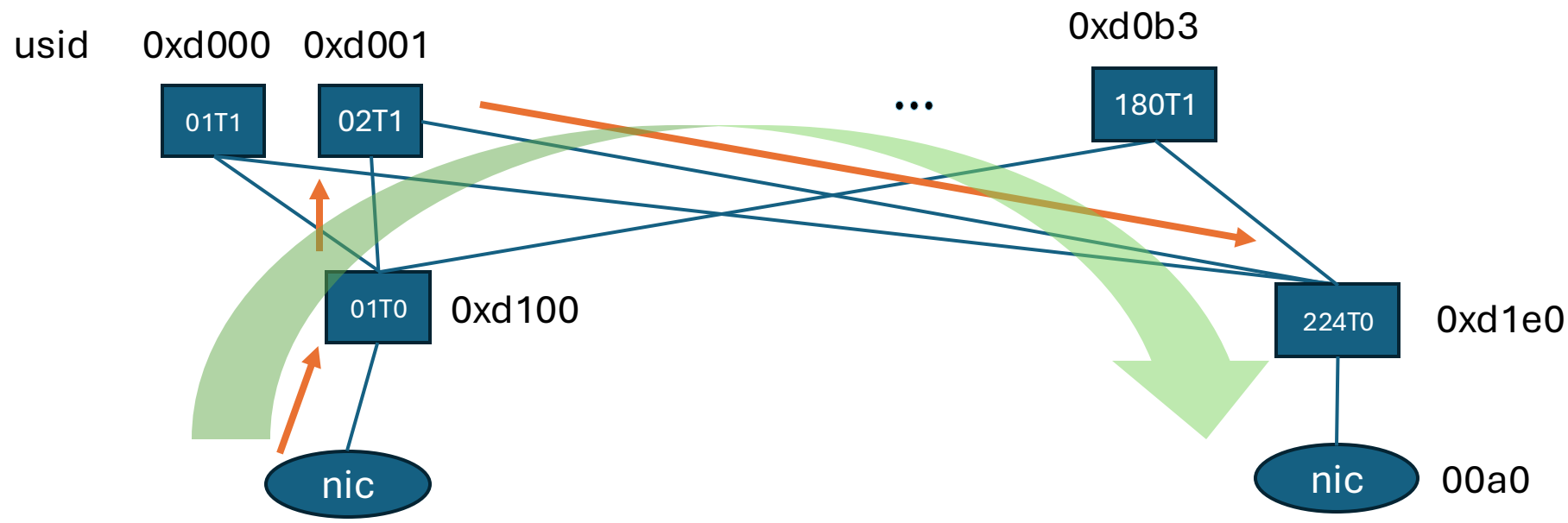
- Long lasting and large flows of training data
- Large bursts of data sent synchronously
- Long training time demands reliable networks
- Retries of failed jobs increased costs
  - Efficient traffic management, monitoring and visibility
- AI applications need fast processing and responses
  - Lossless traffic with low latency
- Traffic using RoCEv2 has low entropy for ECMP
  - Traffic engineering technology for AI backend network

SRv6 for 2-layer topology



- Source Routing and Network Control
- Enhanced Traffic Management
- Scalable and Flexible
- Reliable and Redundancy

