

Open. Together.

OCP SUMMIT

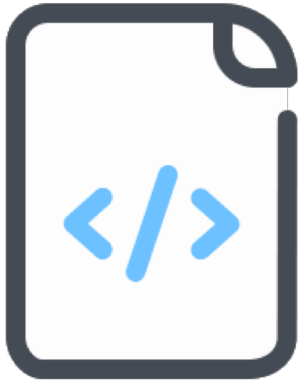
Minipack and F16, Software

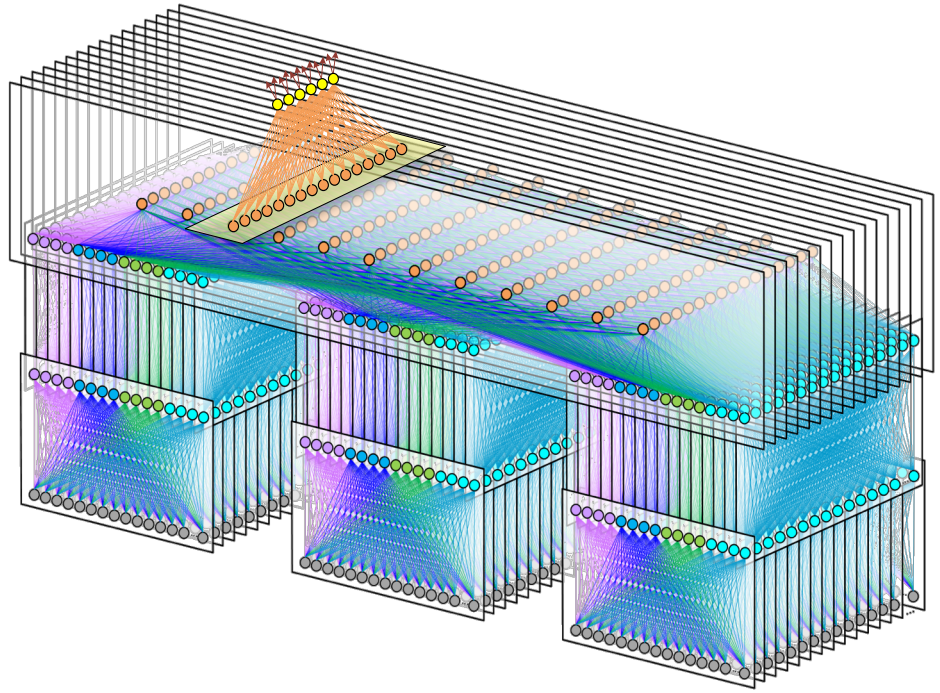
Alex Eckert, Software Engineer
Facebook



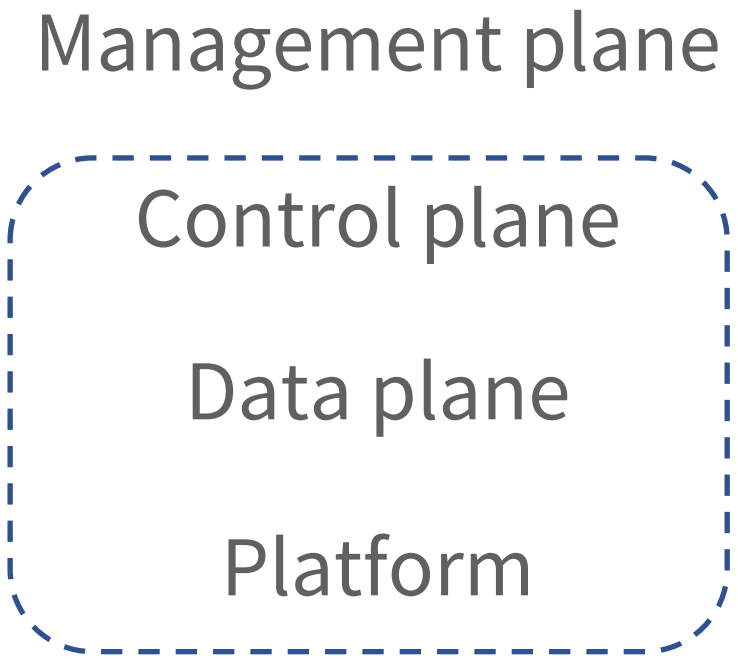
Minipack and F16, Software







Software everywhere



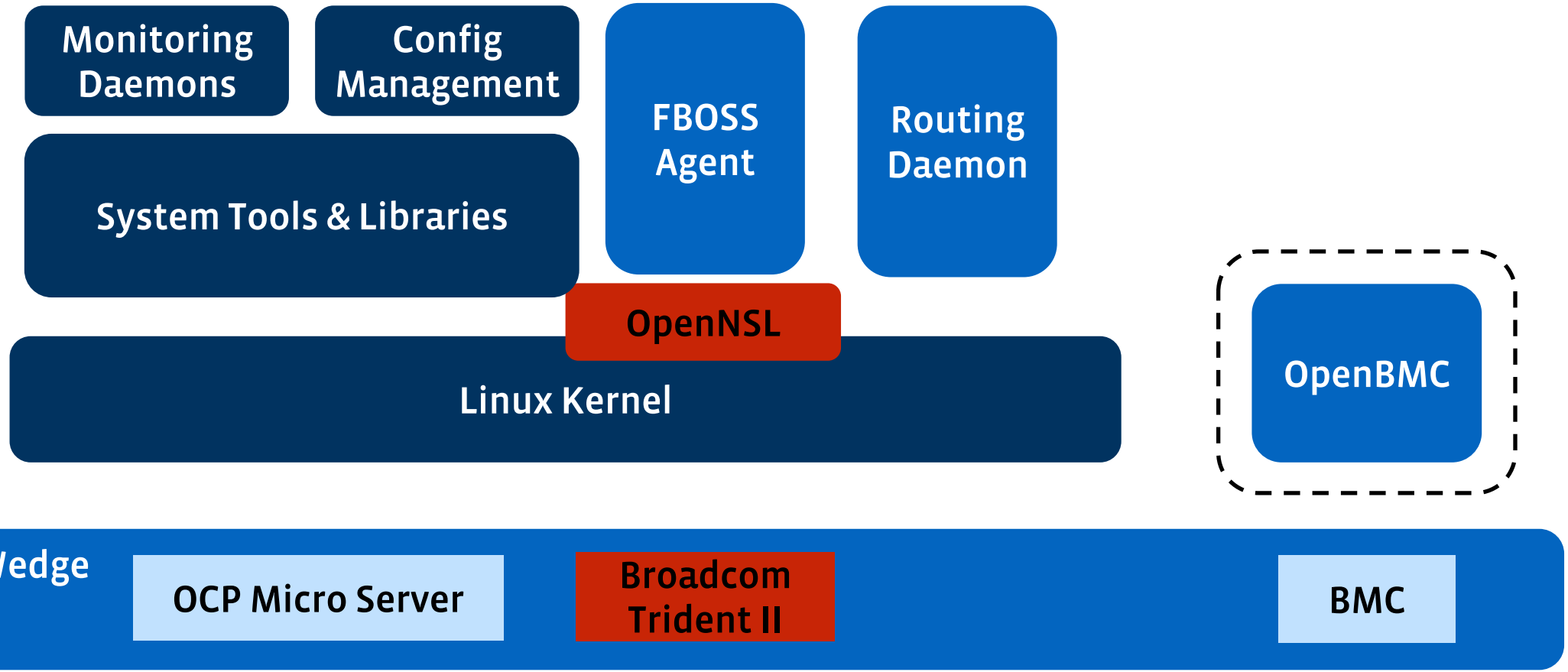
Wedge

FBOSS



16 x 40G - RSW

FBOSS Overview



Sixpack

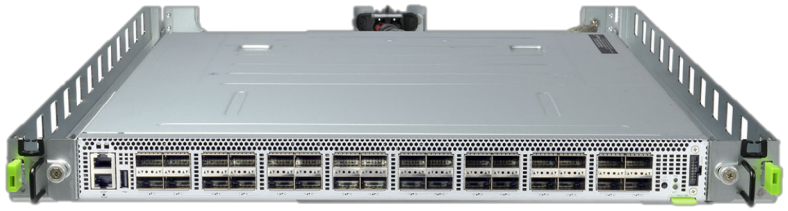
FBOSS



128 x 40G - FSW

Wedge 100

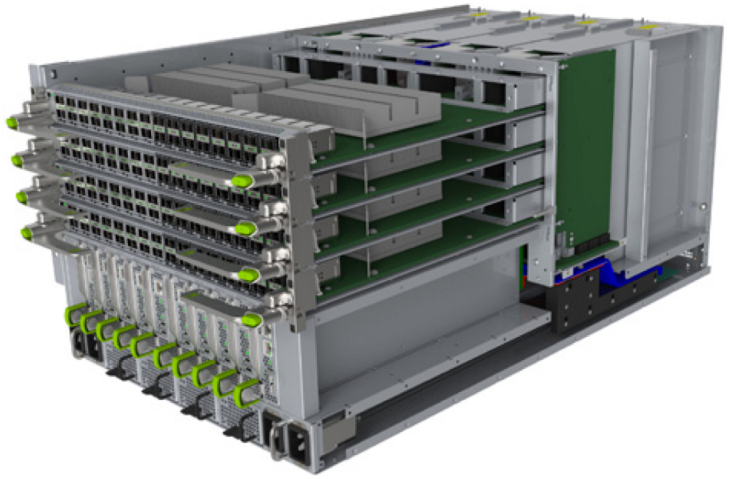
FBOSS



32 x 100G - RSW

Backpack

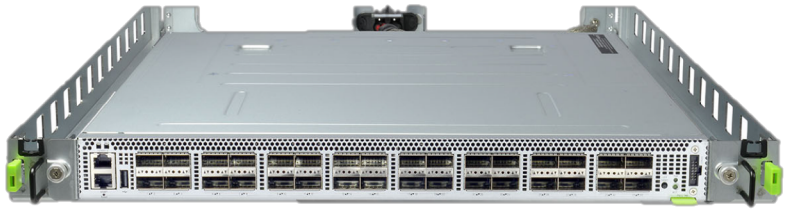
FBOSS



128 x 100G - FSW

Wedge 100S

FBOSS

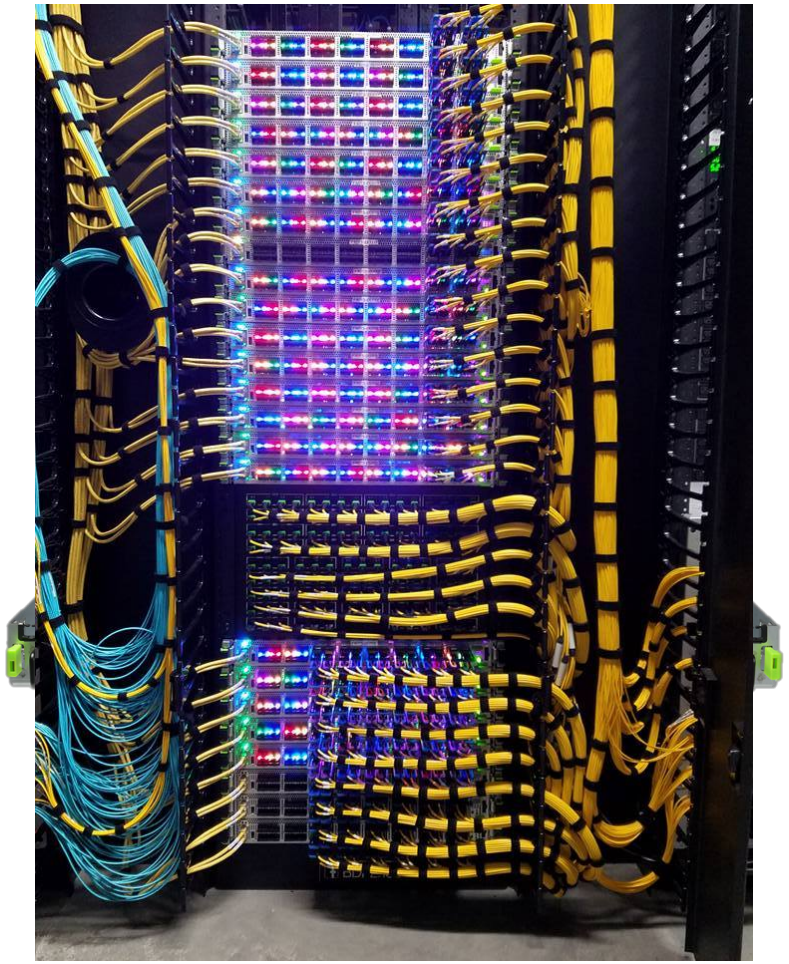


32 x 100G - RSW

Fabric Aggregator



FBOSS



N x 100G - FA

Minipack



128 x 100G – FSW, SSW, FA

New challenges (Minipack)

- Modular switch, single control plane
- Different PIM types, variety of port speeds
- External phy support (Broadcom)
