NGBASE-T$_{SR}$

“Application Specific PMDs That Interoperate”

Response to Concerns

Dan Dove

Sr Director of Technology

Applied Micro

Supporters

- Steve Carlson – High Speed Design
- Alan Flatman - LAN Technologies
- Andrew Jimenez - Anixter
- Pedro Reviriego - Universidad Nebrija, Madrid, Spain
- Brad Booth – Dell
- Valerie Maguire – Siemon
- Mike Bennett – LBNL
- Yakov Belopolsky – Bel Stewart
- Scott Kipp – Brocade
- AndrewJimenez – Anixter
- Stephen Bates - PMC-Sierra

High Level Description (review)

New approach to Data Center Twisted-Pair Networking

- Define cable parameters for key applications
  - TOR, EOR, Uplinks
- 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
- **Common signaling** with defined functional reductions
- **Compatibility between PHYs of different reach** as long as link meets minimum criteria of both PHYs

- NGBASE-TSR approach to allow compatible TOR and EOR solutions at much lower power

Response to Concerns Raised In Geneva

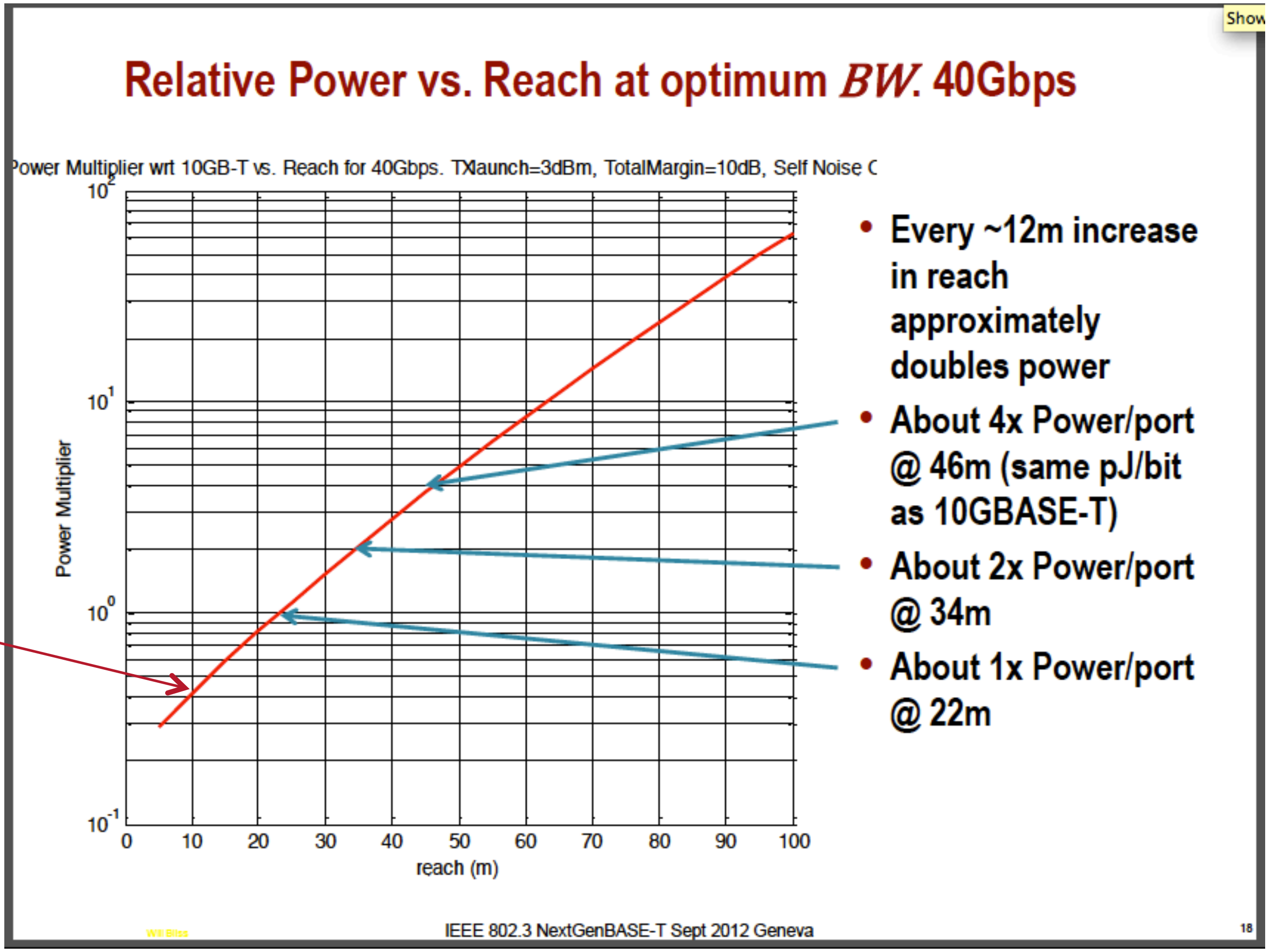
- Power vs Market adoption for 40G devices
- Market adoption for 40G and 10G
- Impact of 10GBASE-TSR on existing 10GBASE-T
- Customer Confusion on reach capabilities
- How Auto-Negotiation works

