NGBASE-T$_{SR}$

A Scalable BASE-T Approach

Dan Dove

Sr Director of Technology

Applied Micro

Supporters

- Steve Carlson – High Speed Design
- Robert Wagner - Panduit
- Paul Vanderlaan - Nexans
- Alan Flatman - LAN Technologies
- Andrew Jimenez - Anixter
- Pedro Reviriego - Universidad Nebrija, Madrid, Spain
- Brad Booth – Dell
- Valerie Maguire – Siemon
- Bernie Hammond - TE Connectivity
- Mike Bennett – LBNL
- Ron Nordin – Panduit
- Yakov Belopolsky – Bel Stewart
- Scott Kipp - Brocade

High Level Description

New approach to Data Center Twisted-Pair Networking

- Define cable parameters for key applications
  - Rack to Top of Rack Reach (5-10m?)
  - Rack to End of Row Reach (20-30m?)
  - **M**aximum **P**ractical **R**each [aka: **MPR**] (XXm?)
- Define Auto-Negotiation to allow shorter-reach-only PHYs
  - Lower TX and RX power required, can reduce AFE
    - As CMOS steps reduce, AFE dominates
- Define *optiona*l mechanism for PHY to back-down advertisement based on self-determination that link exceeds capability
- Common signaling with defined functional reductions
- Compatibility between PHYs of different reach as long as link meets minimum criteria of both PHYs
- Rather than define MPR only, let market decide where volume belongs
- Identify a 10GBASE-TSR approach to allow compatible TOR and EOR solutions at much lower power

Do's and Don'ts

What this presentation does

- Assert that twisted-pair PHYs at very high data rates require relatively high power (compared to fiber/DAC) to achieve **M**aximum **P**ractical **R**each.
- Assert that shorter-reach applications can benefit from alternative PMD requirements to advance market adoption
- Assert that compatibility between longer/shorter reach PMDs has value to the market
- Identify key applications that can benefit from this approach
- Assert that 10GBASE-T could potentially benefit from this approach

