In support of PSM4 for 100GbE

Presenter: Kapil Shrikhande, Dell

802.3bm Task Force meeting

July 2013

Contributors and supporters

- Tom Issenhuth, Microsoft
- David Warren, HP
- Kapil Shrikhande, Dell
- John D'Ambrosia, Dell
- Oren Sela, Mellanox
- Oded Wertheim, Mellanox
- Piers Dawe, Mellanox
- Rick Rabinovich, Alcatel-Lucent
- Mike Dudek, Qlogic
- Scott Kipp, Brocade
- Andy Bechtolsheim, Arista
- John Petrilla, Avago
- Tom Palkert, Molex
- Brian Welch, Luxtera
- Kiyo Hiramoto, Oclaro
- David Lewis, JDSU
- Arlon Martin, Kotura

Paul Kolesar, Commscope
Rick Pimpinella, Panduit
Steve Swanson, Corning
Sharon Lutz, US Conec
Alan Ugolini, US Conec
Adit Narasimha, Molex
Jack Jewell, CommScope
Stephen Bates, PMC-Sierra

Data-center >100m need

- Data-centers have evolved around the 300m 10G-SR reach over MMF for intra-DC, with SMF for inter-building / campus
- Reach challenges became apparent soon after 40/100GE introduction
- At 40GE, partly solved by introduction of proprietary ~300m MMF QSFP+
  - Initial use 4x10GE, then for 40GE once both ends move to 40GE
- Interest in use of 40G-LR4 and emergence of PSM4 technology for intra-DC
- At 100GE, no solution between SR10 (150m) and LR4 (10km) reach; larger step in cost from MMF to SMF solution

Data-center >100m need

- 100GE 500m objective set with intra-DC links in mind
- Solution that addresses 500m, and is cost-optimized for shorter reaches (where larger volume of links resides) is most attractive
  - Link distributions in kipp_01_0112_NG100GOPTX.pdf, kolesar_02_0911_NG100GOPTX.pdf
  - Cost-centroid length concept in kolesar_01b_0512_optx
- Need does not end at 100GE. Same set of questions at 400GE and 4x100GE
  - 400GE Study Group underway
  - Decision in 802.3bm has impact well into the future

