The 448G Generation from a Network System Vendor's Perspective

David Ofelt – Sr. Distinguished Engineer

2025-04-15

JUNIPER NETWORKS | Driven by Experience™

Architect's View of 200Gb/s links

Architect's View of upgrade to 400Gb/s links

The Details Don’t Matter!*,**

* To many of us

** For the near term

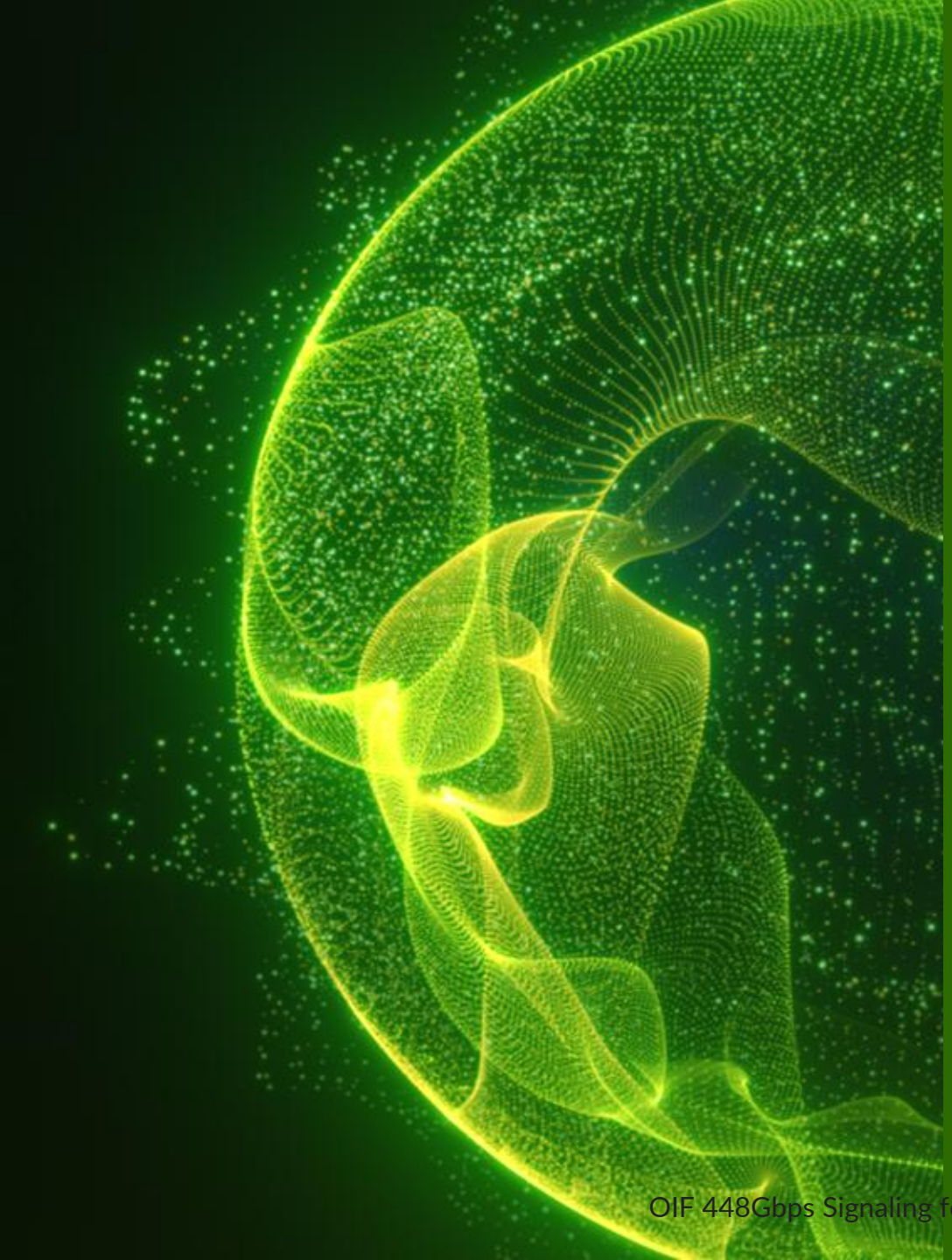
What Are Some Things That Do Matter?