SRv6 for 2-layer topology using uSID



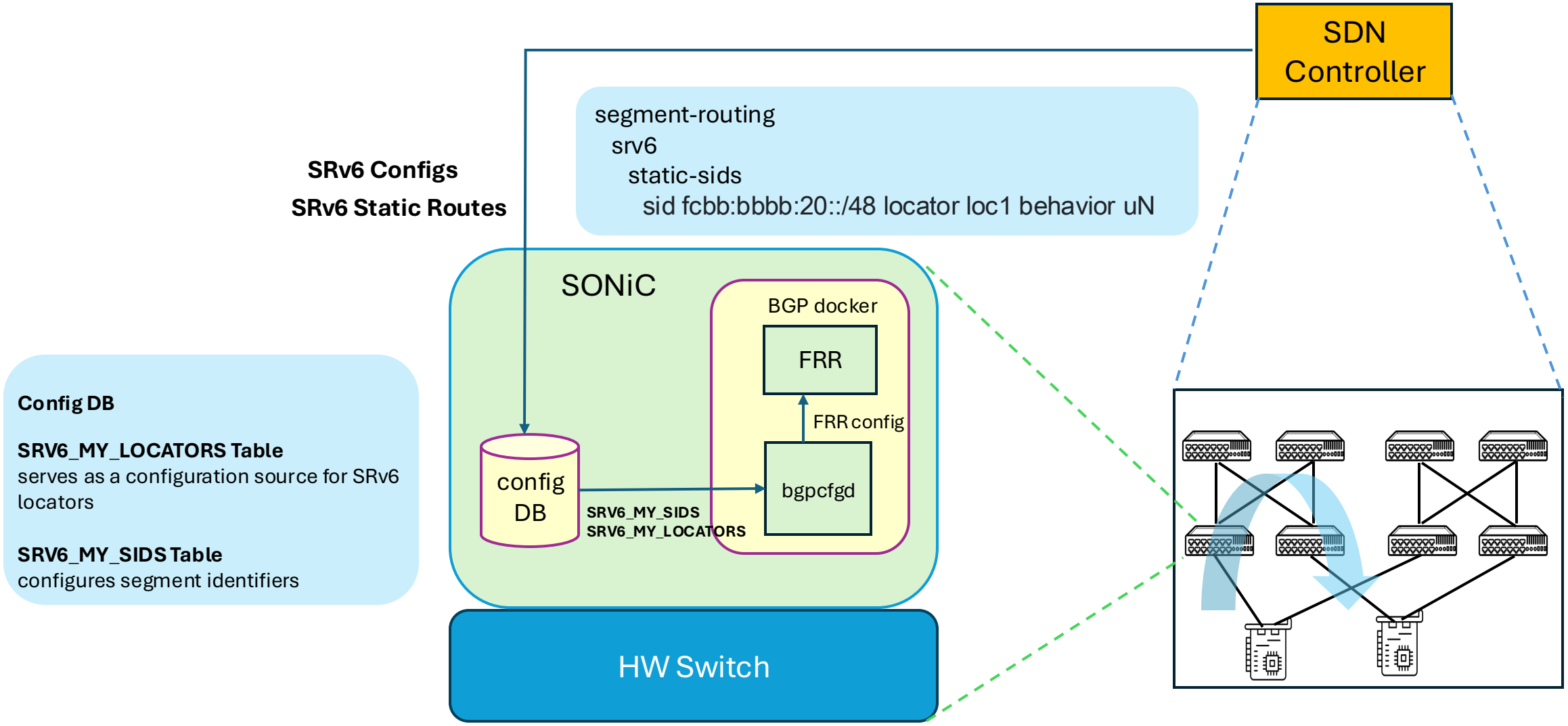
SRv6 packet from Nic

DstIPv6 *fcbb:bbbb:d100:d001:d1e0:00a0::*

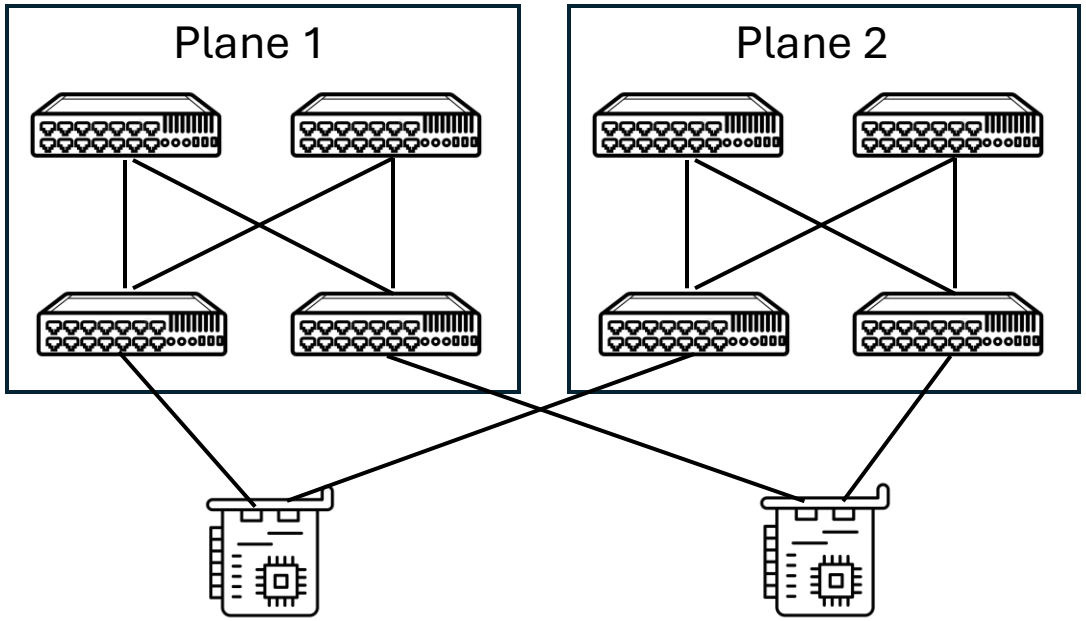
32-block

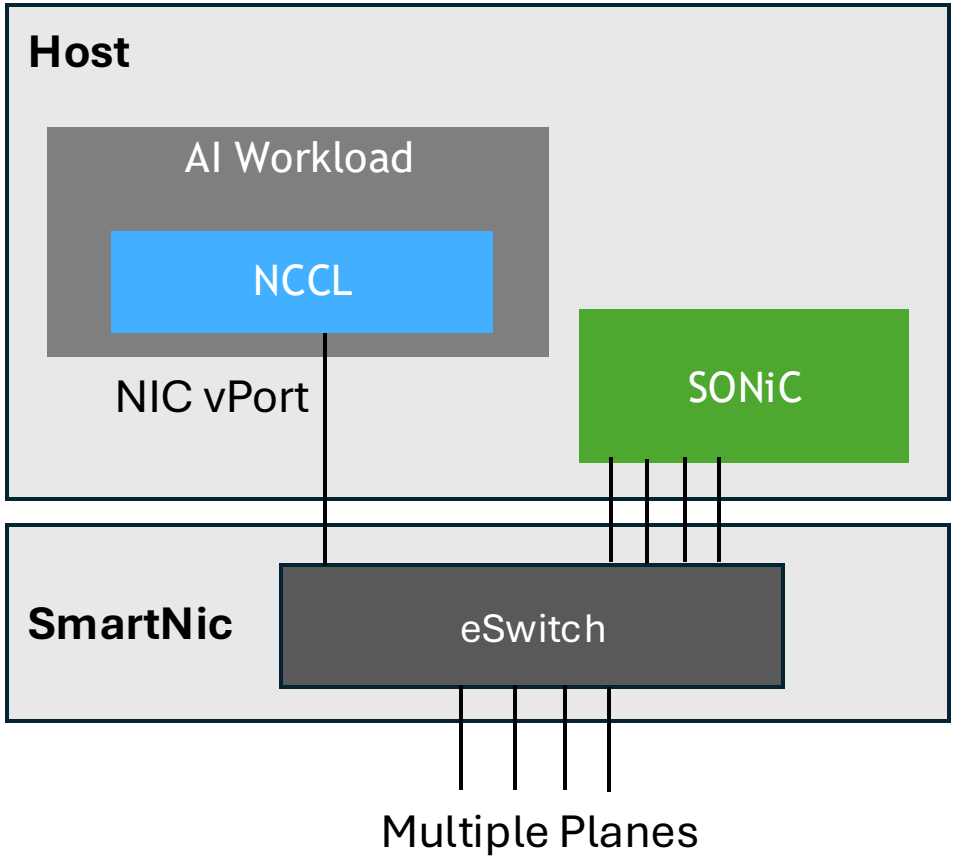
1st hop 2nd hop 3rd hop 4th Hop

Static uSID with SRv6 in SONiC



End-to-End Control





The NIC/DPU has many switch functionalities in multi-plane networks

Inviting contributions to SONiC community in all areas

- SONiC/SAI
- Hardware platforms
- Testing and tooling
- Download, test, deploy!

- Project Wiki with latest specification : https://sonicfoundation.dev/
- Source Code: https://github.com/sonic-net/SONiC/blob/master/sourcecode.md
- Becoming a contributor: https://github.com/sonic-net/SONiC/wiki/Becoming-a-contributor
- Mailing list: https://lists.sonicfoundation.dev/g/sonic-dev
- SONiC Community meeting: https://sonic-net.github.io/SONiC/Calendar.html
- SONiC Routing Workgroup: sonic-wg-routing@lists.sonicfoundation.dev | Home

Thank You