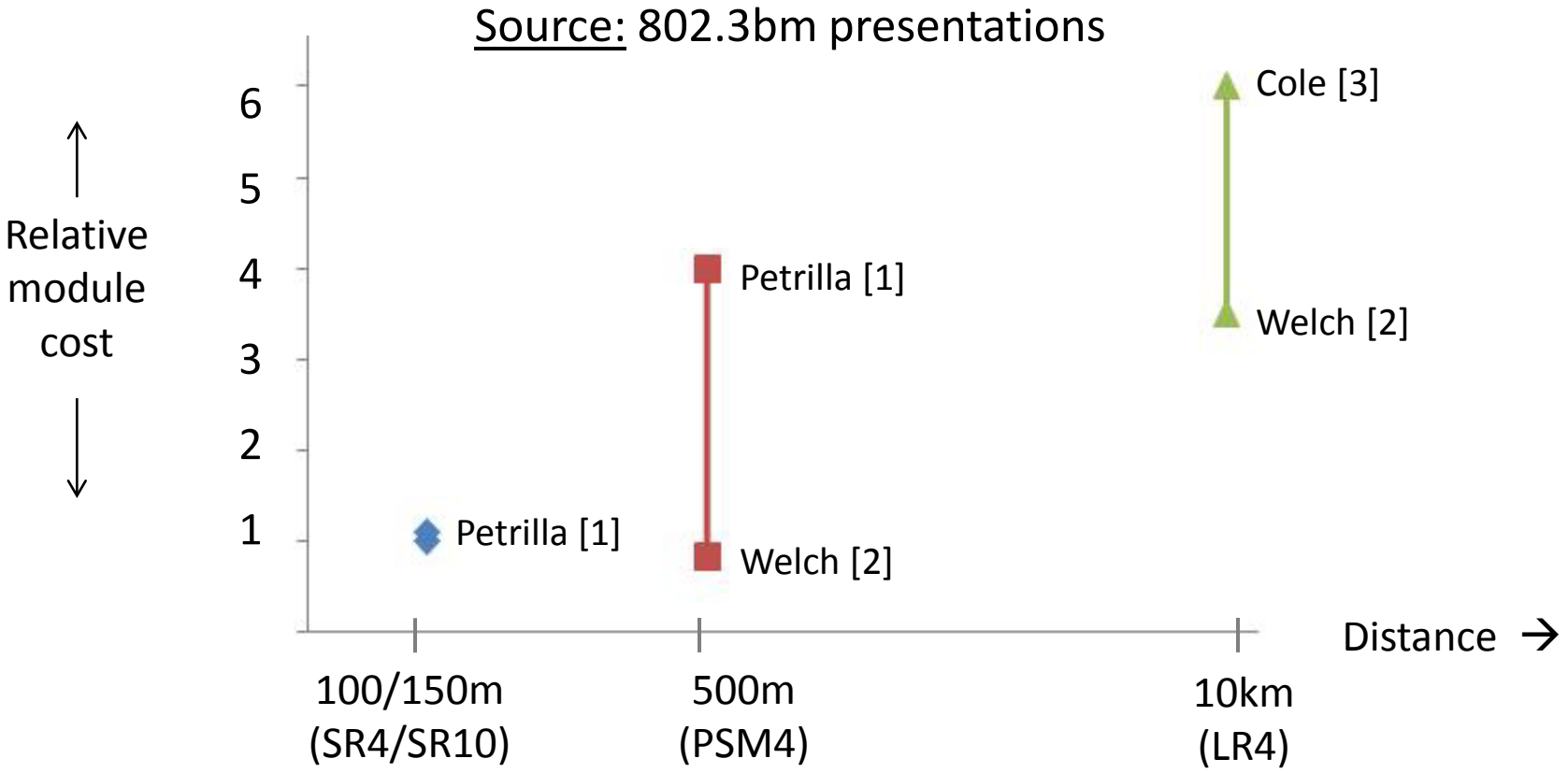
500m objective - where are we

- 6 Task Force meetings (including this one), 6 Study Group meetings, 22 SMF Ad Hoc meetings
- Large number of presentations
- Solutions under consideration: CWDM, DMT, PAM8, PSM4
- This meeting – likely the last opportunity to pick a 500m SMF proposal in 802.3bm

Why PSM4?