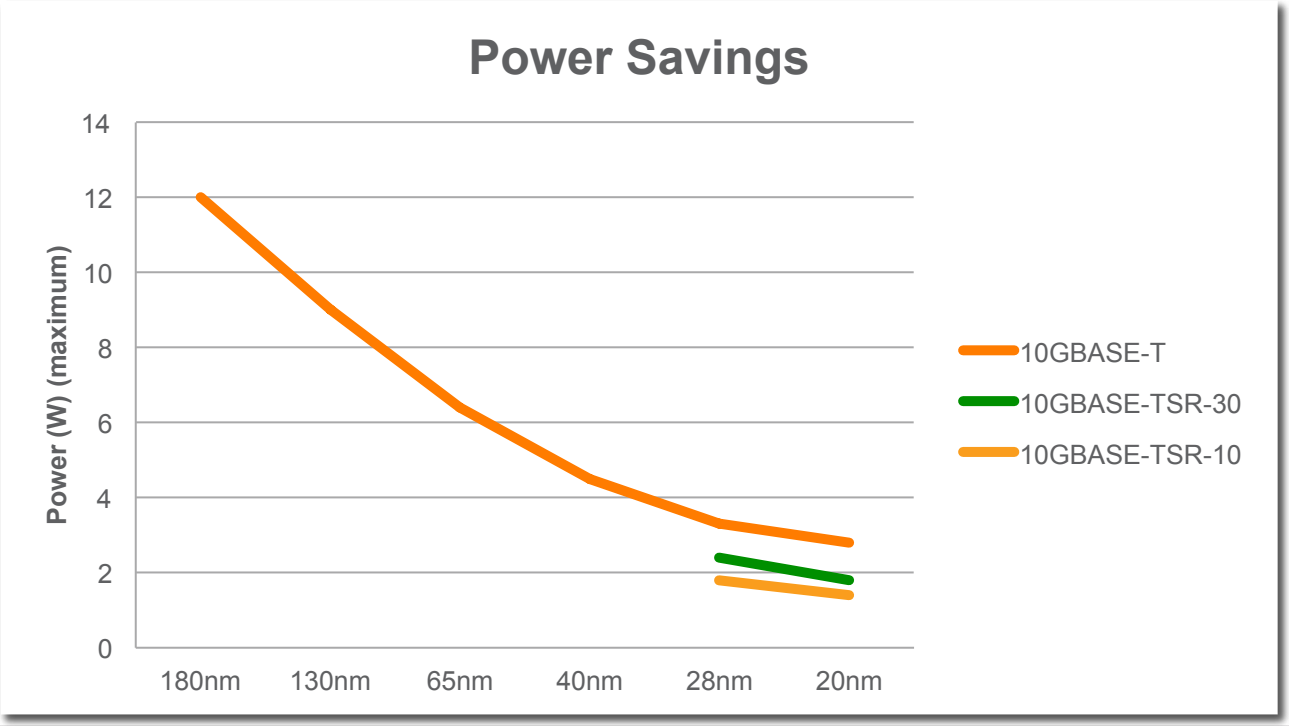
What this presentation doesn't do

- Assert absolute values for reach, this is an SG activity
- Assert how to implement a lower-power NGBASE-T PHY

Why Do We Need This?

Using the example of 10GBASE-T, complex encoding, DSP, and significant SNR challenges lead to a very high-power (relative to alternative 10G solutions).

Proposed approach allows power envelope to track application requirements. Shorter reach applications demand higher density.



Note: Key difference in “TOR Reach” power reduction due to AFE simplification

Application Models (TOR Reach 5-10m)

- Top of Rack Server-Switch
- Maximum 5-10m required
- Patch cord only solution acceptable?
- Shielded Cable acceptable
  - Simplifies EMI design
  - Reduces Echo/Next challenges of multiple connectors
  - Reduces TX power requirement
  - Virtually eliminates ANEXT
- Server likely to use MPR due to flexibility requirements
- TOR switch likely to use 5-10m to reduce power, cost and increase density of BASE-T solutions
- Reduced power allows higher density TOR switches

Application Models (EOR Reach 25-30m)

- End of Row Server-Switch, Switch-Switch
- Maximum 25-30m required
- Two Connector Model Required?
- UTP Required?
  - More complex EMI design
  - Requires Echo/Next challenges of multiple connectors
  - Increases TX power requirement
  - Requires attention to ANEXT
- Server likely to use MPR due to flexibility requirements
- TOR/EOR switch likely to use 25-30m to increase flexibility at optimized power, cost and density

Application Models (MPR)

- End of Row Switch-Core Switch
- Maximum Practical Reach Required
- Four Connector Model Required?
- UTP Required?
  - More complex EMI design
  - Requires Echo/Next challenges of multiple connectors
  - Increases TX power requirement
  - Requires attention to ANEXT
- Server likely to use MPR due to flexibility requirements
- MPR must offer a cost effective solution relative to fiber
  - Still offers an advantage of backward compatibility to lower speeds
- MPR must offer a reasonable power consumption to be acceptable

Auto-Negotiation

A tool to enable “Application Specific Reach”

