GP4L labs: Evaluation of High Performance Software Dataplane

GNA-G Community VC1 Q1-2025 – February 11, 2025

Marcos Schwarz – R&D Manager in Cyberinfrastructure

Acknowledgements

Contributions representing RNP

With technical support and collaboration with:
- GÉANT/RENATER RARE/GP4L/NMAAS projects
- GNA-G Data Intensive Science WG
- GNA-G AutoGOLE / SENSE WG
- FABRIC testbed
… And all it's collaborating institutions and teams

Motivation

1. How can we increase the rate of evolution of Research Networks without interfering with production?

2. How to develop and operate end-to-end / multi-domain orchestration services?
   - Resource reservation (guaranteed bandwidth)
   - Resource provisioning (Circuits, VRFs)
   - Underlay observability
   - Dynamic traffic steering/engineering
   - Dynamic creation of L3 VPNs
   - Closed loop multi-domain visibility/intelligence/controllability

3. How can we create/sustain an integration initiative/platform to propose and validate next generation protocol and services?

Proposition: Build programmable platforms to experiment on persistent pre-production networks leveraging industry/R&E open ecosystems

GP4L – Global Platform for Lab
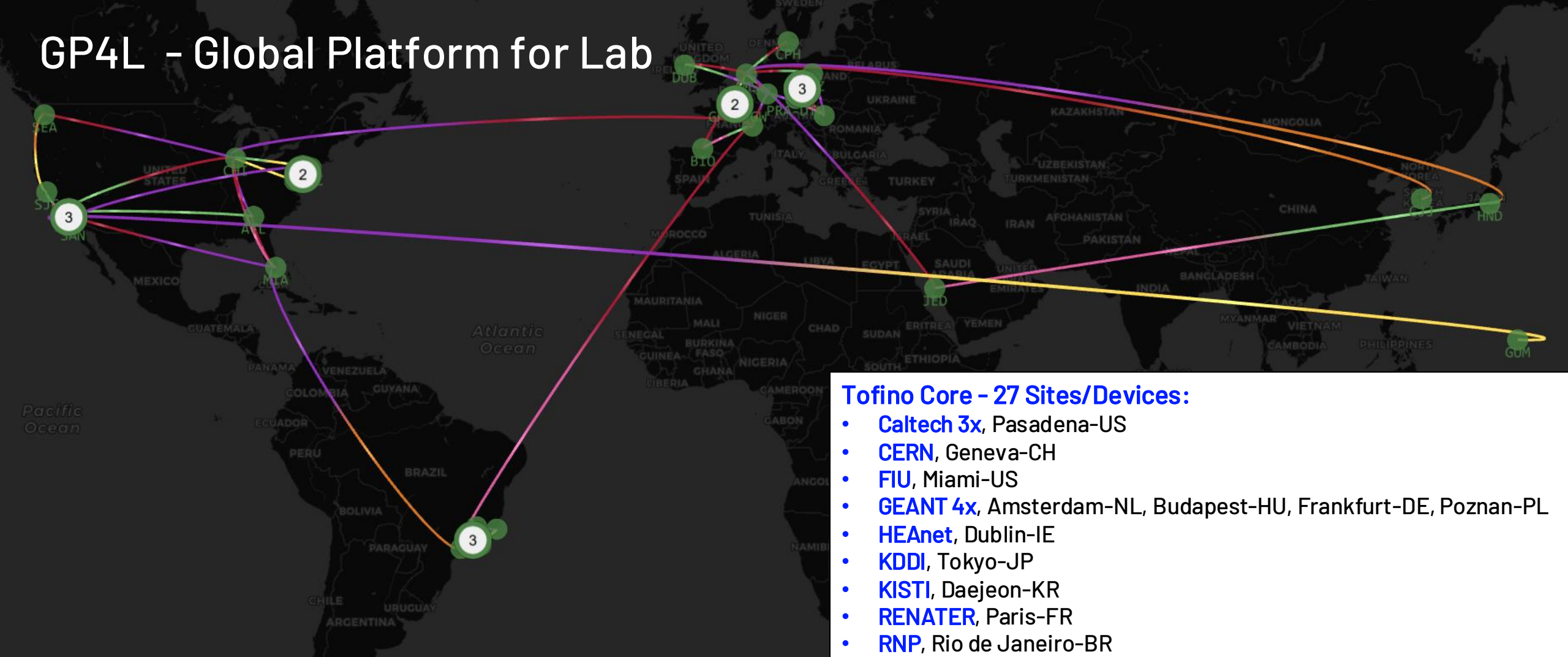
Tofino Core – 27 Sites/Devices:
- **Caltech 3x**, Pasadena-US
- **CERN**, Geneva-CH
- **FIU**, Miami-US
- **GEANT 4x**, Amsterdam-NL, Budapest-HU, Frankfurt-DE, Poznan-PL
- **HEAnet**, Dublin-IE
- **KDDI**, Tokyo-JP
- **KISTI**, Daejeon-KR
- **RENATER**, Paris-FR
- **RNP**, Rio de Janeiro-BR
- **SC24 [New]**, Atlanta-US
- **SouthernLight**, São Paulo-BR
- **StarLight**, Chicago-US
- **SWITCH 6x**, Geneva-CH
- **Tennessee Tech**, Cookeville-US
- **UFES**, Vitória-BR
- **UFMG [New]**, Belo Horizonte-BR
- **UMd/MAX**, College Park-US