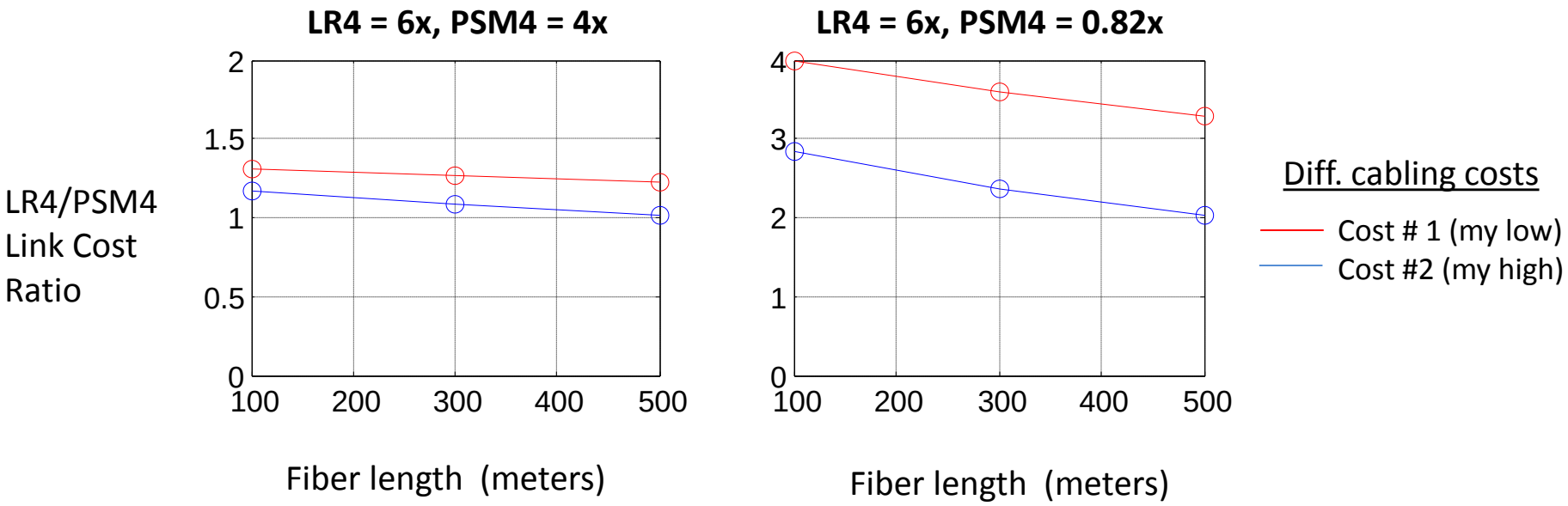
- Lowest cost module likely to be PSM4
  - welch_01b_0113_optx, petrilla_03a_0113_optx, cole_01_0313_optx, shen_01a_0313_smf
- Lowest link cost for multiple scenarios in the <= 500m application space
  - shrikhande_01_0613_smf.pdf, welch_02_0613_smf.pdf
- Lowest power module likely to be PSM4
  - anderson_01_1212_smf.pdf, welch_01_0313_optx.pdf, petrilla_03a_0113_optx.pdf
- Smallest 100G module FF (QSFP28) in nearer-term, at lowest risk
  - Power -> Density -> Cost
- Broad support from module manufacturers
  - Implementations feasible in near-term

Modules relative costs vs. Reach (assuming SR4/10 → PSM4 → LR4)



[1] petrilla_03a_0113_optx : SR4 CFP4 (1.1x), PSM4 CFP4 (4x)
[2] welch_01b_0113_optx : PSM4 (0.82x) and LR4 QSFP28 (3.5x) using SiPh
[3] cole_01_0313_optx : LR4 CFP4 Gen3 (6x)

Link cost analysis: PSM4 and LR4



- LR4/PSM4 ratio ~ 1 (equal cost) for cable cost #2
- LR4/PSM4 ratio ~ 1.3 @ 300m for cable cost #1

- LR4/PSM4 ratio > 2 for both cable costs (PSM4 links significantly cheaper)

- Cabling cost clearly matters, only a few presentations discussing cabling costs, compared to modules costs
- Results from cable cost #2 match other analyses in 802.3bm quite well (Cole, Kolesar) – cable cost #2 used for further analysis

Link to detailed presentation from SMF Ad Hoc at shrikhande_01_0613_smf.pdf

Link cost analysis: summary

- PSM4 links are lower cost than LR4 for the target application
- PSM links remain lower cost than WDM (LR4 or CWDM) over a wide range of WDM module costs
  - Duplex WDM v. parallel cost in shrikhande_01_0613_smf.pdf
- PSM4 provides lowest cost at shorter reaches where larger volume of links reside
- PSM4 remains the lower cost alternative for the application space over a long period of time

Module power / size

- Lowest power module likely to be PSM4
  - LISEL based PSM4 transceiver not including CDR ~ 2W (anderson_01_1212_smf.pdf)
  - Si Photonics based re-timed PSM4 module < 2.5W (welch_01_0313_optx.pdf)
  - DFB discrete TOSA based re-timed PSM4 module ~ 3.76W (petrilla_03a_0113_optx.pdf)

- Technology and power projections indicate strong probability of fitting in smallest FF -- QSFP28

Market potential for PSM4

- Increasing use of parallel starting with 40GE
  - Use of parallel MMF >150m likely at 40GE (~300m QSFP+)
  - Use of PSM technology for 4x10G, and for 40G when link cost lower than 40G-LR4 (or when cabling is present)
- PSM4 + LR4 provides a more distinct choice to users compared to CWDM + LR4
  - Users can leverage different cost trade-offs for Parallel v. Duplex : lower cost in modules, higher cost in cabling
- Broad support from module manufacturers
- Availability of modules in the near term is expected
- Systems integrators interested in supporting PSM4
- Opportunity to standardize PSM4 and ensure inter-op!

