

# The Impact of Scalability from a Systems Perspective

Joel Goergen / Subi Krishnamurthy  
Force10 Networks

November 2006  
Dallas, Tx

# Supporters

- Mike Bennett, LBNL
- Jim Deleskie, VSNL International
- Dan Dove, Procurve Networking by HP
- Toshinori Ishii, Internet Multifeed Co
- Mark Nowell, Cisco
- Lane Patterson, Equinix
- Henk Steenman, AMS-IX
- Satoru Tsurumaki, Softbank BB
- Masato Yamanishi, Softbank BB

# Introduction

- What is meant by scalability?
- Ethernet is scaleable already
  - 10Mb/s > 100 Mb/s > 1000 Mb/s > 10 Gb/s > HSSG
- End Users need additional capacity
  - Scaleable from 10G to some higher speed and points between?
  - From 10G to a Next Speed Jump?
- Need to consider impact on –
  - The End User
  - The System Vendor
  - The Component Vendors

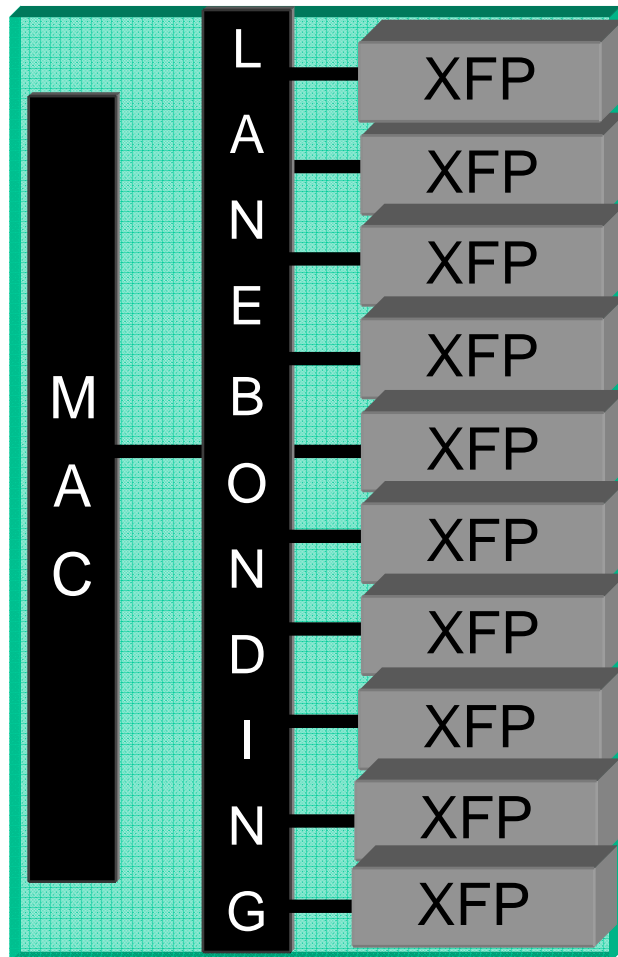
# Scaleable from 10G to Some Higher Speed and Points Between

- Link Aggregation of existing 10G Modules
- Physical Aggregation of existing 10G Modules
- Design New Modules
  - Module 1 - Target some specific rate between 10G and a higher speed
  - Module 2 - Target the entire range by changing # of lanes

# End User Input on LAG

- Using 802.3ad Link Aggregation
- Temporary fix for increased bandwidth demand
- Increased complexity
  - Difficult to plan for capacity and traffic engineering
  - Harder to manage & troubleshoot multiple physical links based on a single logical interface
  - Cable & link management
- Uneven distribution of traffic
  - Limitations in the standard
  - Inefficient distribution of large flows
  - Load balancing requires packet inspection or other knowledge
- Per Mike Bennett (bennett\_01\_0906): “Bottom Line: simplicity = lower operating costs”

# Physical Aggregation of Existing 10G Modules



- Similar to 802.3ad Link Aggregation
- For End Users
  - Overcomes statistical performance problems of 802.3ad Link Aggregation
  - Doesn't overcome complexity associated with multiple fibers and other related issues
  - Port density limited by 10G modules