Power vs Market Adoption

Power Assumptions for 40G Ports

Optimum
Channel
40GBASE-TSR-10
Would be ~1W

Practically speaking
I would expect it to be
at least 2W @ 10m
and 8-9W @ 30m



Relative Power vs. Reach at optimum *BW.* 40Gbps

- Every ~12m increase in reach approximately doubles power
- About 4x Power/port @ 46m (same pJ/bit as 10GBASE-T)
- About 2x Power/port @ 34m
- About 1x Power/port @ 22m

Will Bliss

IEEE 802.3 NextGenBASE-T Sept 2012 Geneva

18

(1) bliss_01_0912 p18 – I am assuming 3W/port for 10GBASE-T (28nm)

Power Budget for 40G Ports (1 of 2)

- Data Center operators are going to install equipment designed to provide optimum price/power/performance for their application needs.
- QSFP+ (40G) may have a power budget of 3.5W, but that applies to optical applications.
- Direct Attach may be driven directly from the switch ASIC with a 500mW premium per channel. (transmitter and equalization power)
- A 1U QSFP+ Switch will typically design for maximum power and density, 36 ports @ 3.5W = 126W of PHY power.

Power Budget for 40G Ports (2 of 2)

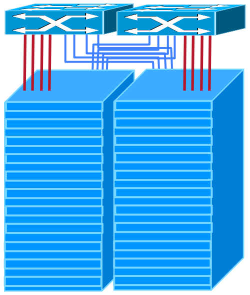
- 1U QSFP+ Switch using DACs will consume ~ 36 *.5W ≈ 18W of PHY power.

- An equivalent 40GBASE-T switch will:
  - require 9W/port(1) at 30m which will limit it to 18 ports in 1U (162W of PHY power)
  - or force a 2U form factor with additional fans and power supply.
  - Less ports/box => Higher cost/port for chassis, PS, fans, etc.

(1) See Slide 6

Power vs Market adoption for 40G devices

96 port switches feed two racks of dual attach



Assuming 40 servers/rack, the 96 port TOR switch thermal requirements are *very challenging*

QSFP = 96 *.5W = 48W of PHY power for passive DAC

40GBASE-T = 96 * 9W = of PHY power 864W for 40GBASE-T

40GBASE-TSR = 96 * 2W = 192W for 40GBASE-TSR [1] when configured for 10m which fits well in a 2U form factor

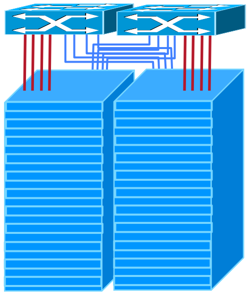
Conclusion: 40GBASE-TSR competes well with QSFP+ for Top Of Rack applications.