- I2C to 128 optics, MDIO, FPGA
- New microserver (MiniLake), New ASIC (Broadcom TH3)
- Cooling 128 optics in only 4 RU



128 x 100G – FSW, SSW, FA

Arista 7368X4

FBOSS



128 x 100G – FSW, SSW, FA

New challenges (FBOSS on Arista)

- First time operating FBOSS on non-FB hardware
- Different external phy (Credo), Different FPGA
- Unfamiliar microserver, no off-the-shelf UEFI BIOS
- Must manage as either FBOSS or EOS in production
- Complex conversion process b/w EOS / FBOSS

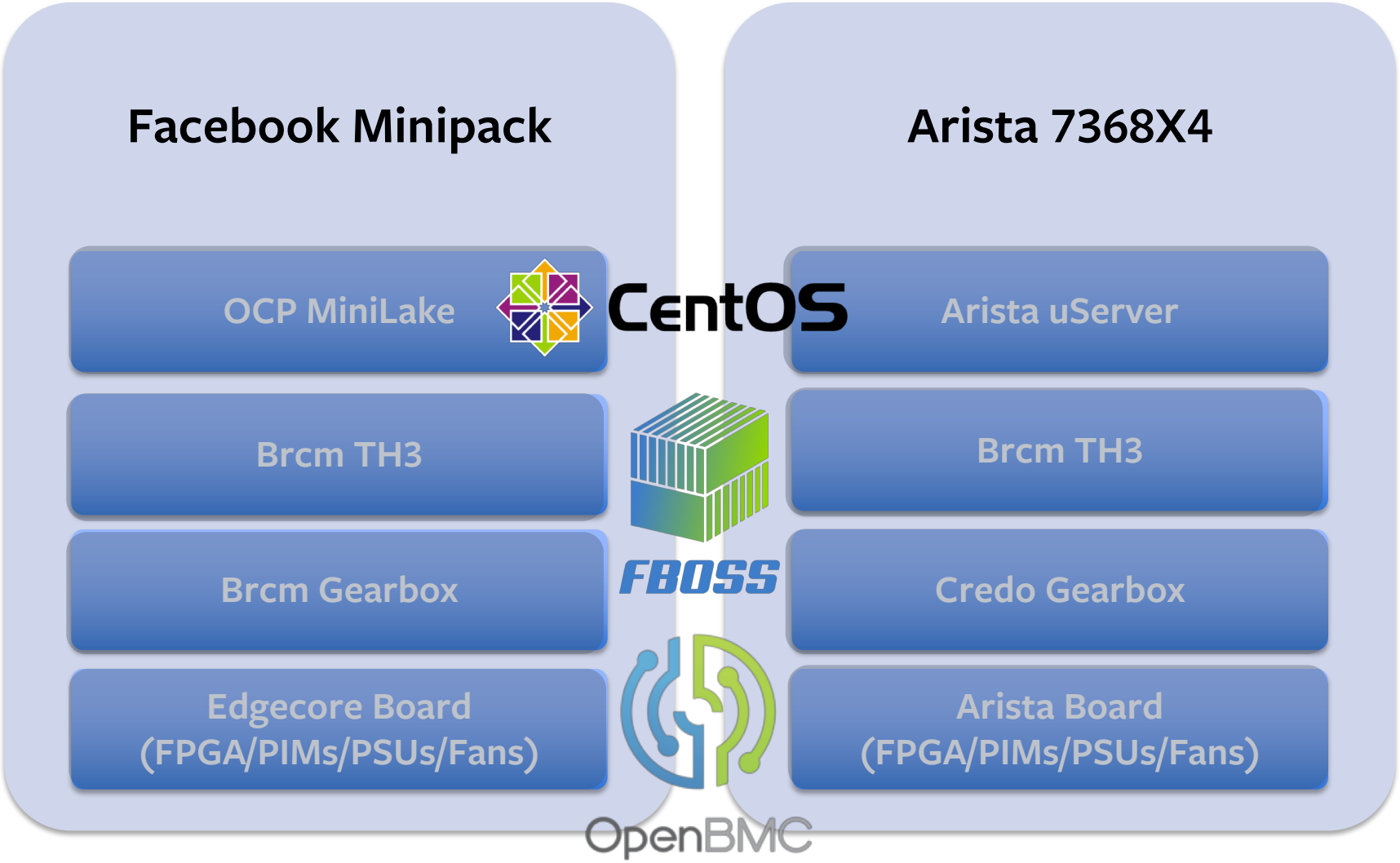


128 x 100G – FSW, SSW, FA

Combinations

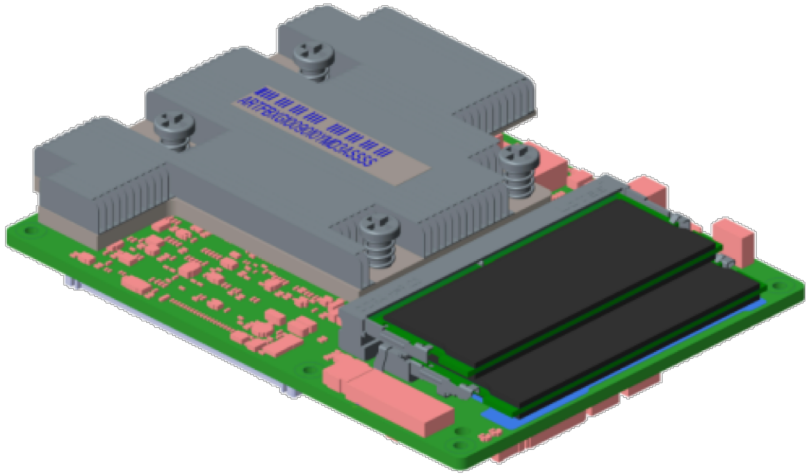
- Two switch models
- Modularity
- Three layers of F16 network
- Same binary everywhere

Common abstractions


Facebook Minipack
Arista 7368X4
OCP MiniLake
CentOS
Arista uServer
Brcm TH3
Brcm TH3
FBOSS
Brcm Gearbox
Credo Gearbox
Edgecore Board
(FPGA/PIMs/PSUs/Fans)
Arista Board
(FPGA/PIMs/PSUs/Fans)
OpenBMC

New microservers!