- What are the implications of 400Gb/s signaling on our system construction?
- What system changes will be visible to the end user?
- What assumptions can/need to be changed?
- What use cases need the technology and when?

And …

Of Course We Need Details :)

- Many, many more in-depth talks over the next two days on this topic!



Evolution…

System Evolution

SONET Focused

Ethernet Focused

OC48C (2.5G)

OC192C (10G)

OC768C (2.5G)

100GbE

400GbE

800GbE

1998 | 2000 | 2002 | 2007 | 2009 | 2009 | 2011 | 2011 | 2013 | 2015 | 2019 | 2021 | 2024

Time

M40 | M160 | T640 | T1600 | MX,SRX | TX Matrix+ | T4000 | PTX5000 | MX2000 | QFX,PTX,MX | QFX,PTX,MX | MX10K8 MX304 | PTX10k

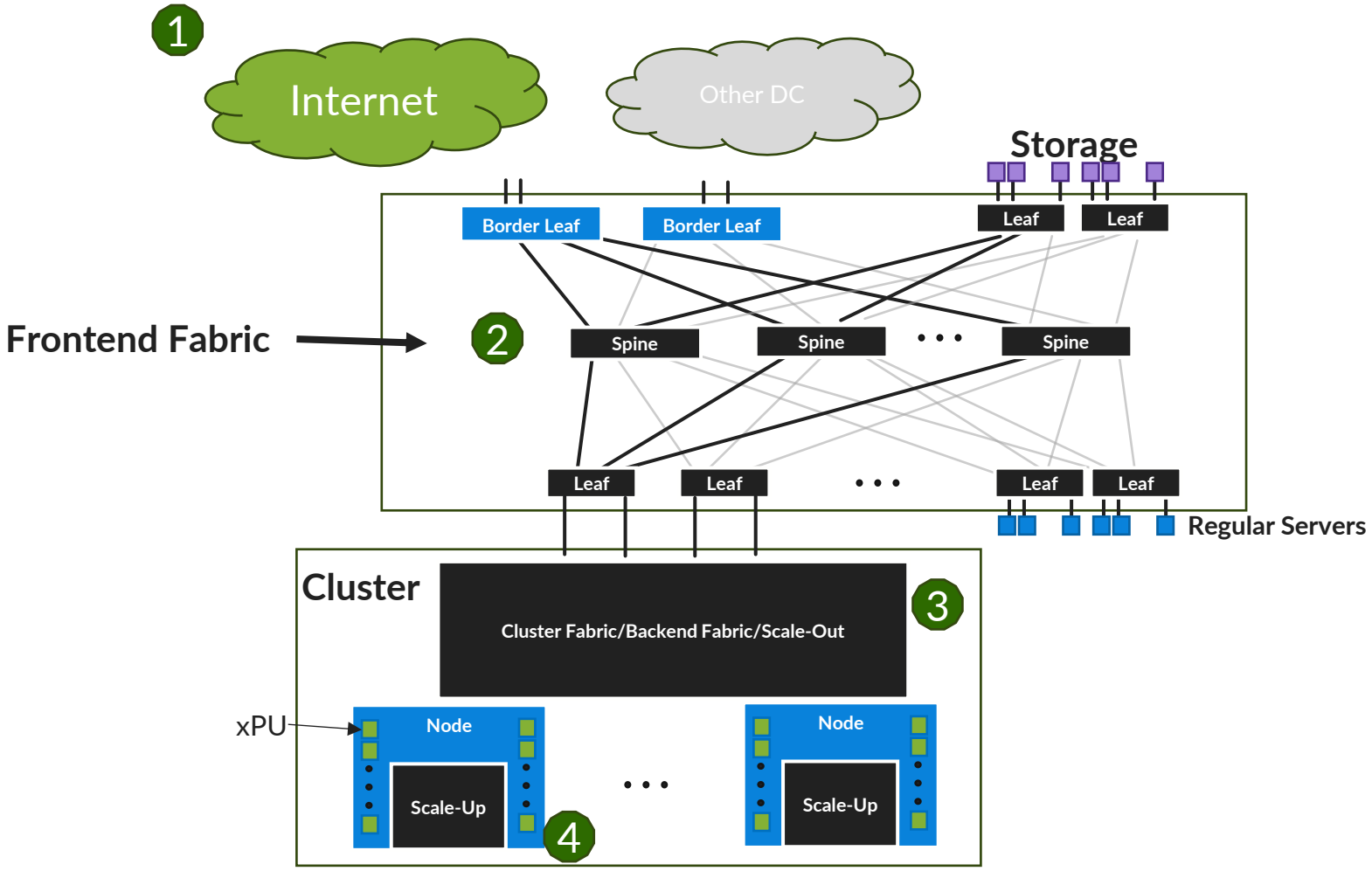
Powered by Juniper ASICs – >10,000x Line Card Bandwidth Improvement!