Auto-Negotiation Of Reach

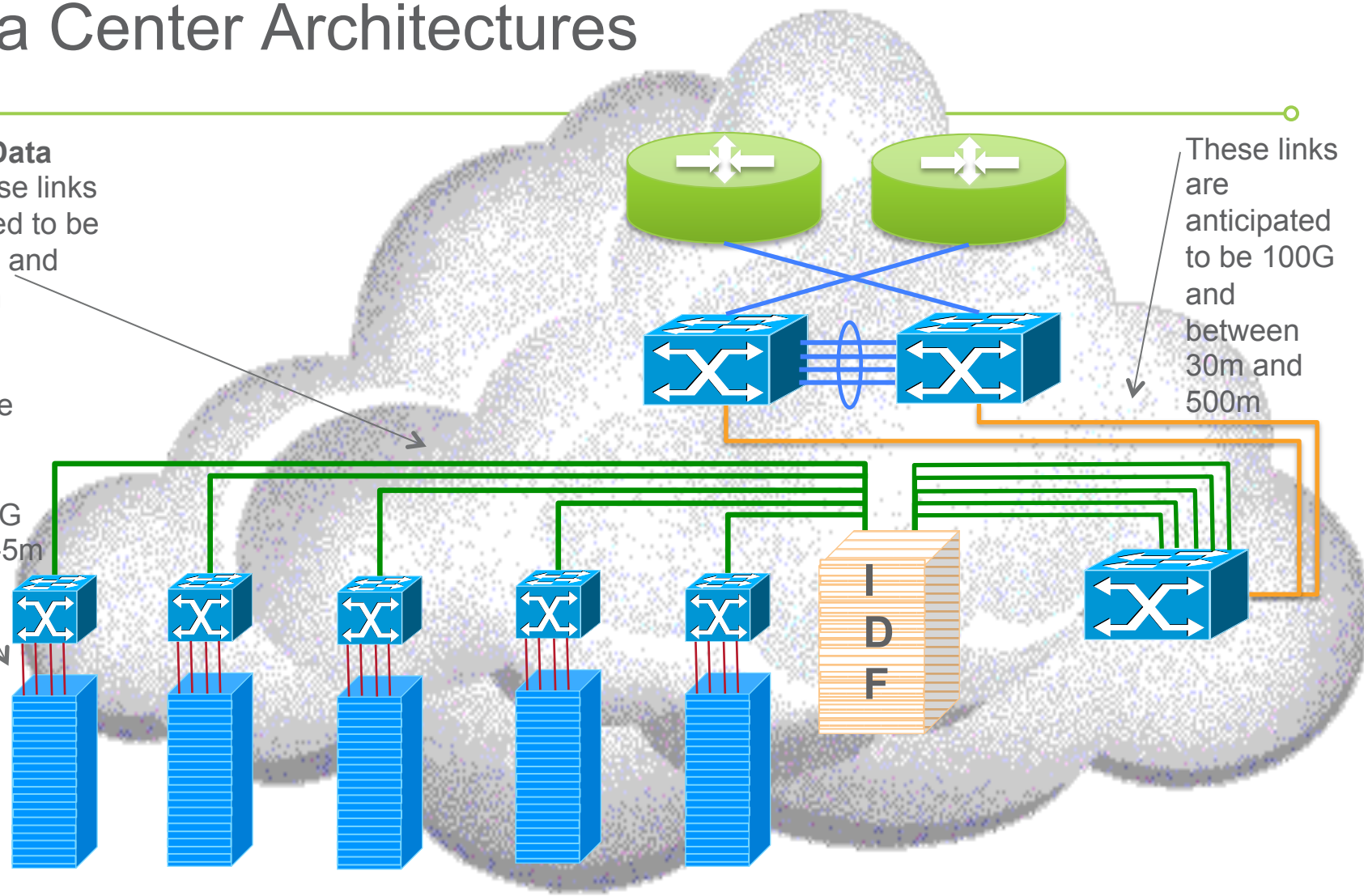
- Advertise all possible supported reaches and speeds
- Add a bit in AN to signal “Link Not Supported (LNS) at higher speed”
- Use proprietary method to evaluate link
  - DSP based PHYs typically have this ability
  - Can assess noise, length, NEXT, FEXT, prior to link establishment
- If link not supported by PHY, assert LNS bit, reduce advertised speed
  - Eliminates link-toggling and/or infinite loop on link establishment
  - Eliminates need for proprietary speed-dropping algorithms
    - Some algorithms lead to toggling which upset higher layer protocols like spanning tree
  - Provides clear information to management entity on why a lower speed was arrived at

Data Center Architectures

Traditional Data Centers, these links are anticipated to be 40G or 100G and 25m to 100m