MiniLake

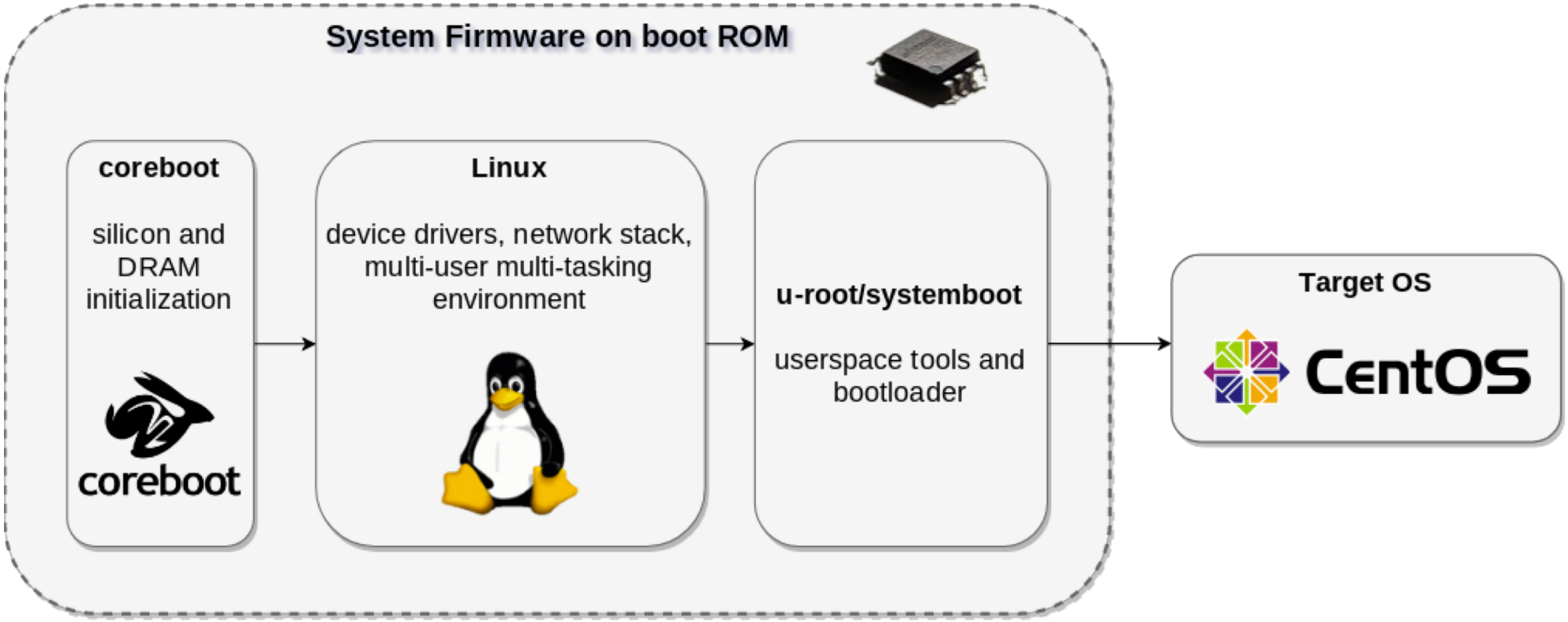


New microservers!