Looking beyond 100GE

- Adopting PSM4 for the 100GE 500m objective directly helps with the introduction of 400GE in the data-center
  - E.g.1: PSM4 + LR4
  - E.g.2: PSM4 + Serial 100G

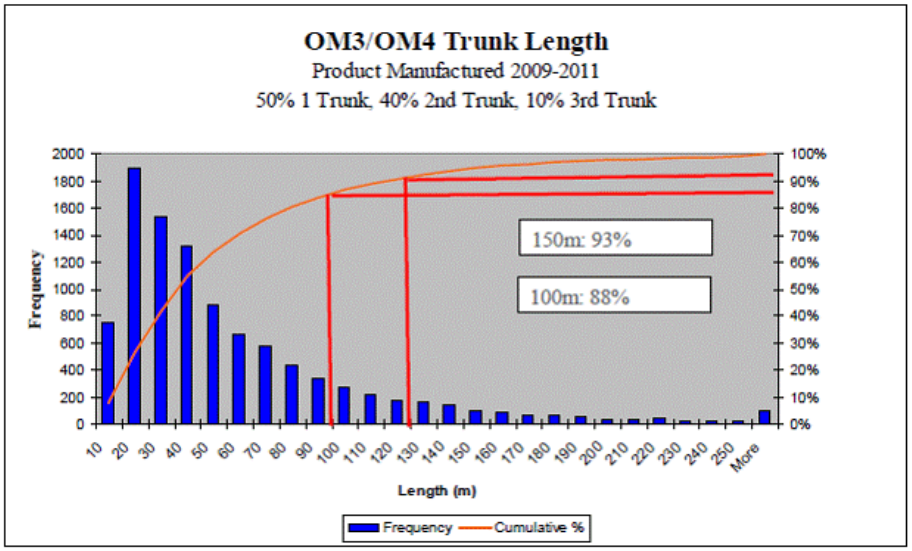
- PSM infrastructure is a building block for 400GE and necessary for 4x100GE breakout in the data-center

Summary

- PSM4 has great potential for driving down cost in the target application space
  - Module cost, link costs, power, density
- Having PSM4 + LR4 provides more choice to the DC user and will enhance 100GE market potential
- PSM4 has broad support from the eco-system
- PSM infrastructure will play an important role in introduction of 400GE and high-density 100GE
- Recommend that 802.3bm adopt the PSM4 baseline proposal for the 500m SMF reach objective