Lot of links
Cost sensitive

These links moving to 10G
Typically 3m-5m
Very Cost Sensitive



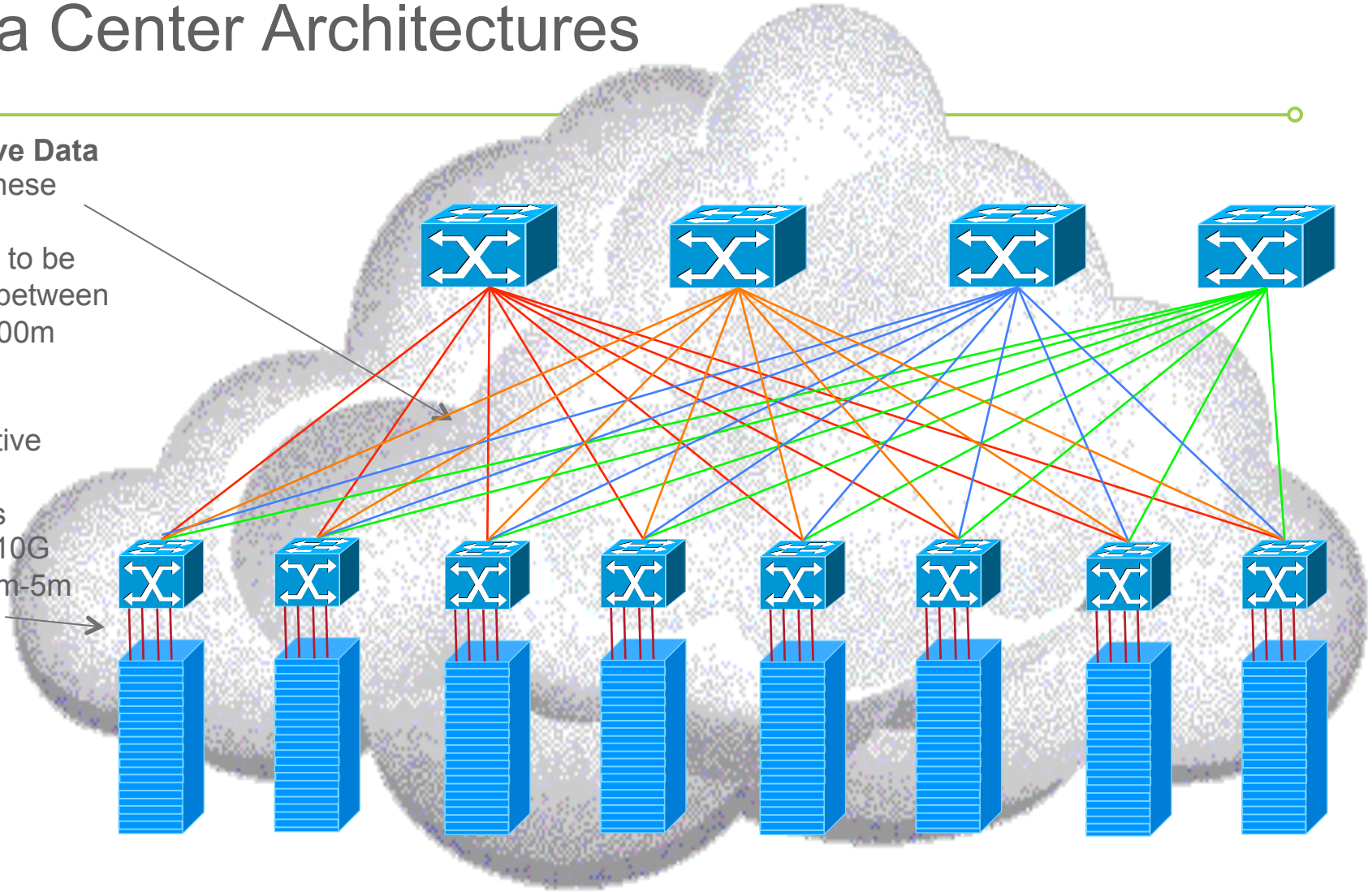
These links are anticipated to be 100G and between 30m and 500m

Data Center Architectures

For **Massive Data Centers**, these links are anticipated to be 100G and between 30m and 500m