Arista control module



OCP Open System Firmware



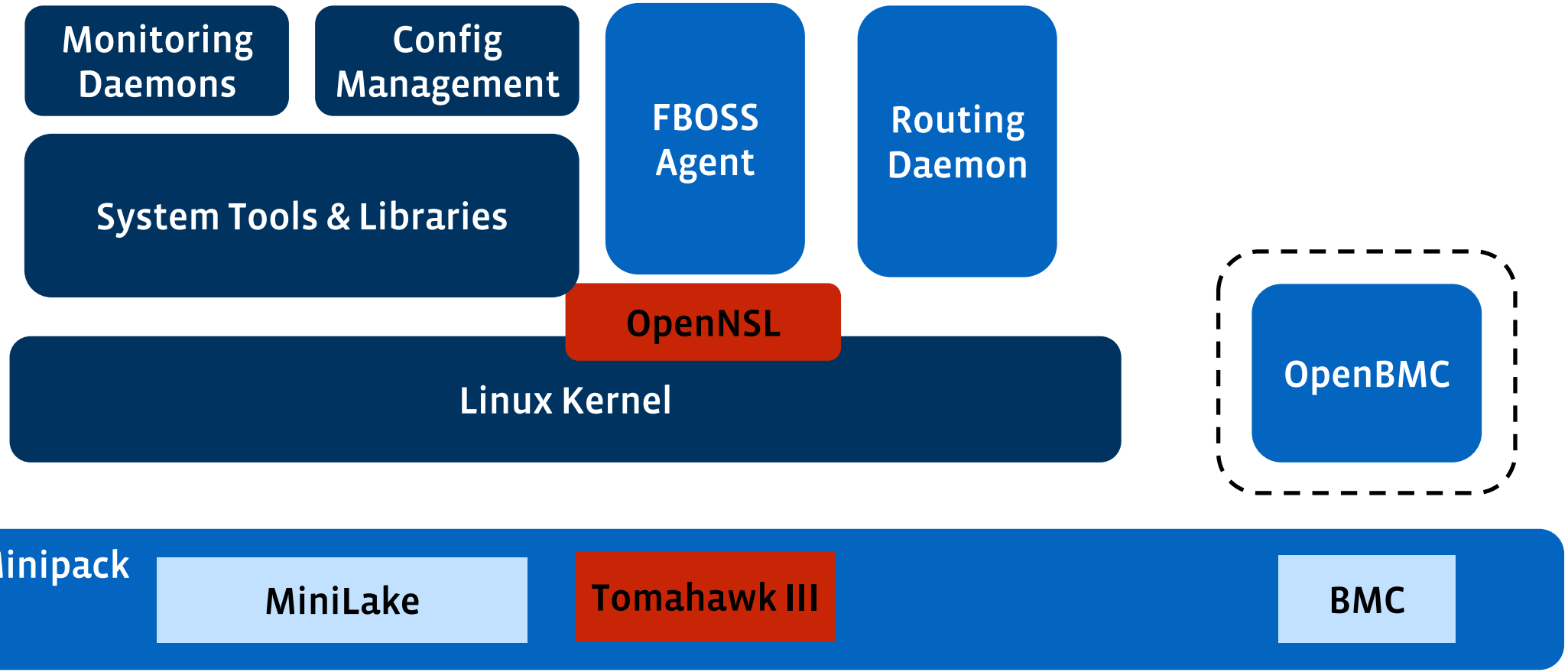
Conversion

ARISTA EOS

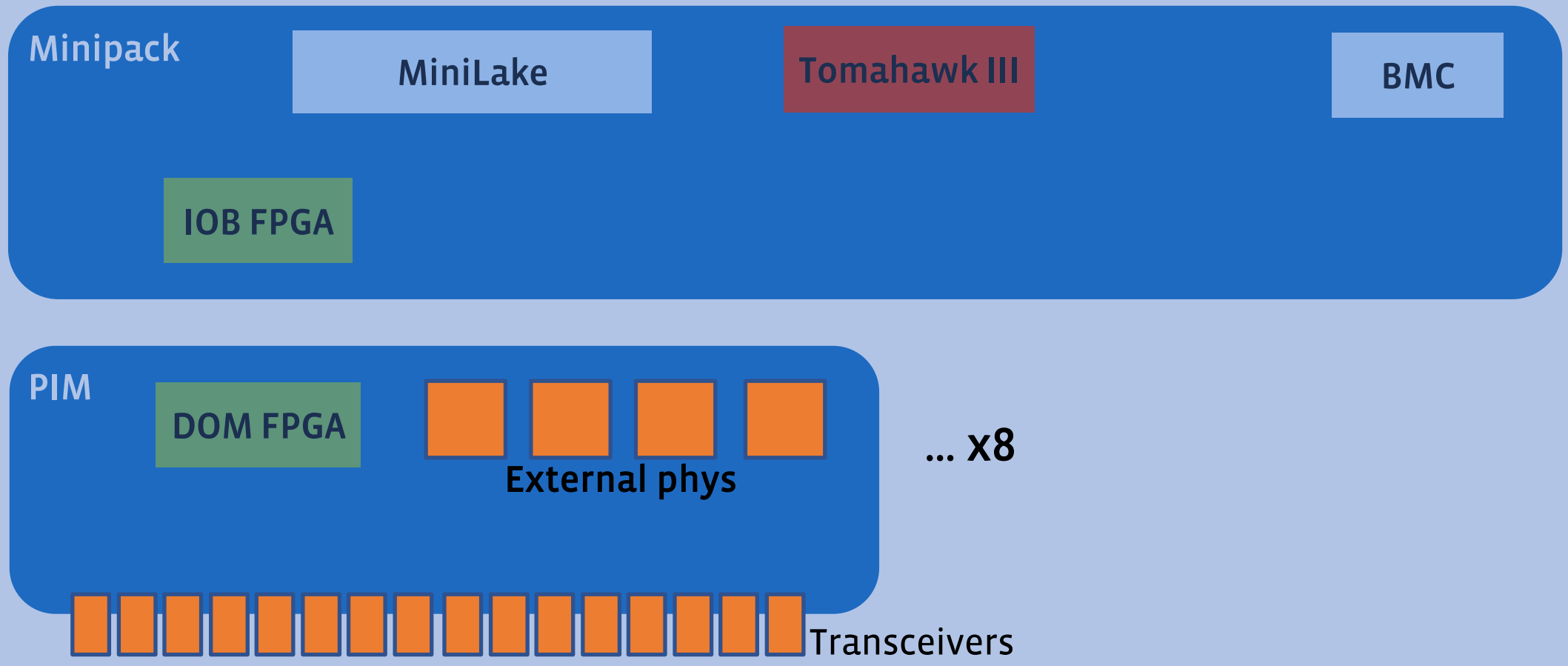




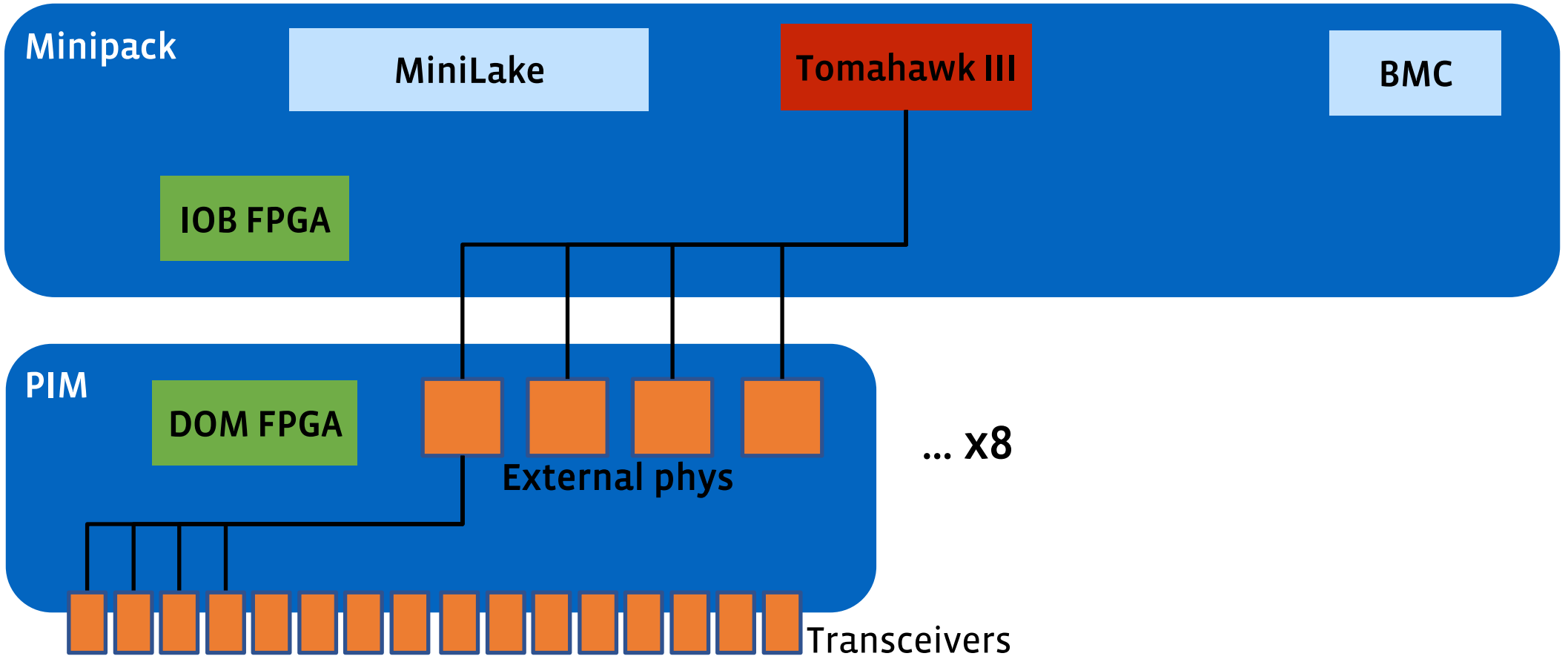
FBOSS Overview