[1] See Slide 6

Power vs Market adoption for 10G devices

Assuming 40 servers/rack, the 96 port TOR switch thermal requirements are *very challenging*

96 port switches
feed two racks
of dual attach



SFP+ = 96 *.15W = 14.4W of PHY power for passive DAC

10GBASE-T = 96 * 3W = 288W for 10GBASE-T

10GBASE-TSR = 96 * 1.5W = 144W for 10GBASE-TSR when configured for 10m

Conclusion: 10GBASE-TSR competes well with SFP+ for Top Of Rack applications.

Market adoption for 10G and 40G

Market adoption for 10G and 40G

10GBASE-T suffered a delayed market adoption

- Great Technology, required substantial DSP and analog precision
- 2006 standardization saw 130nm parts at 9W max
- Three geometry spins 130n, 65n, 40n, and power is still ~4W max
- Port Density rendered it non competitive against Direct Attach Cables (DACs) for Top Of Rack switches
- Moore's Law was unable to bring power down sufficiently through geometry shrinks
- 1.5W/port is the trigger point required to over-take DACs

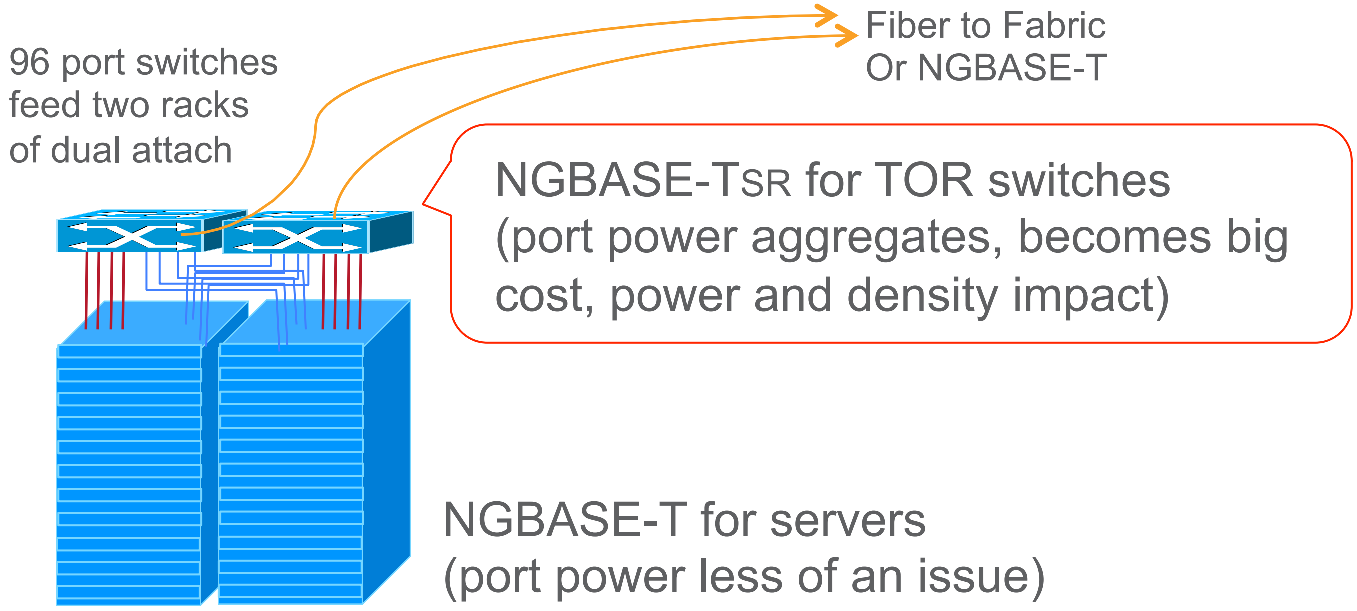
40GBASE-T could suffer a similar delayed market adoption