Martini 250nm	D4/DX	Gimlet 180nm	Stoli 130nm	Trio 65nm	SF 65nm	Trio2 Stinger 40nm	Express 40nm	Trio3, 4 Express2 28nm, HMC	Express3,4 Trio5 16nm, HBM	Trio6 7nm	Express5 7nm

Changes Over Time

- Signaling scaled from 125MHz parallel, through source synchronous, to >100Gb/s SERDES
- Links went from not needing FEC, to using FEC, to having retry capabilities
- Intra-box interconnect started with routed PCBs (backplane/midplane) to ortho-direct connectors
  - and now in some cases- cables
- Becoming more difficult to support forward compatibility over multiple generations
- Packages scaled from small with few high-speed signals to massive with many hundreds of SERDES
- Power scaled from ~10W ASIC to 1500W package
  - With a large amount of the power in inter-die and inter-package links

Network Evolution



Market Evolution

Telco Service Provider

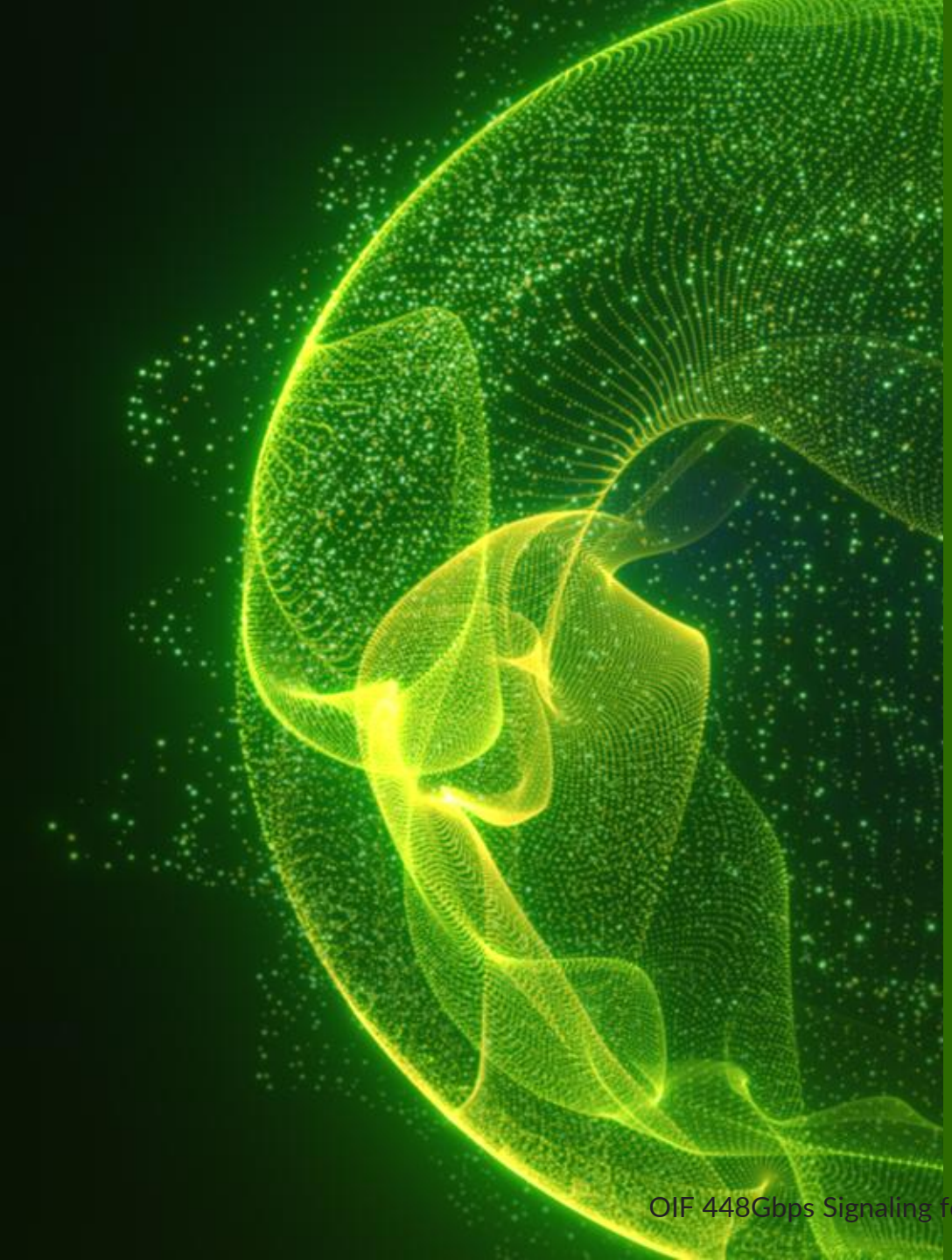
- Boxes Live in network a long time
- Slow deployment ramp
- Upgraded gradually
- Forward/backwards compatibility
- Long Deployment Process
- Large diversity in deployed interfaces