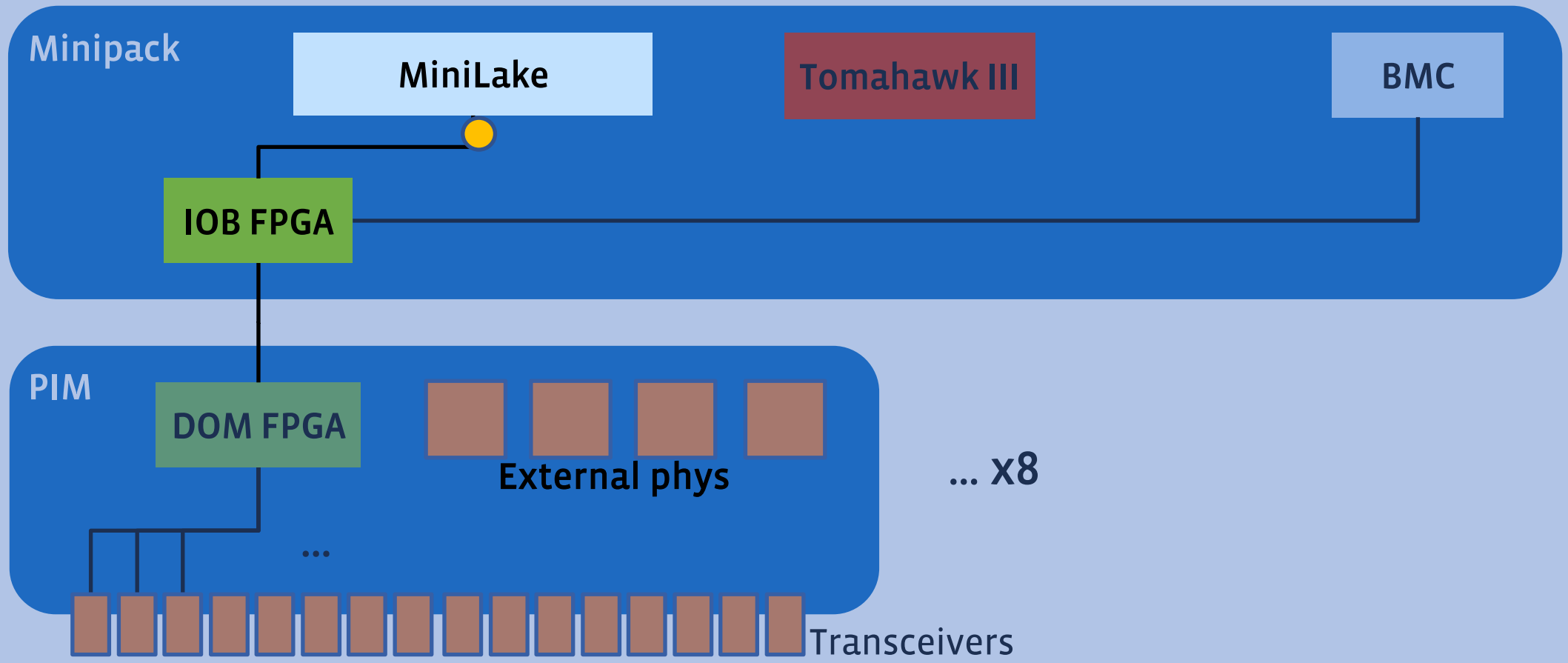
Minipack Hardware Components



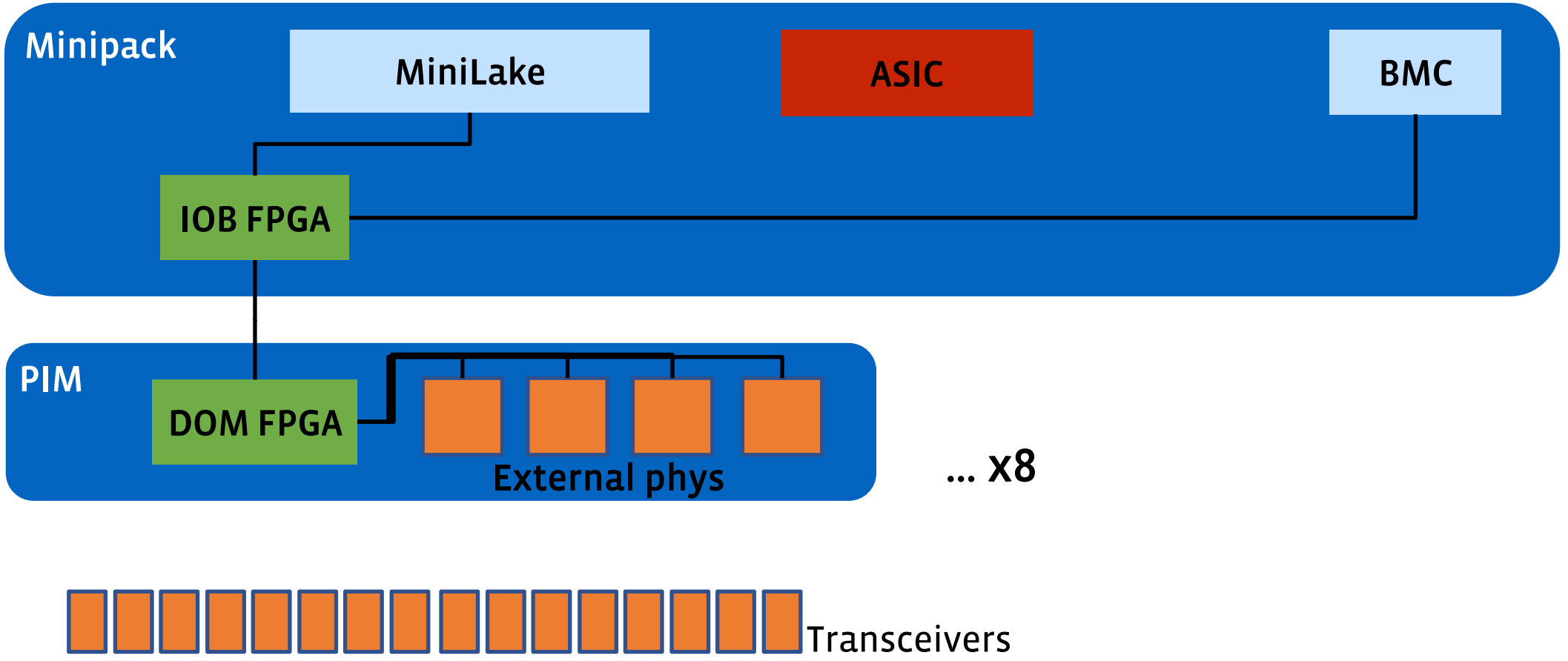
Data path


Minipack
MiniLake
Tomahawk III
BMC
IOB FPGA
PIM
DOM FPGA
External phys
... x8
Transceivers