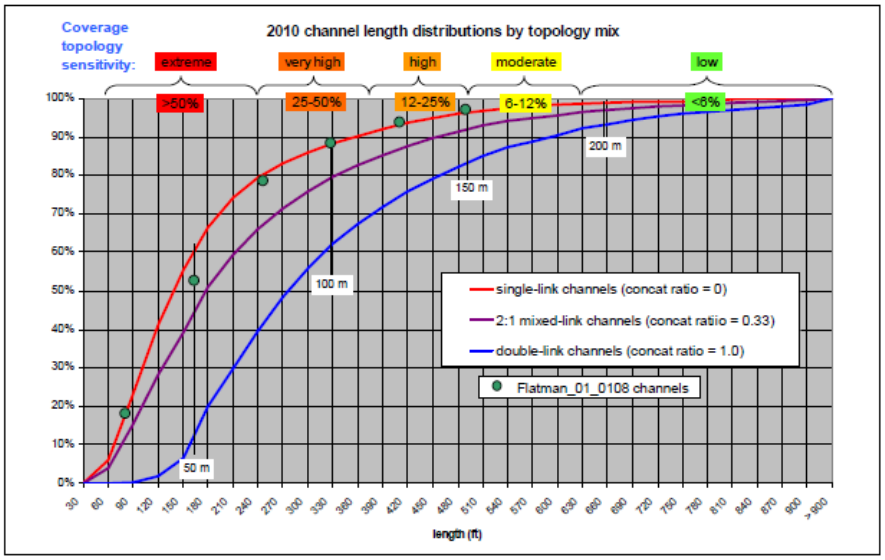
BACKUP

Data-center links statistics: snapshots



Source: Corning data from [kipp_01_0112_NG100GOPTX.pdf]

- At least 10% links beyond 100m



Source: [kolesar_02_0911_NG100GOPTX.pdf]

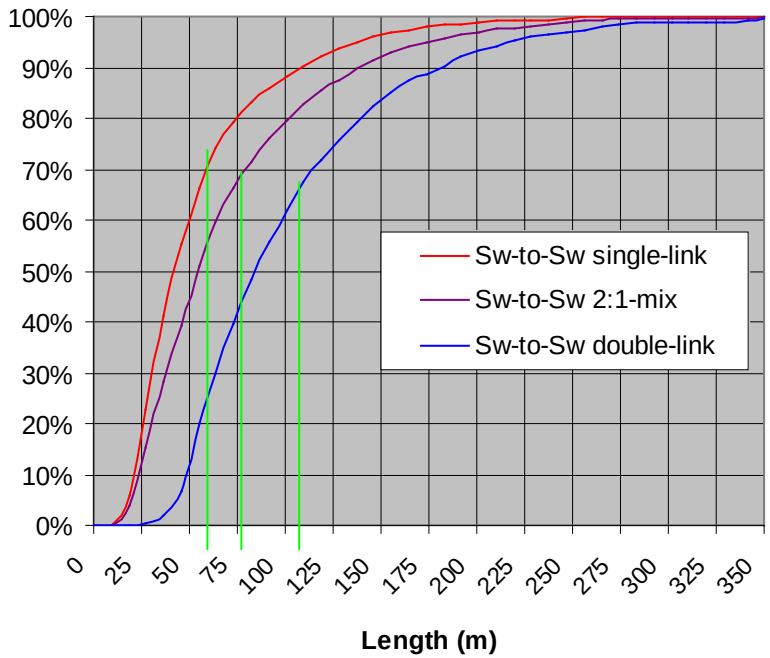
- 10% single-link channels beyond 100m
- Also seen in flatman_0108 channels
- 20-40 % of double-link channels > 100m depending on double-link model

Cost-centroid length

Cost-Centroid Lengths [m]

Length Selection	Switch-to-Switch Channels		
Single-mode deployed for	Single Link	2:1 Mix Link	Double Link
All Lengths	59	75	106
> 100 m	148	157	163