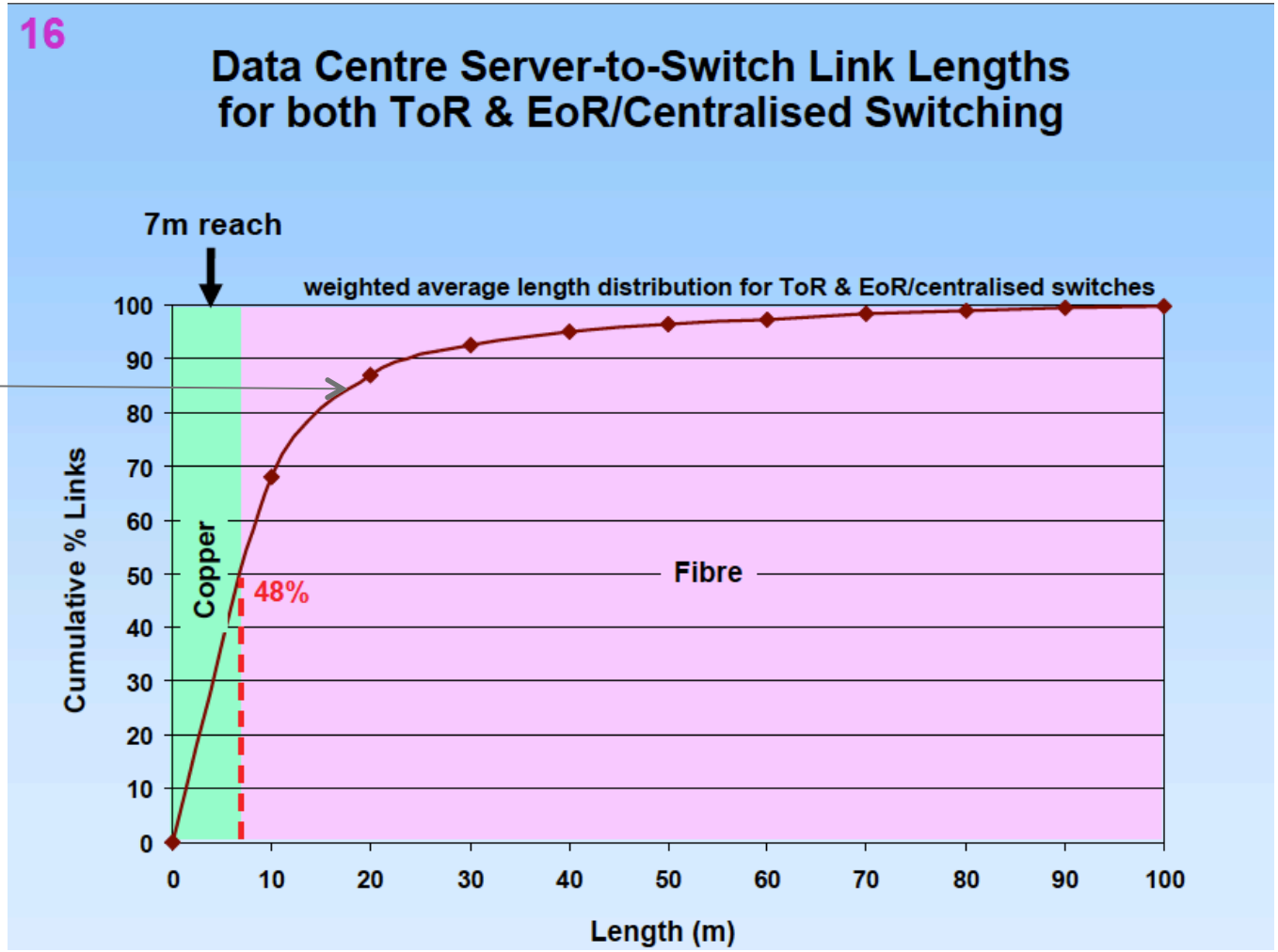
Lot of links
Cost sensitive

These links moving to 10G
Typically 3m-5m
Very Cost Sensitive



nGBASE-T$_{SR}$ Addresses > 85% of The Market

85% or more server/switch ports are < 20m

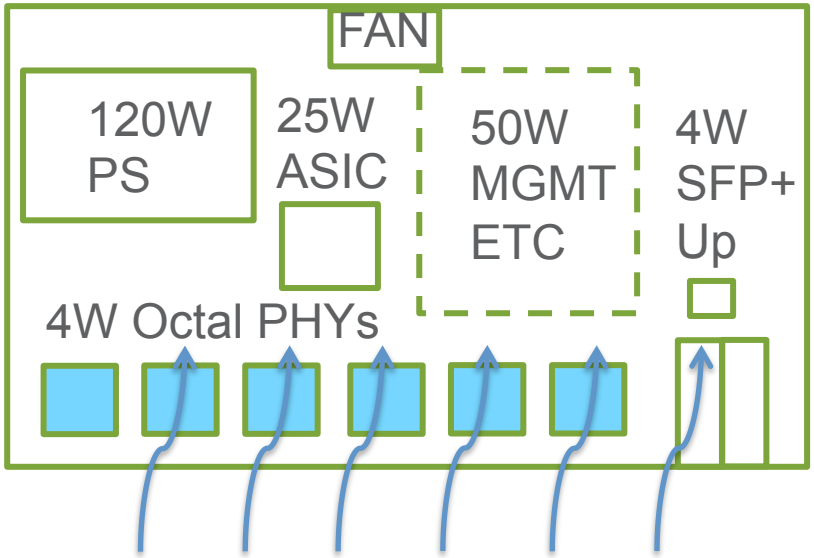