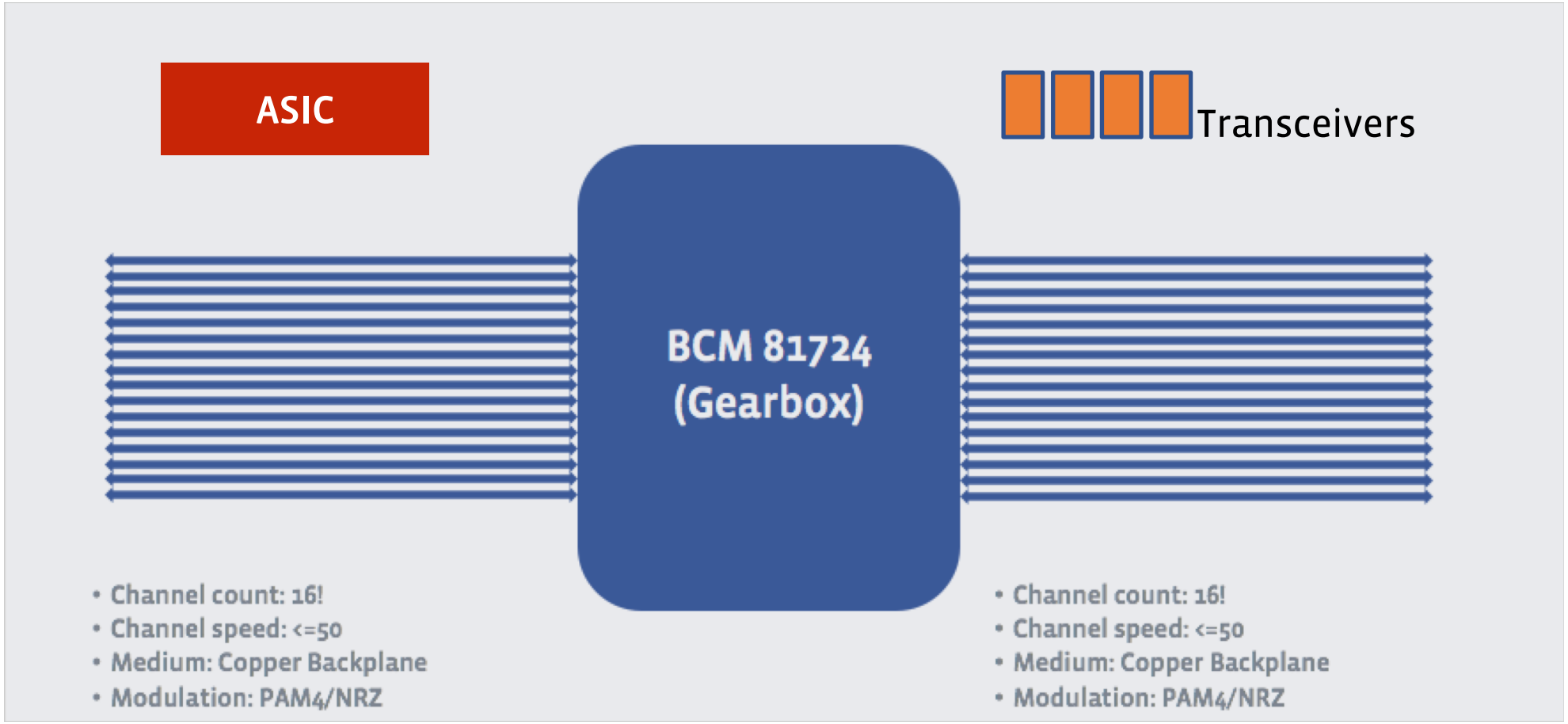
Management path – I2C



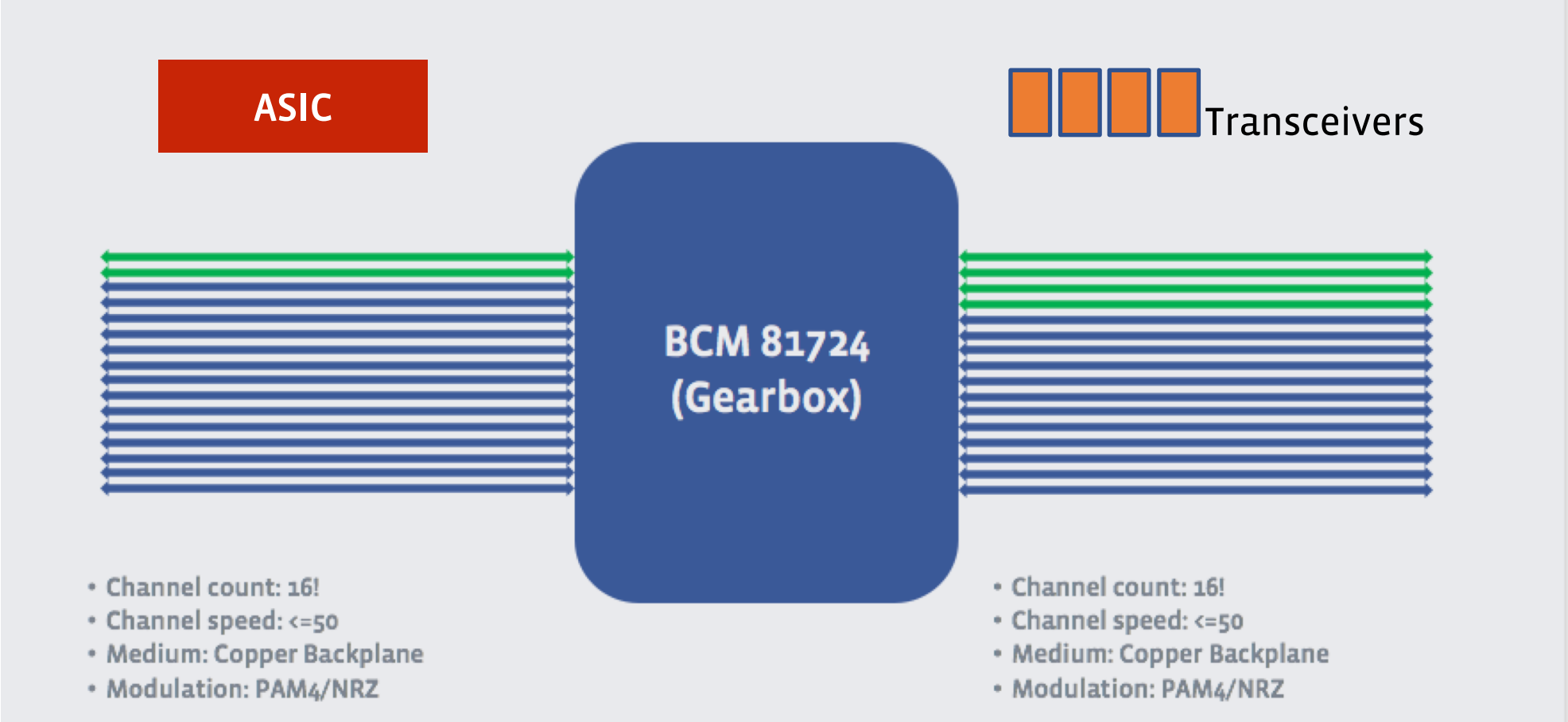
Management path – MDIO



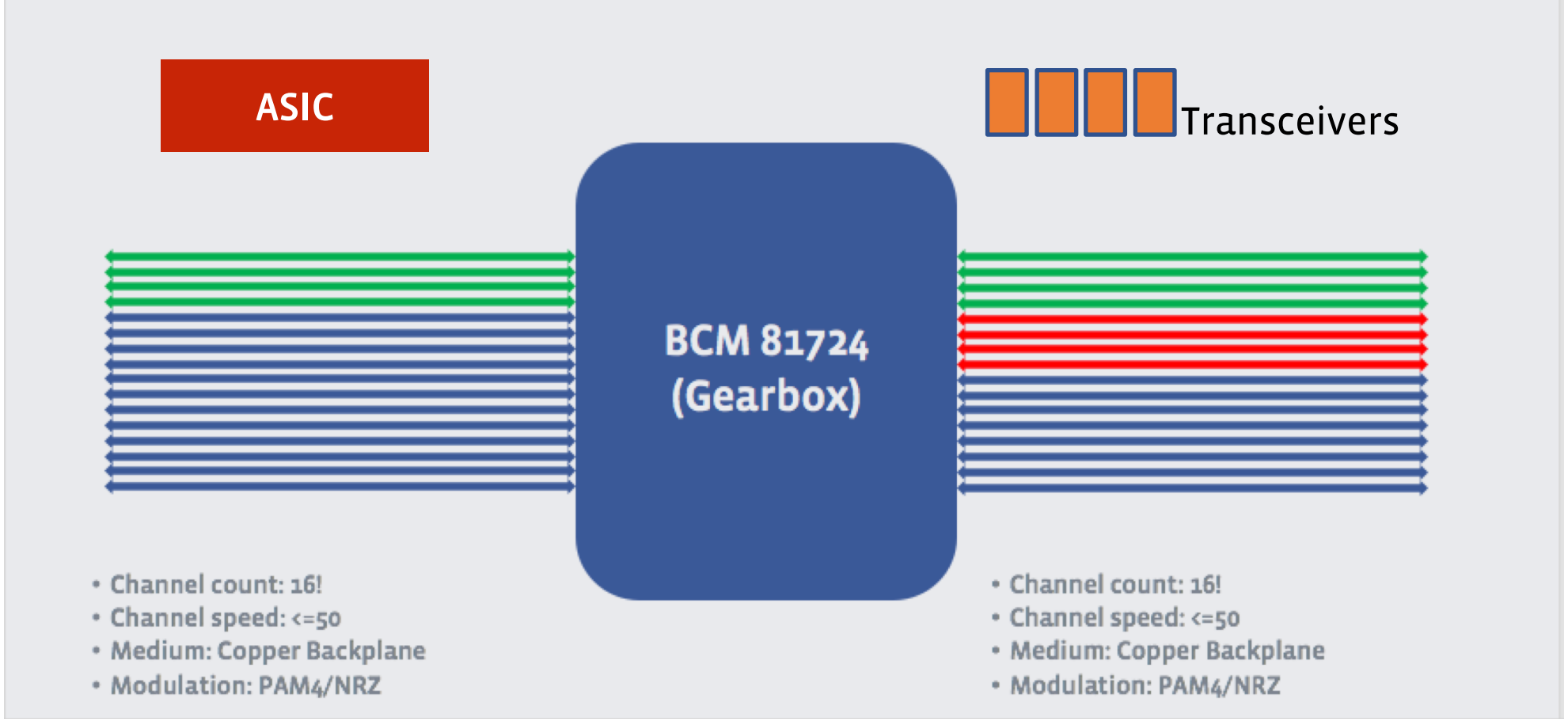
Programming External PHYs



Programming External PHYs

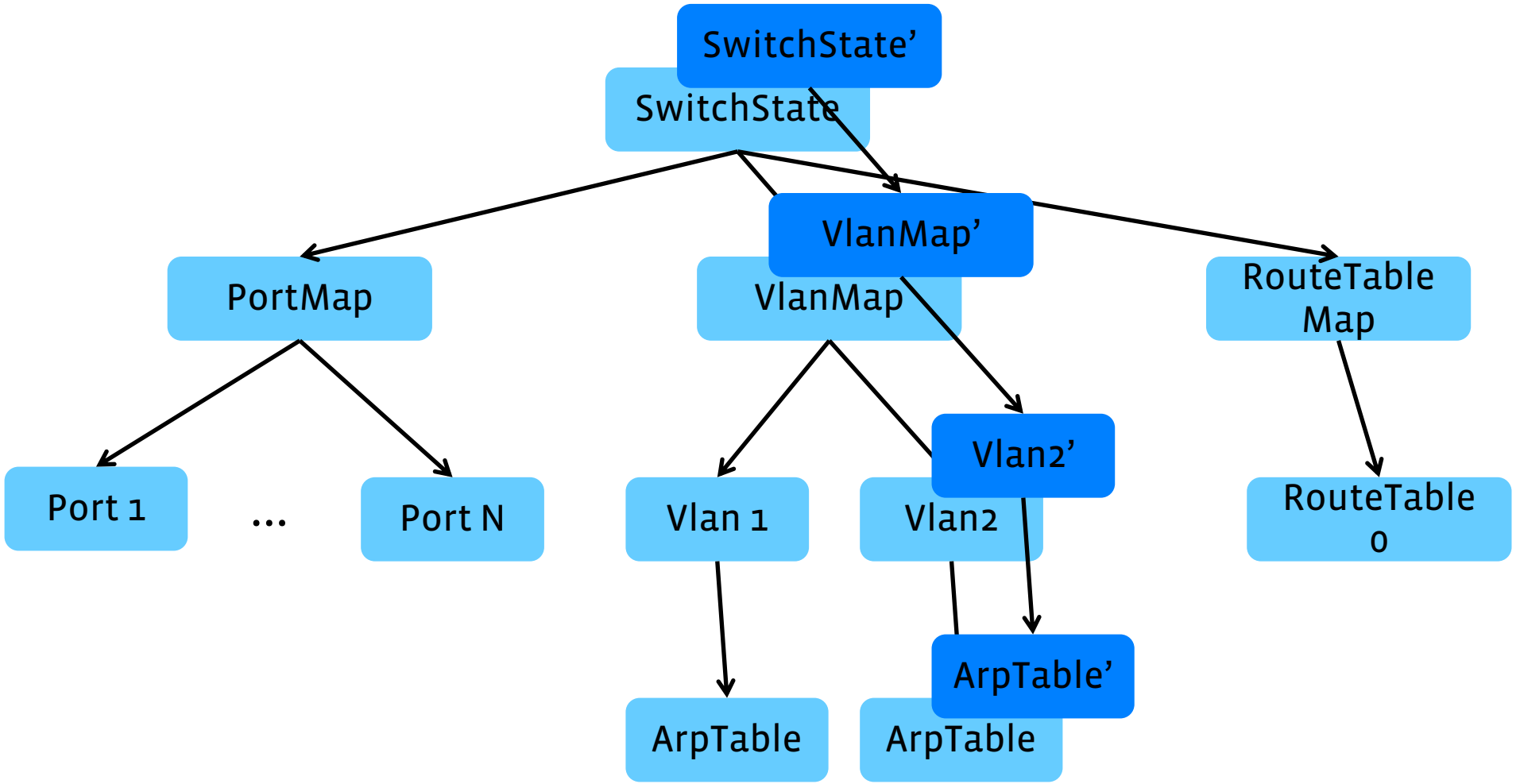


Programming External PHYs

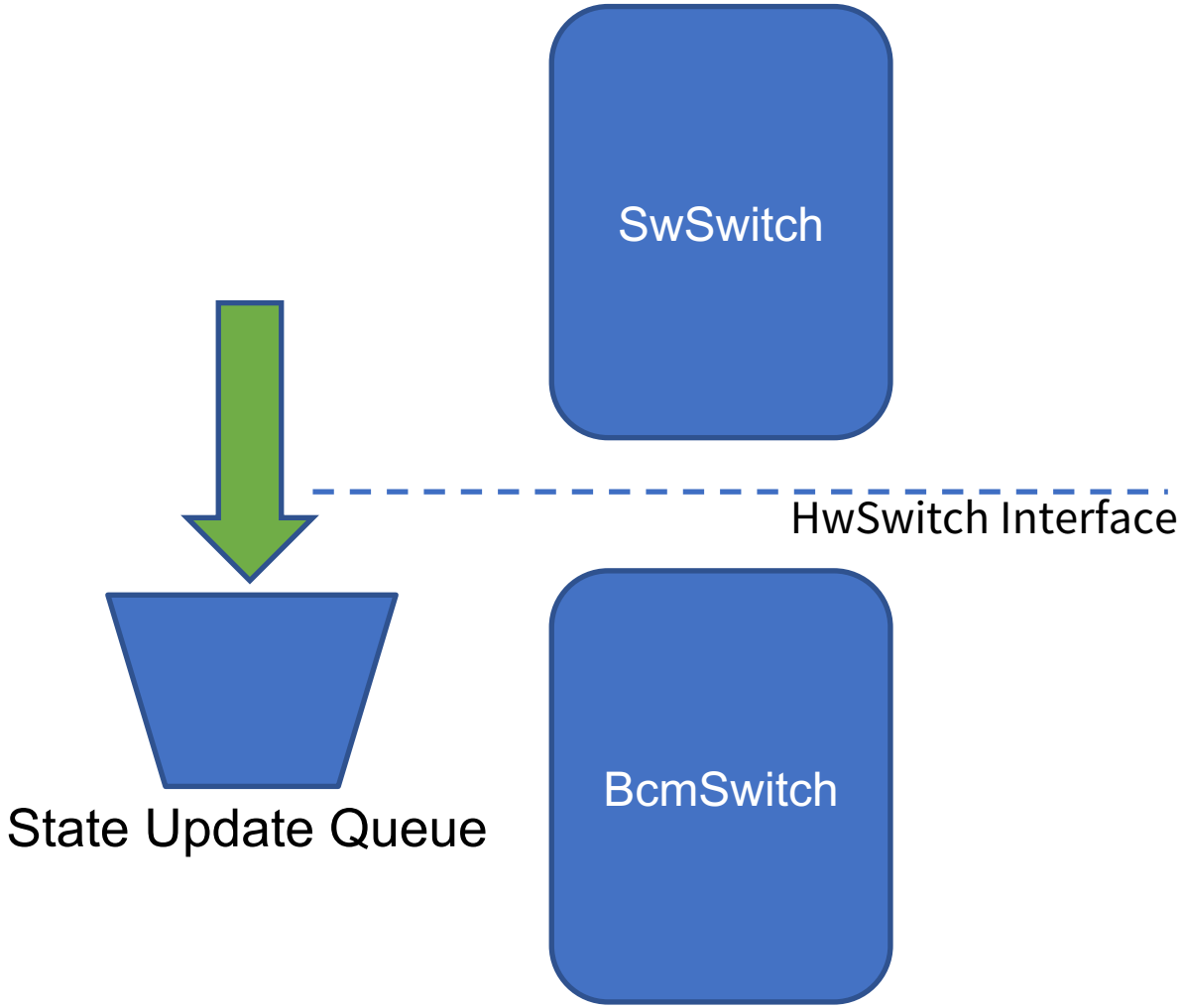


State management problem

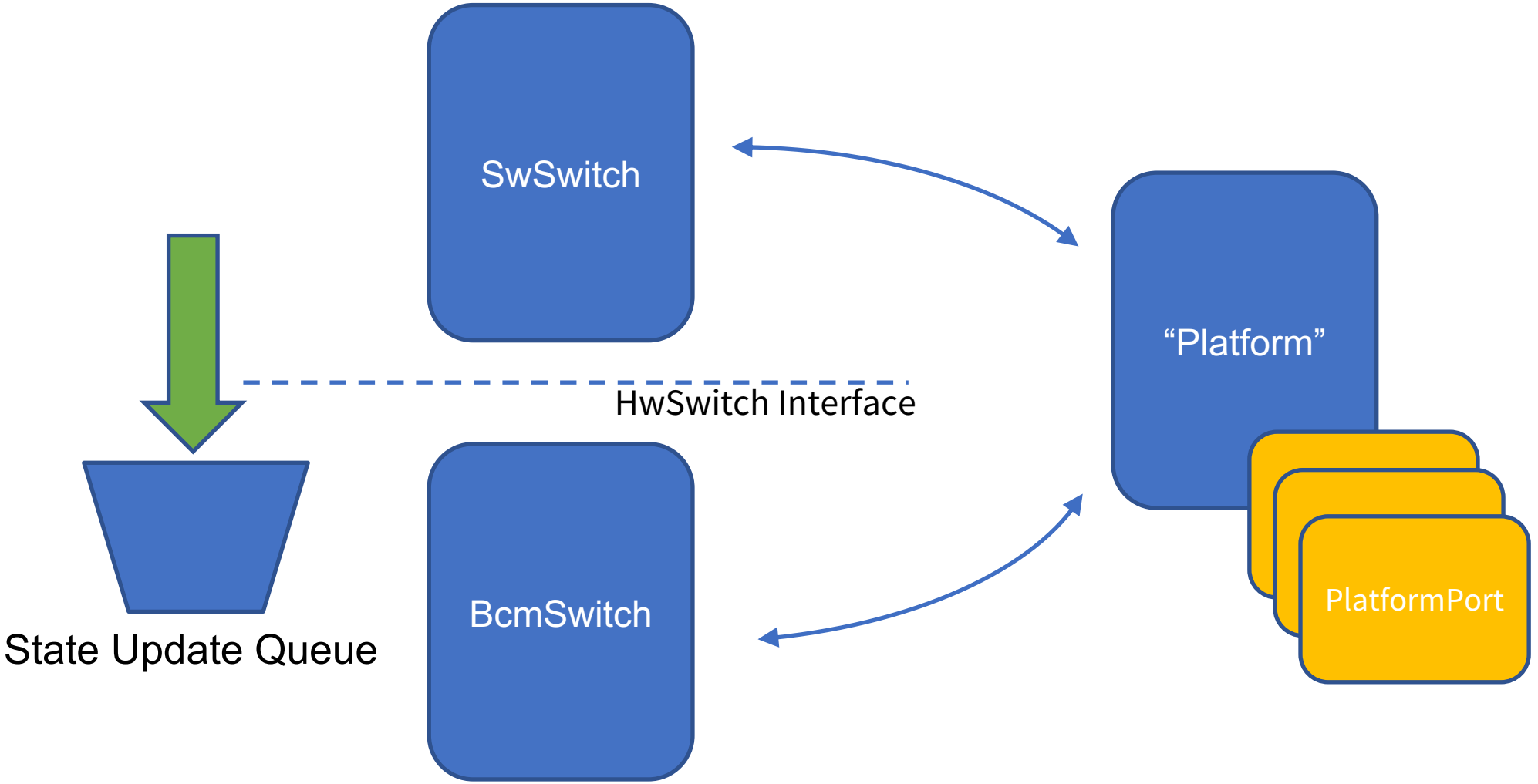