Data Center Channel Length CDFs and Cost Centroid Lengths for Channels > 0 m



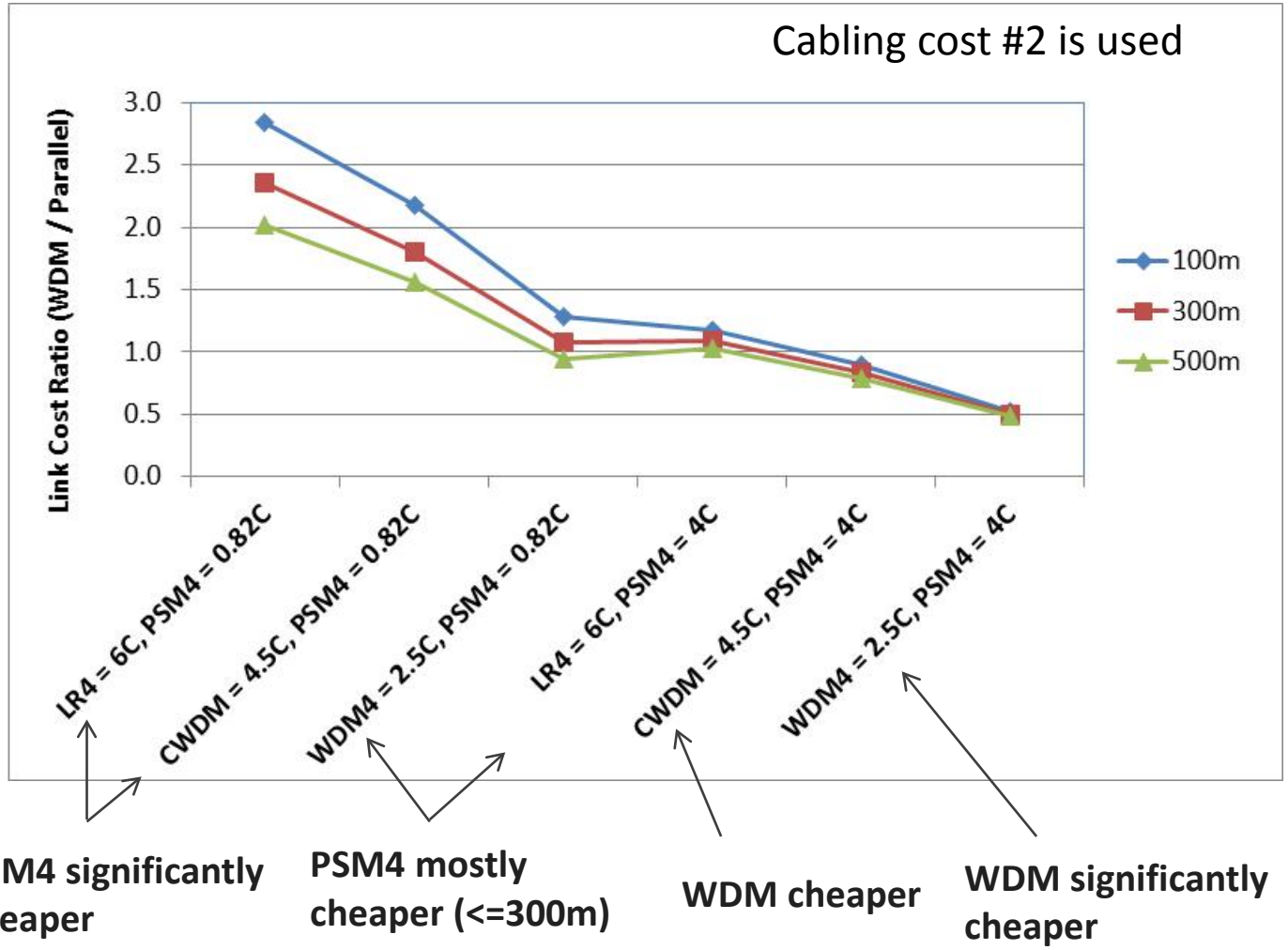
source: kolesar_01b_0512_optx

Link cost analysis (1)

- Analysis method : similar to cole_01b_0213_smf
- Total link cost ratio = (2*Duplex module + 2f_DL) / (2*Parallel module + 8f_DL)
- Double Link model as described by P. Kolesar
  - Exception: MPO-LC cassettes, MPO-LC cables (PSM module), LC-LC cables (duplex module) used at end points
- Assumed 24f trunk cables : carries 3 x PSM4 circuits or 12 x duplex circuits
- 2 cabling costs considered
  - #1) my low end : chose lower cost cabling components
  - # 2) my high end : chose higher cost cabling components
- Module relative cost used – next slide

Link cost

Different scenarios of WDM and PSM relative module costs (from earlier slide) show PSM4 links can be cheaper in many scenarios



Link cost analysis (2)