Ref: flatman_01_0311_NG100GOPTX.pdf

Why is Power Such A Critical Parameter? (1)

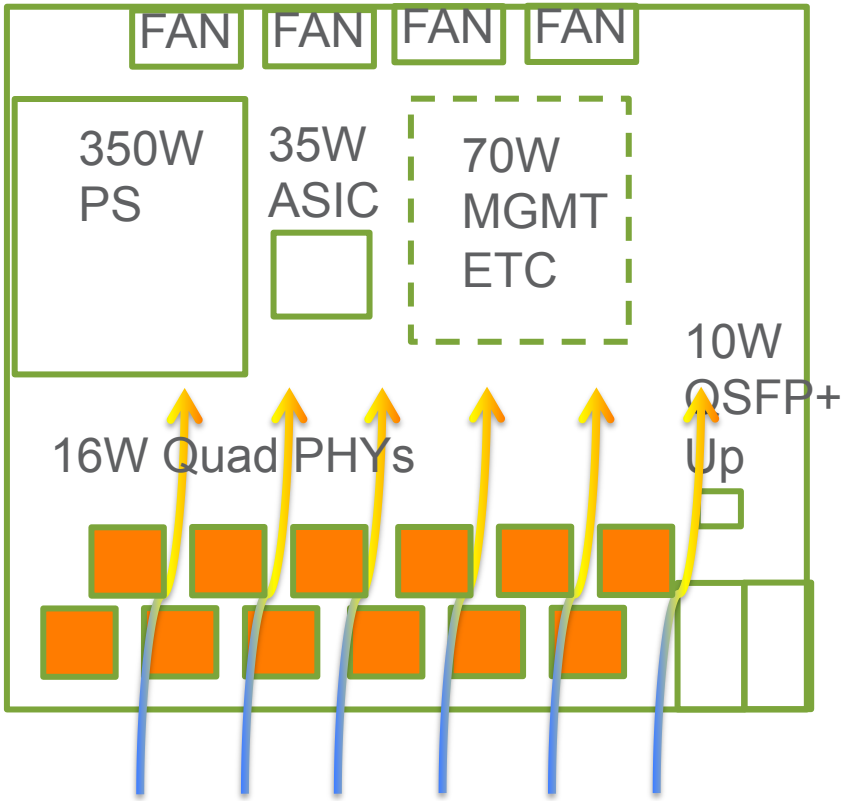
Power Consumption ~ Thermal Management has a non-linear impact on both cost and density for high port-count switches.