BlueField-2/DPDK Islands – 7 Sites/Devices [New]:
- **Pacific Wave/UCSD**, Chicago-US, Guam-GU, Los Angeles-US, New York-US, San Diego-US, Seattle-US, Sunnyvale-US

x86/DPDK Islands – 4 Sites/Devices:
- **FABRIC [New]**, Miami-US
- **2x GEANT**, Paris-FR, Prague-CZ
- **KAUST [New]**, Saudi Arabia-SA

Context

- **State of P4 hardware, since Tofino was discontinued in 2023**
  - IPUs and DPUs from Intel and AMD/Xilinx are currently the norm
  - Next-Generation P4 ASIC - [Xsight Labs X2 + Oxide P4 compiler](#)
  - Tofino [compiler](#) and [SDE](#) were Open Sourced as part of p4lang

- **We are constantly looking for future stacks that can provide production grade features and deep programmability**
  - freeRtr + Tofino: Base architecture
  - SONiC/PINS/DASH + SAI:
    - Future proof with the biggest community and ecosystem of partners
    - Progress is slow but steady and high barrier to extend/contribute
  - **VPP** + P4-DPDK + Linux Control Plane
    - Everything is in software and easier/faster prototyping
    - Flexible control plane options: FRR, **Holo**, vendors (Arista, Juniper, Nokia...)

Vector Packet Processor (VPP)

- **Rich Feature set**
  - 14+ MPPS, single core
  - Multipath – ECMP and Unequal Cost
  - Network Programming using SR-MPLS and SRv6, Traffic Steering

- **Device Drivers**
  - DPDK, eBPF/XDP, RDMA, device specific (Intel, Marvell)

- **Linux Control Plane**
  - Any control plane

- **Adopted by many high performance routing projects**
  - SONiC, VyOS, Netgate TNSR, Calico-VPP

Current/Future Components and Features

(Open Source)	(Commercial)
freeRtr	Arista cEOS
Holo	Cisco xRD
SONiC (FRR)	Juniper cRPD
VyOS (FRR)	Nokia SR Linux