Switch State



FBOSS Agent


SwSwitch
HwSwitch Interface
State Update Queue
BcmSwitch

FBOSS Agent


SwSwitch
"Platform"
HwSwitch Interface
PlatformPort
State Update Queue
BcmSwitch

Platform Configuration

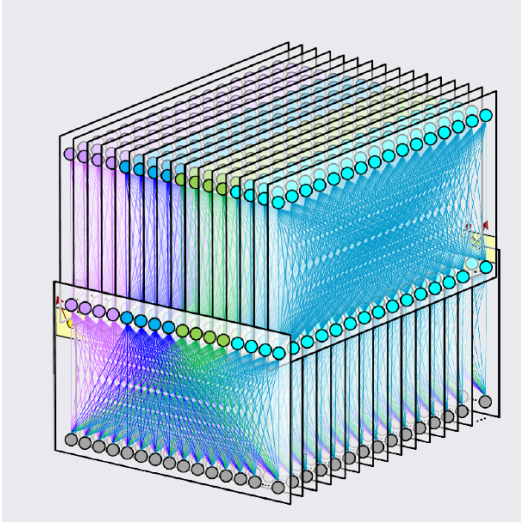
- Each port + speed has unique settings (PlatformPortSettings)
- Model config as a map<speed, PlatformPortSettings>
- Hides complexity from SwSwitch