48 port 1000BASE-T Switch
.5W of PHY, 120W of Thermal Mgmt



Air Flow Slow, Easy

48 Port 10GBASE-T Switch
4W of PHY, 350W of Thermal Mgmt

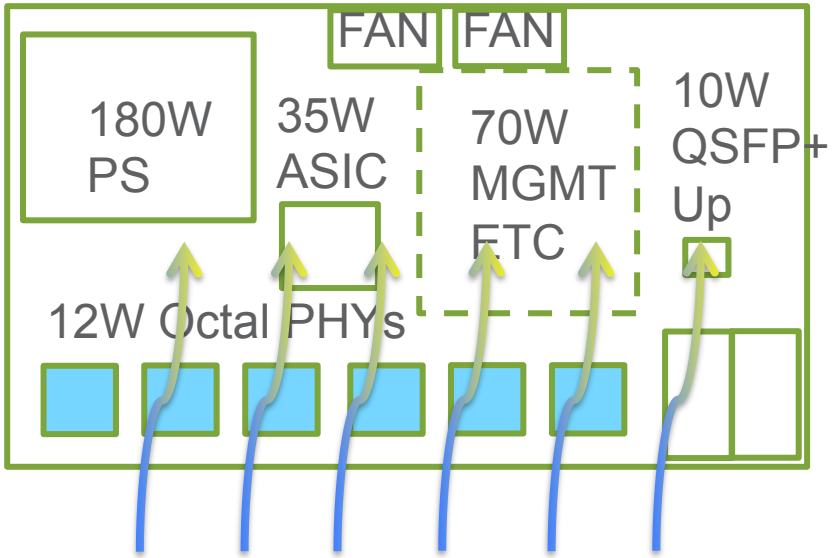


Air Flow FAST, Difficult

Why is Power Such A Critical Parameter? (2)

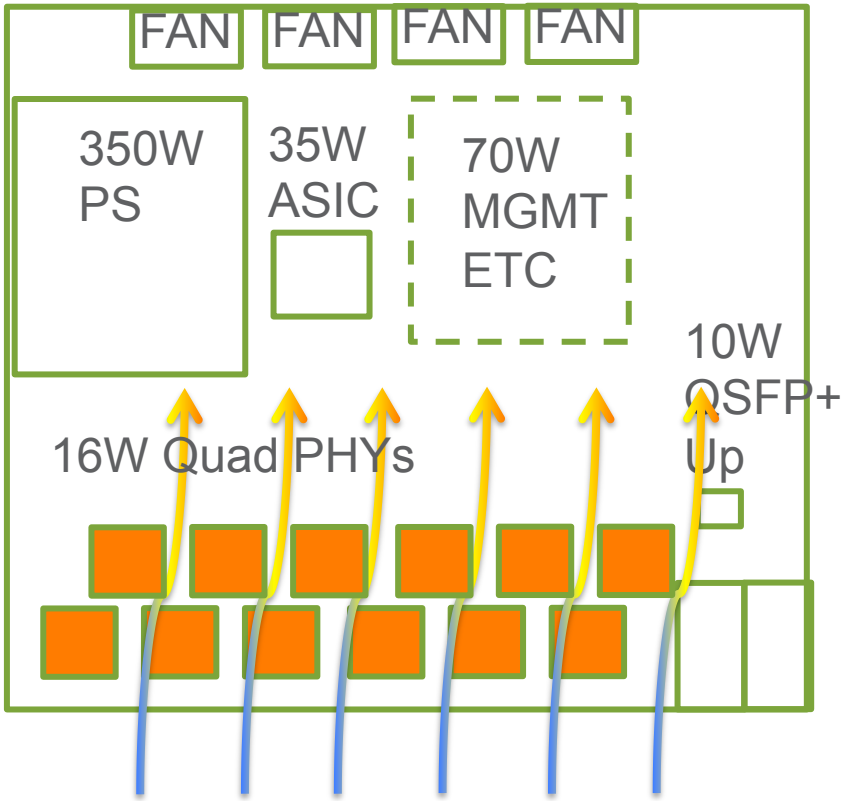
Power Consumption ~ Thermal Management has a non-linear impact on both cost and density for high port-count switches.

48 port 10GBASE-TSR Switch
1.5W of PHY, 180W of Thermal Mgmt



Air Flow Medium, Moderate

48 Port 10GBASE-T Switch
4W of PHY, 350W of Thermal Mgmt



Air Flow FAST, Difficult

Conclusions

We need a new approach to BASE-T PMDs

- Define key applications and necessary reach
  - Don't assume "one size fits all".
- Define Auto-Negotiation approach to allow reach-optimization for power, cost, complexity
- Define Auto-Negotiation approach to allow back-down and communicate that decision to link-partner
  - Note: Existing 10GBASE-T PHYs may be upgradeable via F/W
- Allow market to drive implementations
  - Some applications will take lion's share of volume but those applications are typically the most cost sensitive
  - Some applications will demand MPR, and if there is sufficient volume, implementations will arise to address them.

The result will be faster adoption in applications that demand lower power, cost and higher density.

Thank You