Hyperscaler

- Data Center
  - More homogeneous
  - Upgraded in chunks
  - Shorter Lived
  - Fast deployment ramp
  - Semi proprietary
- Backbone/Core
  - More like classic telco

AI! ML! HPC!!

- Very homogeneous
- No partial upgrade needed (?)
- Single-generation boxes
- System may be single vendor
- System may have proprietary pieces
- Very aggressive technology use
- Not all technologies needed
- Rapidly evolving partitioning
  - Scale-up!
  - Scale-out!
  - …



Questions…
(so many questions)

Ask The Right Questions

- You'll get answers to only the questions you ask – so ask the right ones!

Q: "Do you care about latency? If so, what latency is important?"

A: "Yes! As low as possible!!!"

- vs –

Q: "What facets of your network are important for which applications?"

A: "Resilience, Reach, Beachfront, Power"… "Latency can be traded off for resilience, reach and power"

Source: Ethernet Alliance TEF 2024 Keynote – Ethernet – The foundation of AI @ Meta, Nic Viljoen

System Questions!

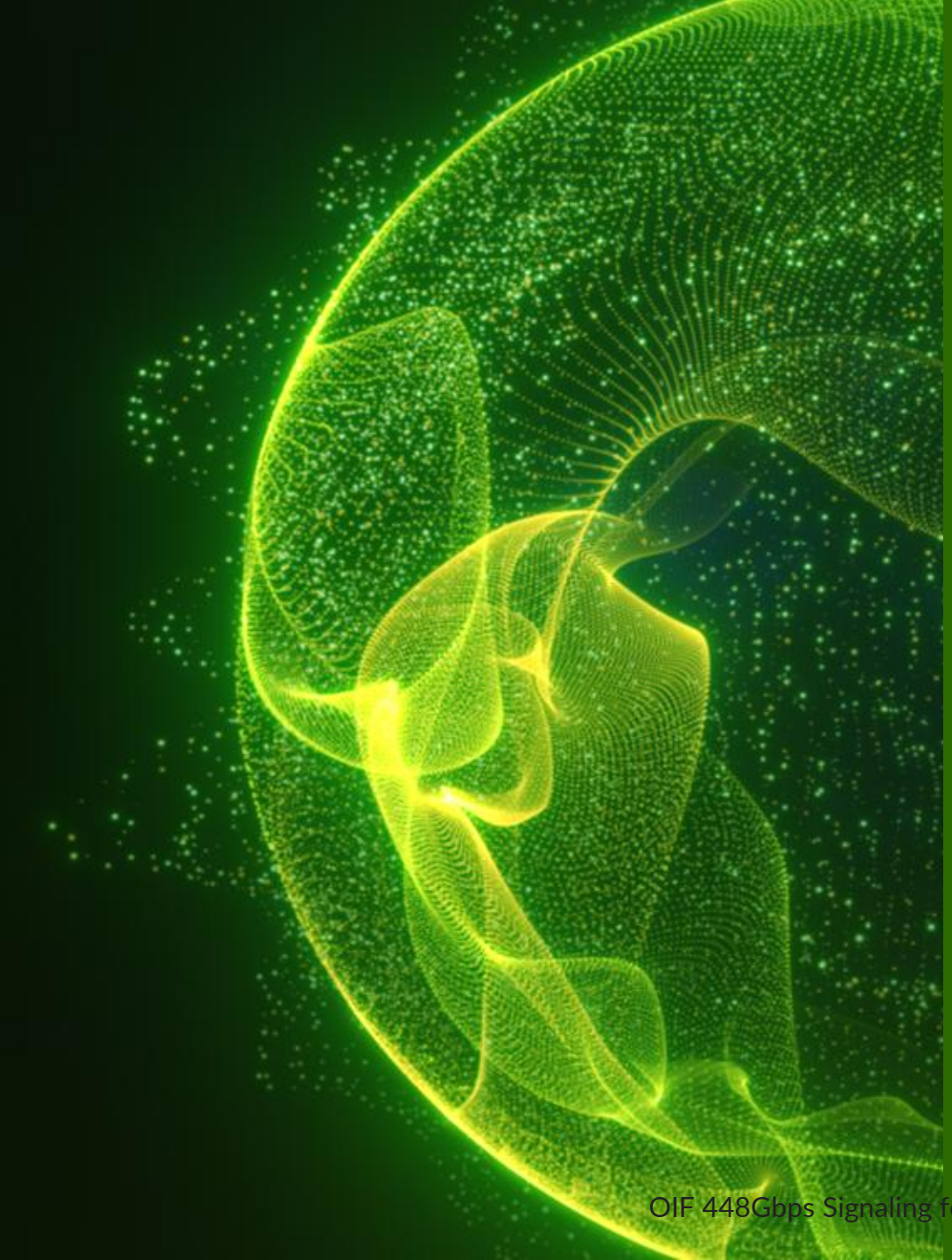
- What do the channels look like?
  - Does PCB work? If so- How many retimers required? Or is everything cabled?
  - Co-packaged copper the only way to get out of the package?
- How do we power and cool these systems?
  - Is liquid-cooling the only viable approach?
- Are pluggable modules supported?
  - Can I re-use existing form factors (with upgrades) like SFP, QSFP(-DD), OSFP?
  - Do I have to invent a new module type?
  - Or is CPO or NPO the most viable future?
- Is it possible to build a multi-generational chassis?
  - If I build a 200Gb/s-based system now- can I upgrade it to 400Gb/s?
  - How about the one I built for 100Gb/s signaling?
- Do I need to decompose the chassis?
  - Modules provide easier multi-generational interop

End-user questions!

- If current modules form factors don’t scale- what is the least bad answer for the customer?
  - Will depend on the use case/network
- How do current chassis evolve?
  - Is there any need for them to?
- What use cases need flexibility and interface diversity?
  - Scale-up and scale-out networks are embedded and appear to have more homogeneous requirements
- Is upgradability important?
- Is there enough benefit to optimize the solution for each sub network?

Summary

- Need to quickly figure out what we can and can’t do in the 400Gb/s generation
- We will need to do All The Things (eventually)
- Need to communicate to the end users about any changes in experience ASAP
- Need to find leverage in any unique facets of the newer network use cases
- Time between technology generations feels like it is getting compressed
- Unclear how long we can keep hitting things with a bigger and bigger hammer to make them work



Thanks!