- The power required to meet maximum practical reach 8-9W[1] will substantially impair port density
- Allow the PHY to negotiate reach (TOR or EOR)
- TOR PHY ~2W could fit in QSFP+ form factor switch
- EOR PHY 9W would not achieve density, but could compete against fiber on cost, cabling, ease of use, backward compatibility

[1] See Slide 6

NGBASE-TSR Use Cases

- NGBASE-T for uplinks
- NGBASE-T for servers
- NGBASE-TSR for TOR switches



96 port switches feed two racks of dual attach

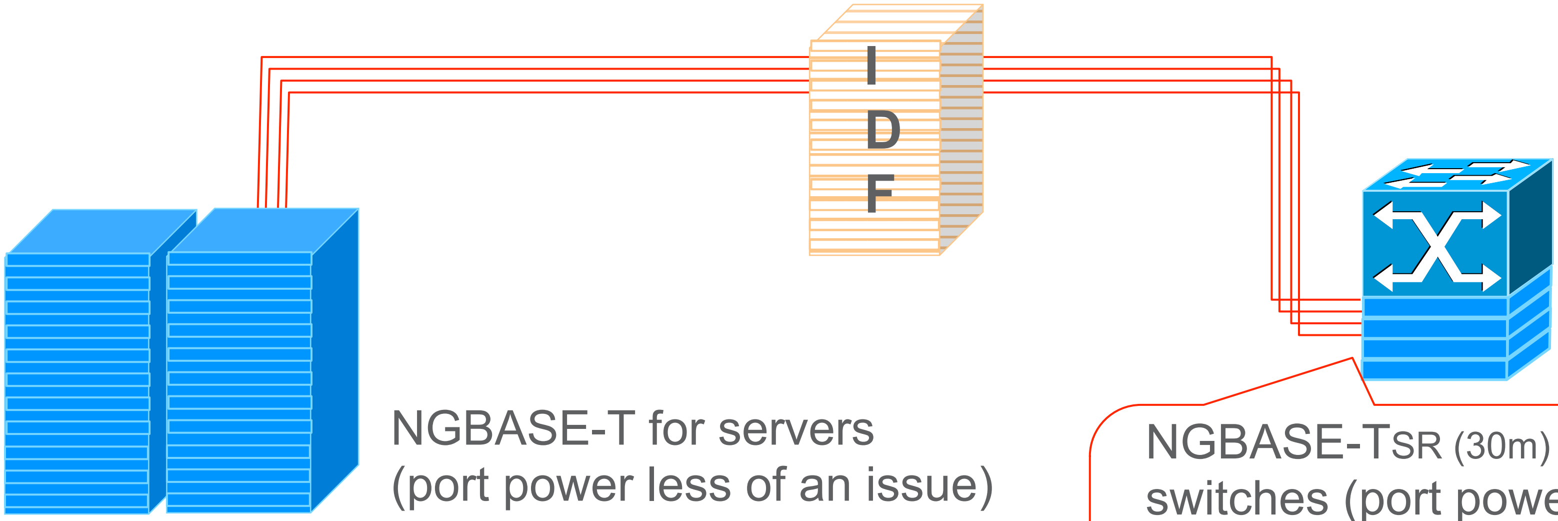
Fiber to Fabric
Or NGBASE-T

NGBASE-TSR for TOR switches (port power aggregates, becomes big cost, power and density impact)

NGBASE-T for servers (port power less of an issue)

NGBASE-TSR Use Cases

- NGBASE-T for uplinks
- NGBASE-T for servers
- NGBASE-TSR for EOR switches



NGBASE-T for servers (port power less of an issue)