Linux Control Plane
VPP

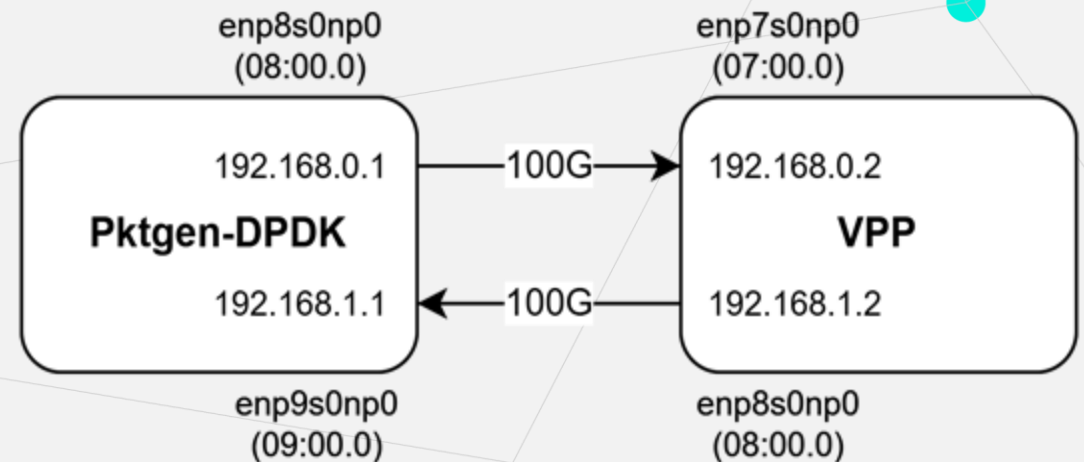| Linux Networking	RDMA	**DPDK**

Nvidia ConnectX-6
Nvidia Bluefiled-2

EGP (BGP)
IGP (BGP, ISIS, OSPF)
NETCONF, RESTCONF, gNMI
Traffic Engineering (BGP, PCEP)

Segment Routing (SR-MPLS, SRv6)
Telemetry (SNMP, sFlow)

Quality of Service (HTB)

Device Driver	Connectx-6
Linux Networking	17 Gbps
RDMA	44 Gbps
DPDK	99 Gbps

Initial Experiment (FABRIC and NRP)

- Evaluate VPP/DPDK performance on Connectx-6 (FABRIC) and BlueField-2 (NRP)
  - 100G interfaces
  - Packet sizes: 64B packet, 1500/9000 MTU, 4096/8192 MSS
 - pktgen-DPDK as traffic generator
  - Back to Back scenario

Packet Size	ConnectX-6	BlueField-2	
	x86_64	x86_64	arm64
64	5G	3G	2G
1518	73G	48G	36G
4154	100G	99G	64G
8250	100G	98G	48G
9018	100G	96G	33G

enp8s0np0 (08:00.0)   enp7s0np0 (07:00.0)

Pktgen-DPDK
192.168.0.1 —100G→ 192.168.0.2 VPP
192.168.1.1 ←100G— 192.168.1.2

enp9s0np0 (09:00.0)   enp8s0np0 (08:00.0)

Second Experiment (FABRIC only)

- Evaluate encapsulation performance on different packet sizes
  - 100G interfaces
  - Size: 64B packet, 1500/9000 MTU, 4096/8192 MSS
  - Encapsulation: None, VLAN, MPLS, SRv6
  - Encapsulation between VPP nodes