# New Module Design 1

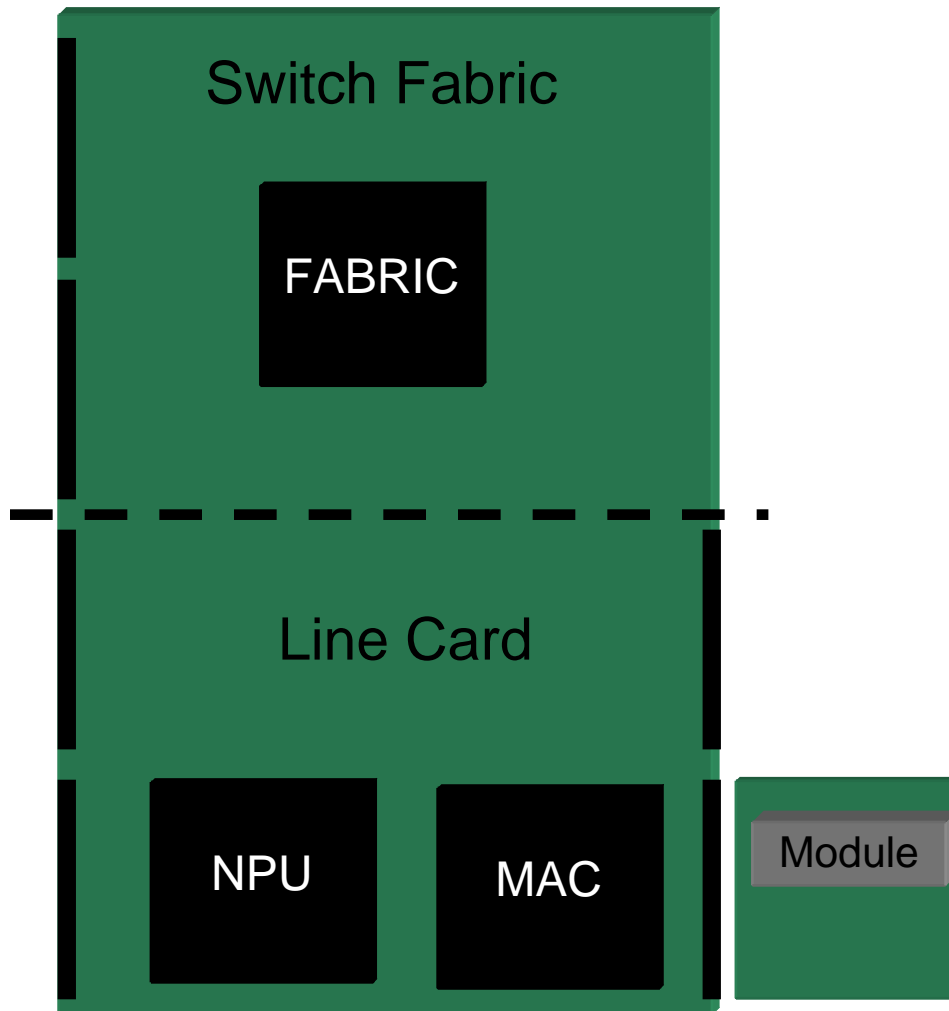
- Proposal - Target some specific rate between 10G and a higher speed
- Observation - Remember 10GbE and the X-Wars?
  - 300 pin MSA, 200 PIN MSA, XENPAK, XPAK, X2, XFP, SFP+
  - Why wouldn't it happen again?
  - Multiple flavors of modules to support

# Impact of New Module Design 1

- End Users
  - Impacts vendor availability
  - If multiple speeds exist, increased cost, complexity, and administration of fiber management
  - Mike Bennett, “Bottom line: simplicity = lower operating costs” (see bennett\_01\_0906)
- System Vendors
  - Cost optimization for Rates
    - Design different boards to support different speeds?
    - If MAC is scaleable
      - Design for a specific rate?
      - Design for a range?
    - Line card rate specific designs?
    - Backplane design for supported speeds?
  - Verification testing of each speed offering
  - Component qualification
- Module Vendors
  - Multiple flavors of modules to support



# New Module Design 2



- Modules supporting different rates by adding lanes (programmable or hardwired) could be put in a common form factor determined by the maximum capacity
  - Doesn't allow port count of lower capacity modules to be maximized
  - Because of different speed options, same issues as noted on fixed rate modules
  - Multiple modules to support
- Backplanes / Line Cards / Fabrics
  - Cost, design driven by maximum capacity
    - PCB Boards / layers
    - Connectors
    - Devices / port integration
    - Thermal management
  - Performance driven by maximum capacity
- CAPEX minus optics driven by max port capacity
- Reducing optics cost doesn't significantly reduce port cost.

# Impact of New Module Design 2

- End Users
  - Impacts vendor availability
  - If multiple speeds exist, increased cost, complexity, and administration of fiber management
  - Mike Bennett, “Bottom line: simplicity = lower operating costs” (see bennett\_01\_0906)
- System Vendors
  - Port density on card edge based on maximum capacity
  - Cost optimization for lower rate cards constrained
  - Verification testing of each speed offering
  - Component qualification
- Module Vendors
  - Multiple flavors of modules to support

# Adding Capacity by a Speed Jump

- Users need something beyond LAG
  - Decreased cable management and complexity
  - Per Mike Bennett (bennett\_01\_0906): “Bottom Line: simplicity = lower operating costs”
- Encourages focus on a given space
  - Example (based on 100G) optical / electrical research
    - 10 lambda / lanes of 10 Gb/s
    - 5 lambda / lanes of 20 Gb/s
    - 1 lanes / 100 Gb/s
- Scalability from one speed jump to another has proven key to Ethernet and will (eventually) happen again

# Conclusions

- Choose a single 100G MAC Rate
- Support different PHY solutions
  - Lambda \* lane speed combinations
- Scalability - Ethernet has a tradition
- Scalability from 10G to some higher speed and points between?
  - Does not address End User expressed needs
  - Complicates system design / testing
  - Multiple flavors of modules to support