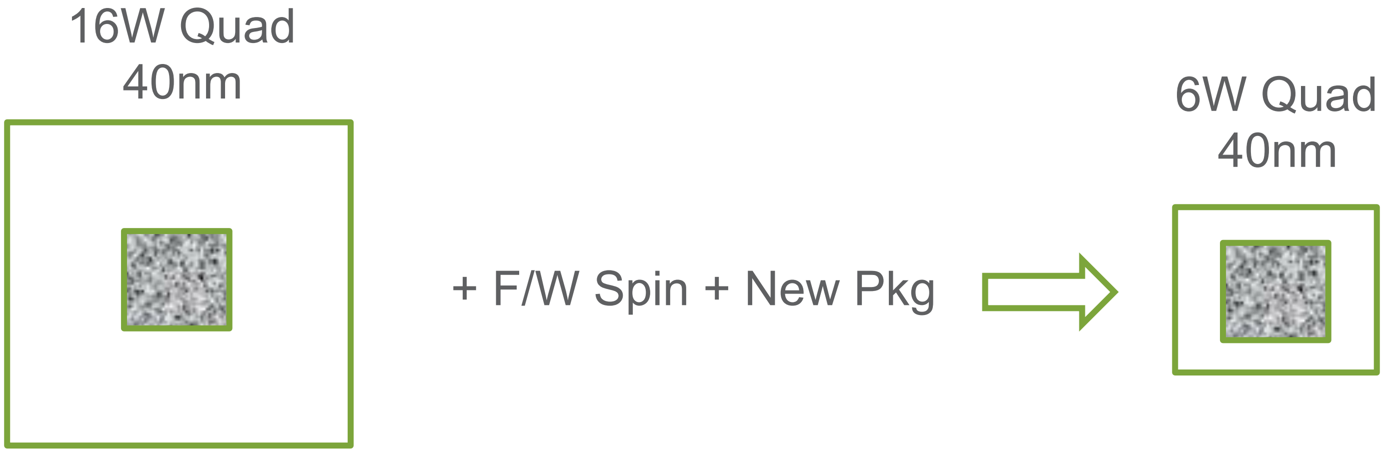
NGBASE-TSR (30m) for EOR switches (port power aggregates, becomes big cost, power and density impact)

*IDF – Intermediate Distribution Frame

Impact of 10GBASE-TSR on existing 10GBASE-T

10GBASE-TSR Market Impact

- 10GBASE-T has now been through 3 generations
- Virtually all PHYs in the market utilize firmware
    - Auto-negotiation parameters subject to modification
    - Training parameters subject to modification
- Existing silicon can be sold as 10GBASE-TSR
    - Firmware revision
    - Smaller, lower cost (or higher port count) packaging
    - Fuse or Firmware lock to prevent over-heating
    - Enables higher density, lower cost switches



16W Quad
40nm

+ F/W Spin + New Pkg

6W Quad
40nm

10GBASE-TSR Market Impact

- System vendors could choose to manage power within their switches by enabling 10GBASE-T PHYs to operate in 100m, 30m, or 10m as power and cooling allows.
  - Switches can utilize thermal measurement
- Uplink ports selectable via management

- 10GBASE-TSR does not cannibalize existing silicon investments, it enables existing silicon into new applications
- 10GBASE-TSR allows lower cost, higher density switches

- 10GBASE-TSR expands the 10G market faster

10GBASE-TSR Interoperability with existing 10GBASE-T

Interoperability with existing 10GBASE-T

- Use cases
  - Compliant 10GBASE-T connected to 10GBASE-TSR
    - If cable > ShortReach, TSR side drops advertised speed to 1000BASE-T, devices link at 1000BASE-T
  - Compliant 10GBASE-T(1)connected to 10GBASE-TSR
    - If cable > ShortReach, both sides drop advertised speed to 1000BASE-T, link at 1000BASE-T.
      - Notify Management of cable incompatibility
      - Switch/Server can flash speed LED or otherwise communicate link incompatibility
  - Compliant 10GBASE-T(1) connected to 10GBASE-TSR
    - If cable < ShortReach, both sides operate at 10G at lower power mode of operation.