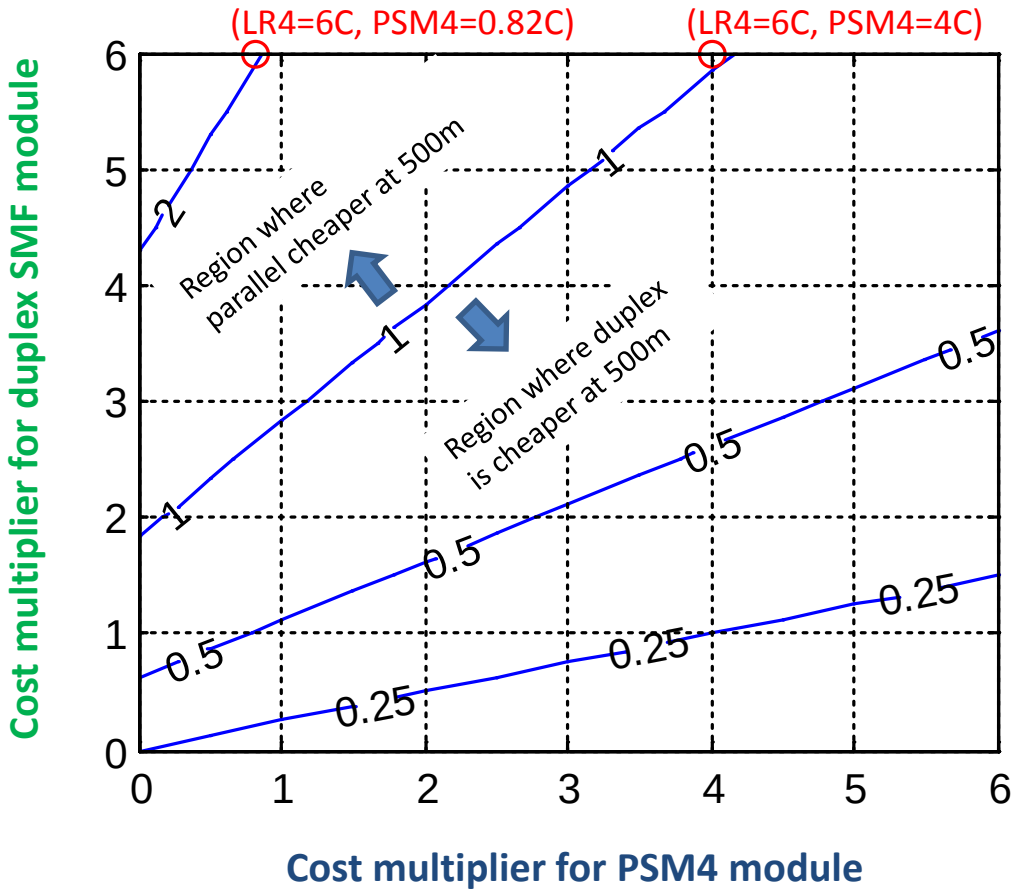
- The cost ratio of a WDM (or any duplex) SMF link to parallel SMF link can be calculated more generally, as a function of the duplex module and parallel module costs

- Duplex module relative cost = X = C * (0, 0.5, 1.0, … 6)
- Parallel module relative cost = Y = C * (0, 0.5, 1.0, … 6)
  - Where C = SR10 CXP cost

- Calculate matrix of link cost ratio (duplex/parallel) for above X, Y values of module costs