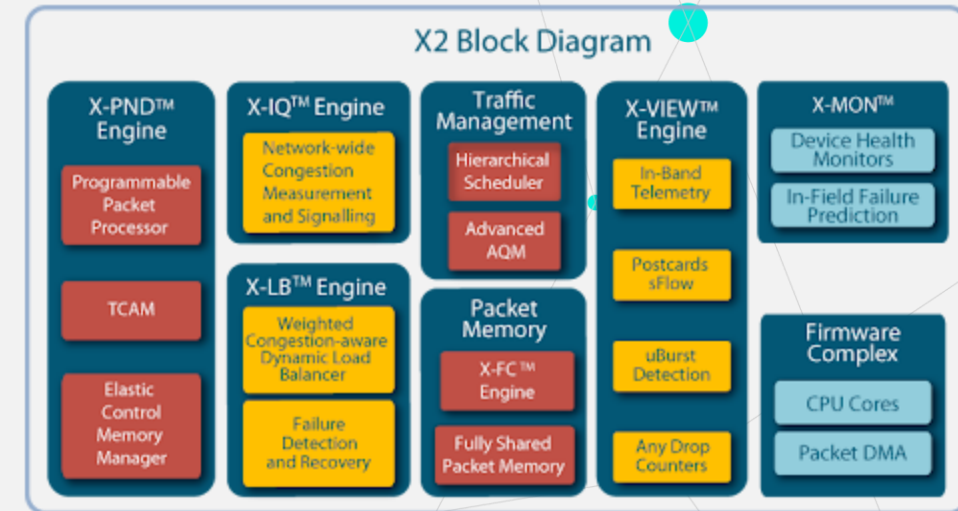| Packet Size	VLAN	MPLS	SRv6
64	4G	5G	8G
1518	65G	73G	73G
4154	100G	100G	100G
8250	100G	100G	100G
9018	100G	100G	100G
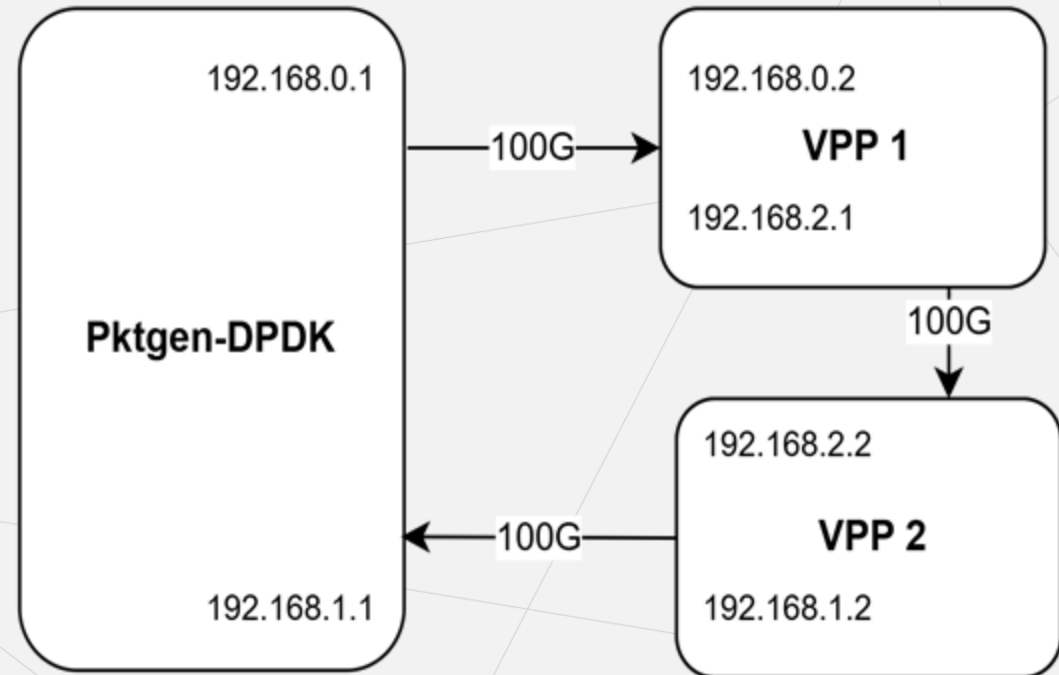
Future Work

- Publish FABRIC Jupyter Notebook
- Control Plane Integration: FRR, Holo, vendors (Arista, Juniper, Nokia…)
- Multiple 100G WAN Multipath + Traffic Engineering Scenarios
- Hardware Accelerated QoS
- P4-DPDK Integration