*ShortReach values to be defined by SG

(1) F/W updated to TSR

Concerns about Customer Confusion on reach capabilities

Customer Confusion on Reach Alternatives

- Customers understand technologies based on their name.
  - **40GBASE-T**: Maximum Practical Reach (to be defined by Study Group) (0 to 30-45m + IDF?)
  - **40GBASE-TSR**: Reach defined for TOR applications (0 to 5-10m)
    - Highest density demands
    - Lowest power demands
    - Alternative to QSFP+ DACs
    - A winning solution!
- I recommend two reach objectives.

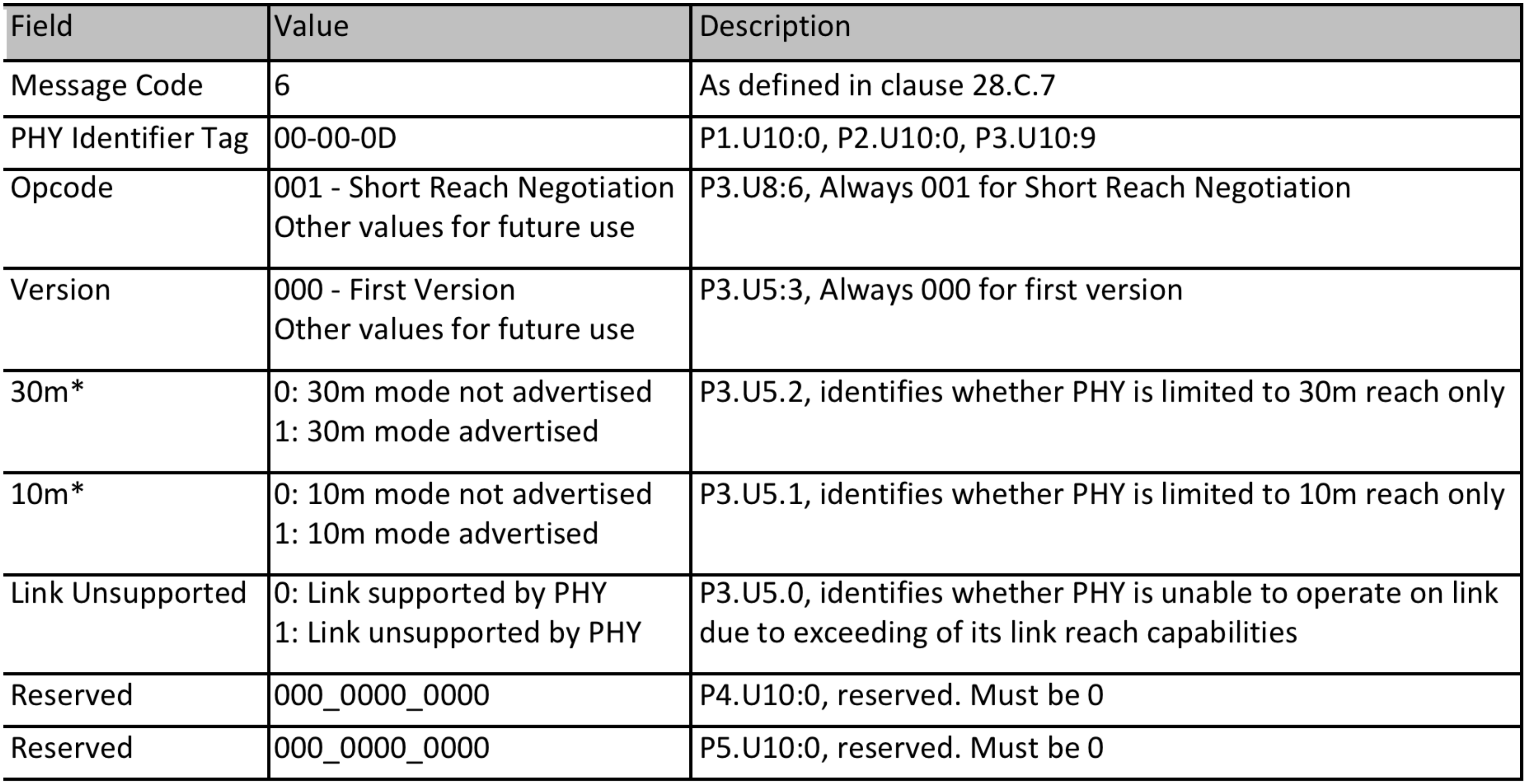
Customer Confusion on Reach Alternatives