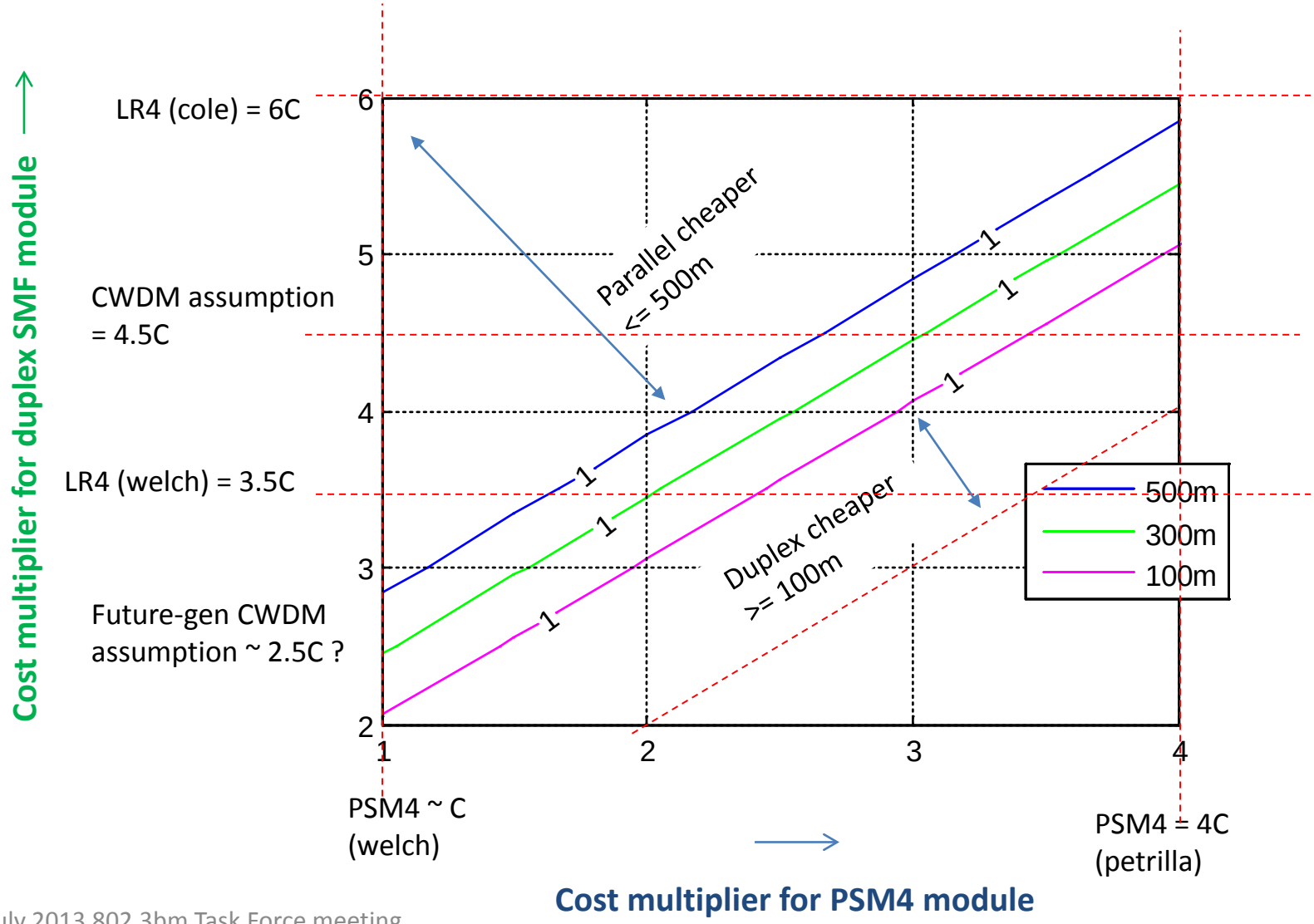
- From the matrix data, trace contour lines on a X-Y plot
  - For e.g. contour lines where duplex/parallel link cost ratio = 0.25, 0.5, 1.0, and 2.0 are plotted on next slide for 500m cable length

Contour plot for 500m SMF



- As a reference, the two points (red circles) match the LR4/PSM4 ratio plotted on slide 5
- Line marked “1” is contour line of equal cost (duplex link = parallel link)
  - Parallel is cheaper above “1” line
  - Duplex is cheaper below the “1” line

Link cost analysis: parallel v. duplex



Link to detailed presentation from SMF Ad Hoc at shrikhande_01_0613_smf.pdf