VPP + P4-DPDK Integration (Draft proposal)
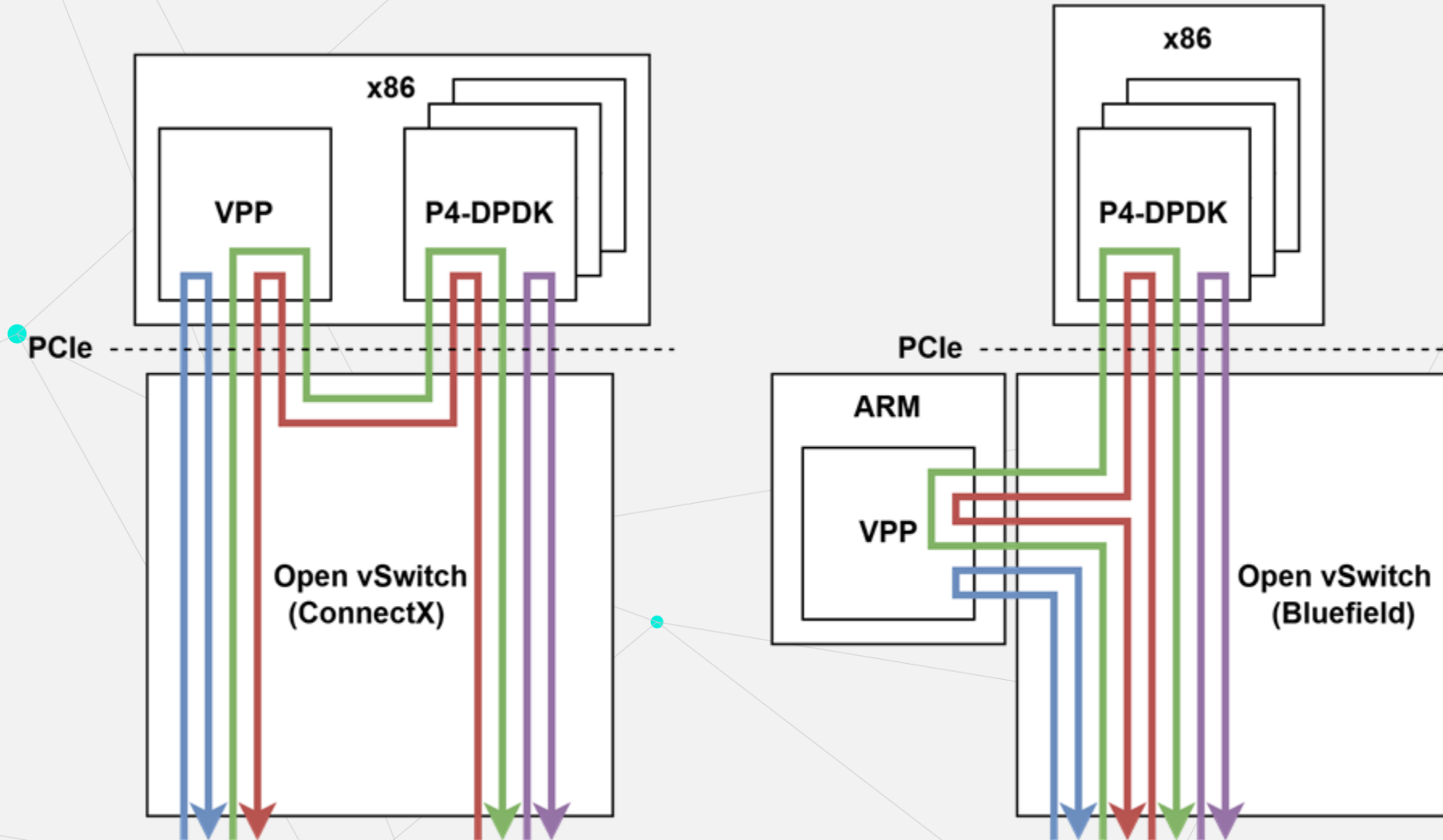
Next Generation Programmable Platforms

- **Xsight Labs + Oxide P4 compiler**
  - X2 Programmable ASIC (possible Tofino successor)
  - Not available yet
  - Open P4 compiler and SONiC support expected

- **Cisco 8102-28FH-DPU**
  - Supports SONiC DASH
  - Programmable ASIC (Silicon One) and DPU (Pensando)
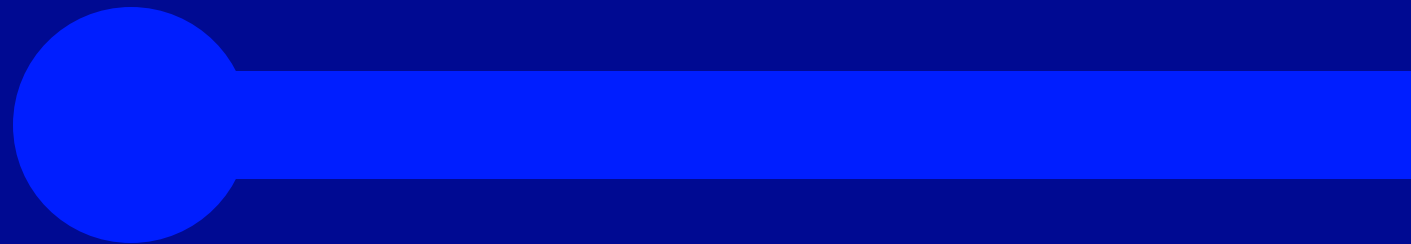   - but not sure if exposed to the user

Questions?