- **10GBASE-T**: Maximum Practical Reach (100m + 4 connector)
  - Servers, EOR and uplinks to remote wiring closets
- **10GBASE-TSR30**: Reach defined for EOR applications (0 to 20-30m)
  - Very high density demands
  - Low power demands
  - Alternative to 4W PHYs
  - May be 4W PHY operating in low-power mode
- **10GBASE-TSR10**: Reach defined for TOR applications (0 to 5-10m)
  - Highest density demands
  - Lowest power demands (1.5W/port)
  - Alternative to SFP+ DACs
  - A winning solution!
- I recommend three reach objectives

How Auto-negotiation Would Work

Auto-Negotiation (example – TBD by TF)

Link Partners use Message Code 6 to communicate capability



Field	Value	Description
Message Code	6	As defined in clause 28.C.7
PHY Identifier Tag	00-00-0D	P1.U10:0, P2.U10:0, P3.U10:9
Opcode	001 - Short Reach Negotiation Other values for future use	P3.U8:6, Always 001 for Short Reach Negotiation
Version	000 - First Version Other values for future use	P3.U5:3, Always 000 for first version
30m*	0: 30m mode not advertised 1: 30m mode advertised	P3.U5.2, identifies whether PHY is limited to 30m reach only
10m*	0: 10m mode not advertised 1: 10m mode advertised	P3.U5.1, identifies whether PHY is limited to 10m reach only
Link Unsupported	0: Link supported by PHY 1: Link unsupported by PHY	P3.U5.0, identifies whether PHY is unable to operate on link due to exceeding of its link reach capabilities
Reserved	000_0000_0000	P4.U10:0, reserved. Must be 0
Reserved	000_0000_0000	P5.U10:0, reserved. Must be 0

Example Next Page Definition

* Reach values to be defined by Study Group

Link Assessment (example – TBD by TF)

- “Link Unsupported” bit for Auto-Negotiation would allow graceful decision to drop speed when the PHY determines a link is unsupported.

- DSP based PHYs contain many sophisticated cable assessment abilities
  - ANEXT/EMI assessment
  - NEXT assessment
  - FEXT assessment
  - ECHO/TDR for reach
  - IL assessment

- Assessment can be done during Auto-Negotiation, or Training
- Preferable performance based on AN (reduces decision time)
- No need to standardize method of assessment
  - Mandate assessment must qualify good cables
  - Vendors will ensure their method meets that requirement

* Reach values to be defined by Study Group

Conclusion: We need a new approach to BASE-T PMDs

- Differential in power between TOR application and EOR application mandates different PHY approaches

- Compatibility between TOR and EOR PHYs on an acceptable link is essential.

- 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
  - back-down and communicate that decision to link-partner
    - Reminder: Existing 10GBASE-T PHYs may be upgradeable via F/W

Conclusion: We need a new approach to BASE-T PMDs

- Allow market to drive implementations
  - Some applications will take lion's share of volume but those applications are typically the most cost sensitive
    - TOR switches
    - EOR switches
  - Some applications will demand MPR, and if there is sufficient volume, implementations will arise to address them.
    - Servers, Uplinks
    - PHY vendors can build multipurpose devices and allow system vendors to purpose them as required.

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

Thank You