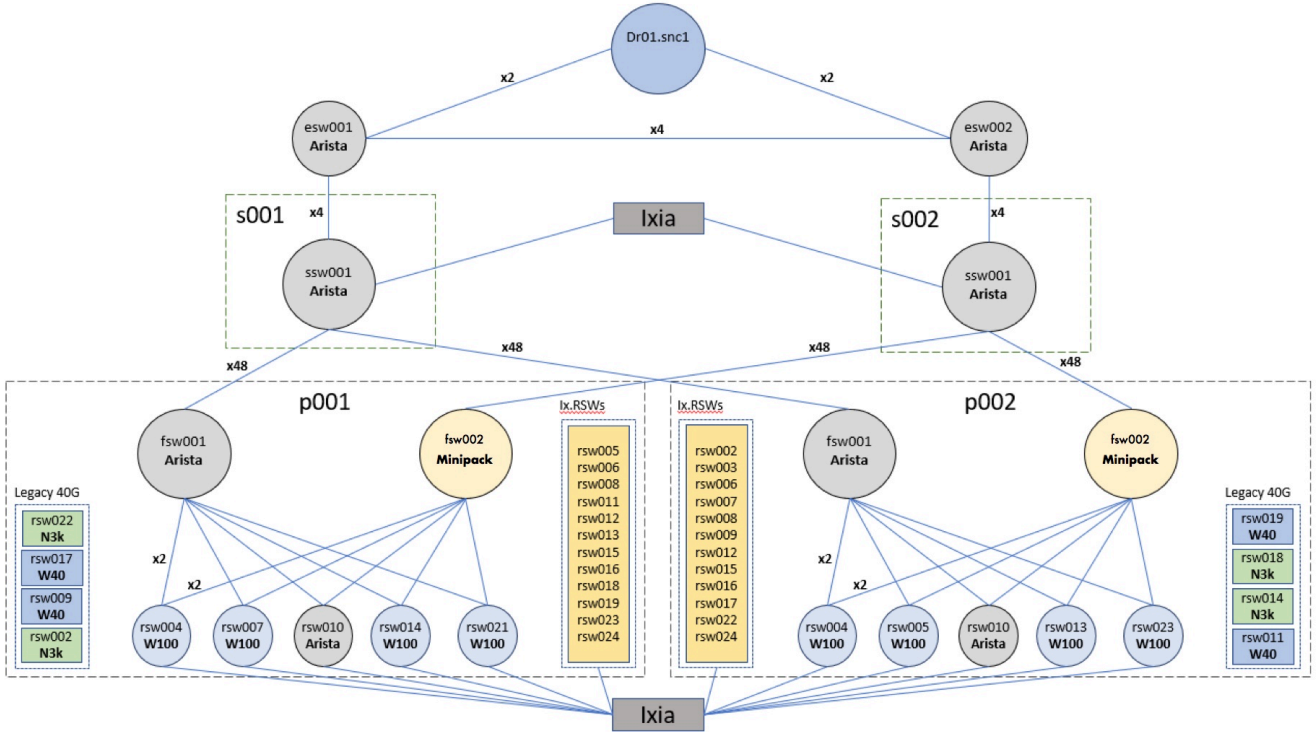
Testing

Testing strategy

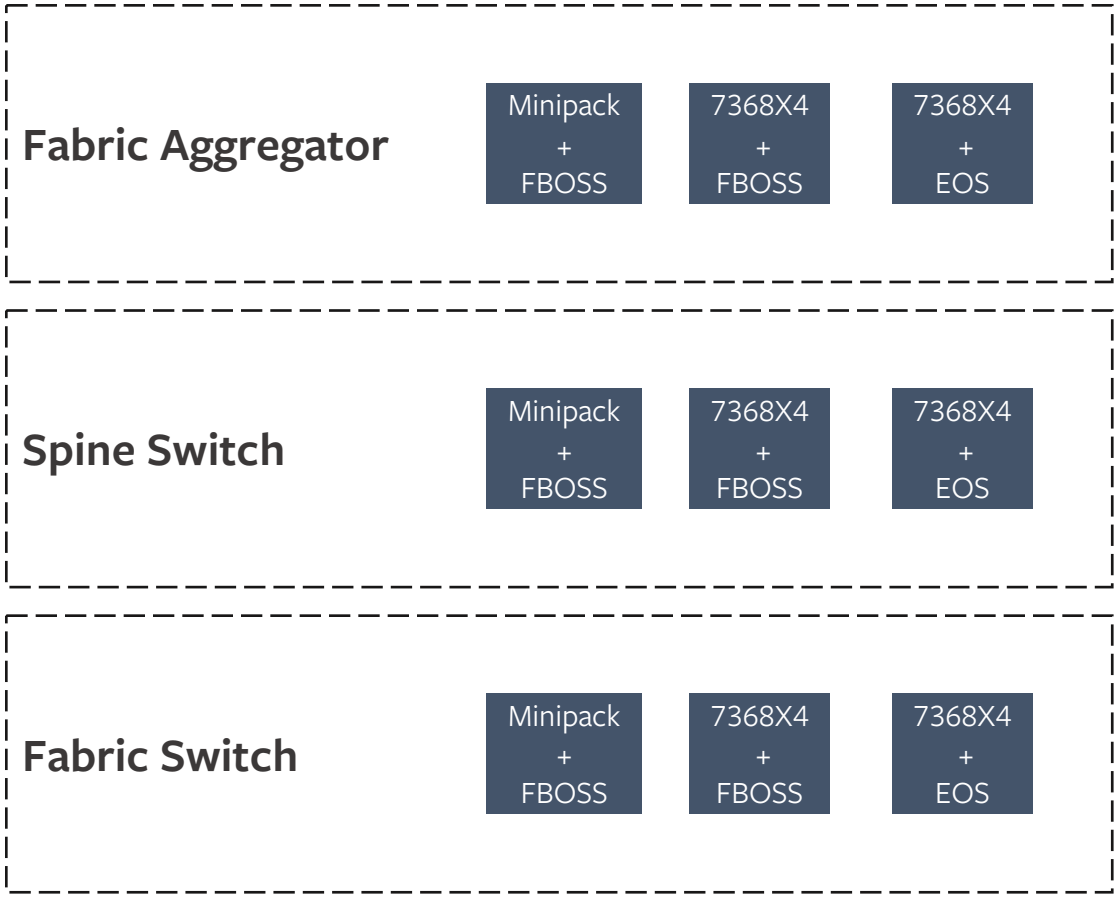
- Targeted tests in realistic lab environments
- Deploy early
- Invest in automation

Scale challenge





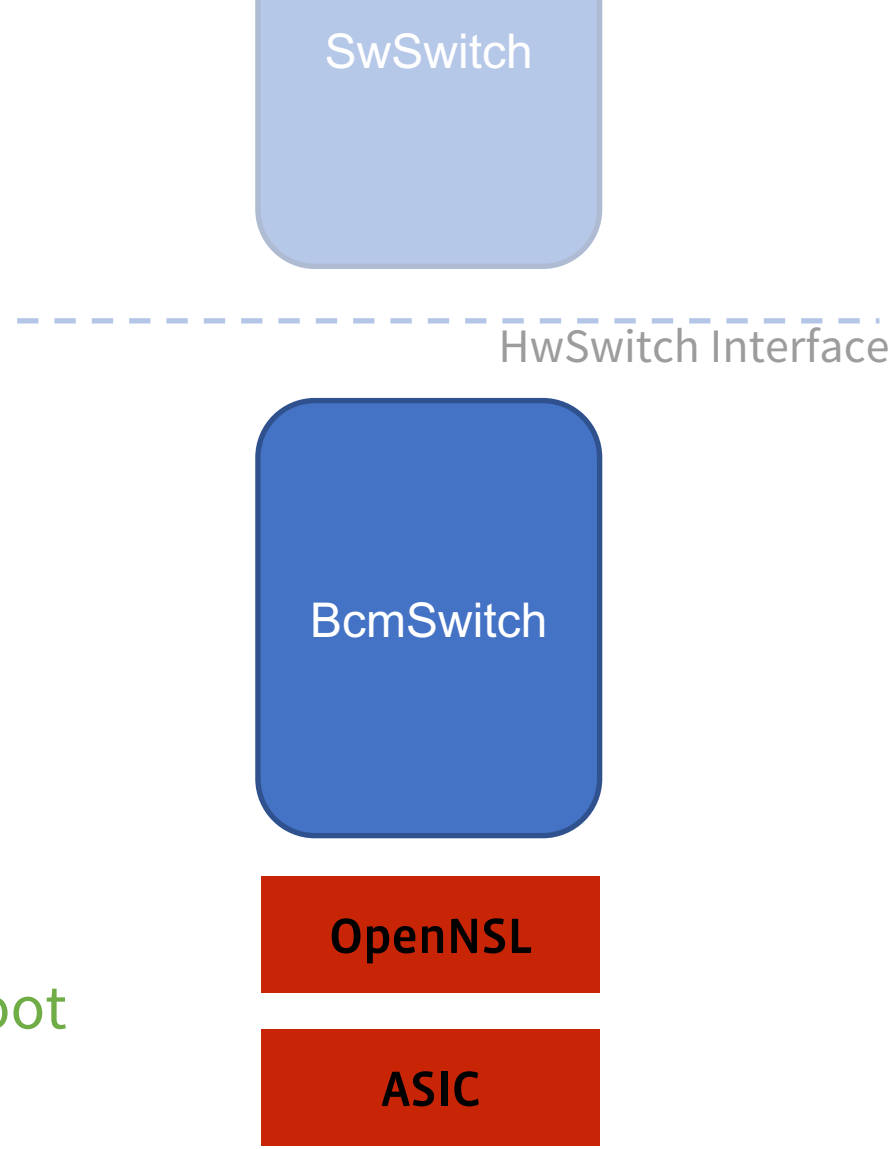
Test all combinations



Deploy early

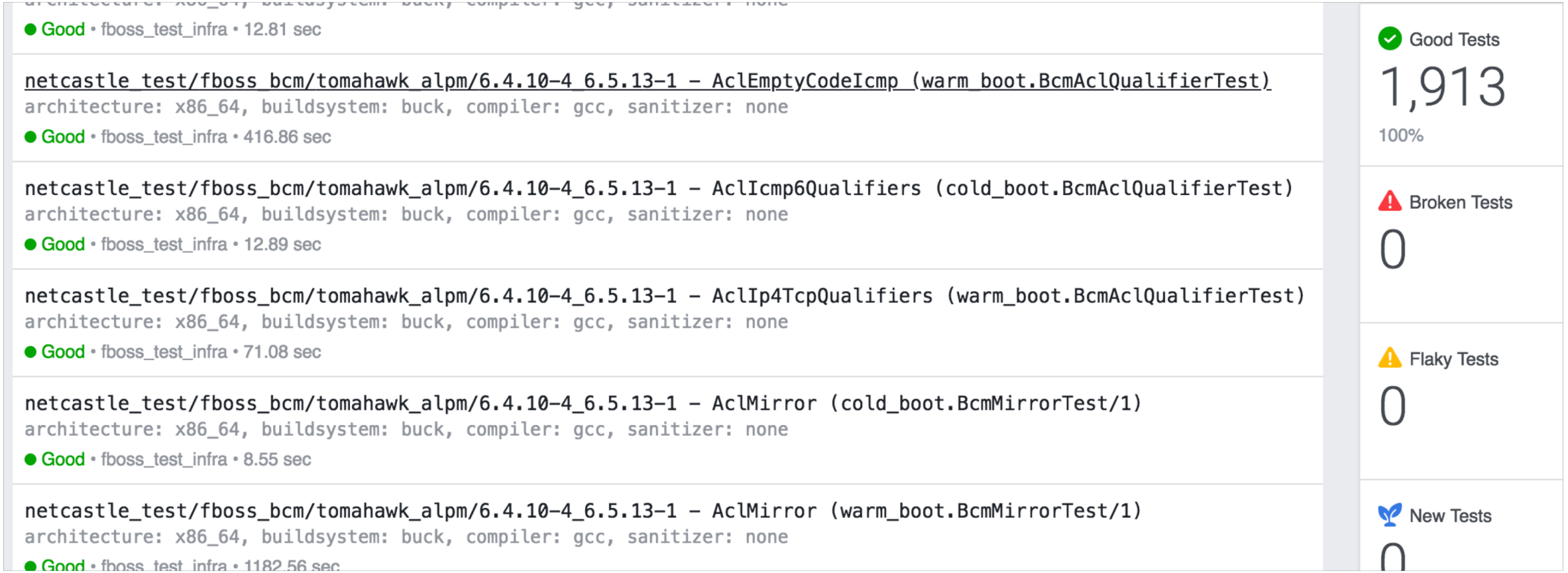
Invest in automation

- Automate as much as possible of our lab testing
- Develop targeted ASIC data plane tests



Test specific asic functionality, verify, warm boot

Utilize common FB testing infra



Takeaways

- Minipack is a powerful modular building block for our networks.
- Software support for modularity has its challenges.
- We were able to overcome these challenges through improved hardware layer abstractions and investing heavily in early deployment and automated testing.

Thank you


Open. Together.
OCP Global Summit | March 14–15, 2019
OCP
SUMMIT