75-GHz spacing for 400G Ethernet over DWDM for DCI

Liang Du (Google)
Tad Hofmeister (Google)
Rich Baca (Microsoft)

Supporters

Rene Schmogrow (Google)

Mark Filer (Microsoft)

Matthew Mitchell (Facebook)

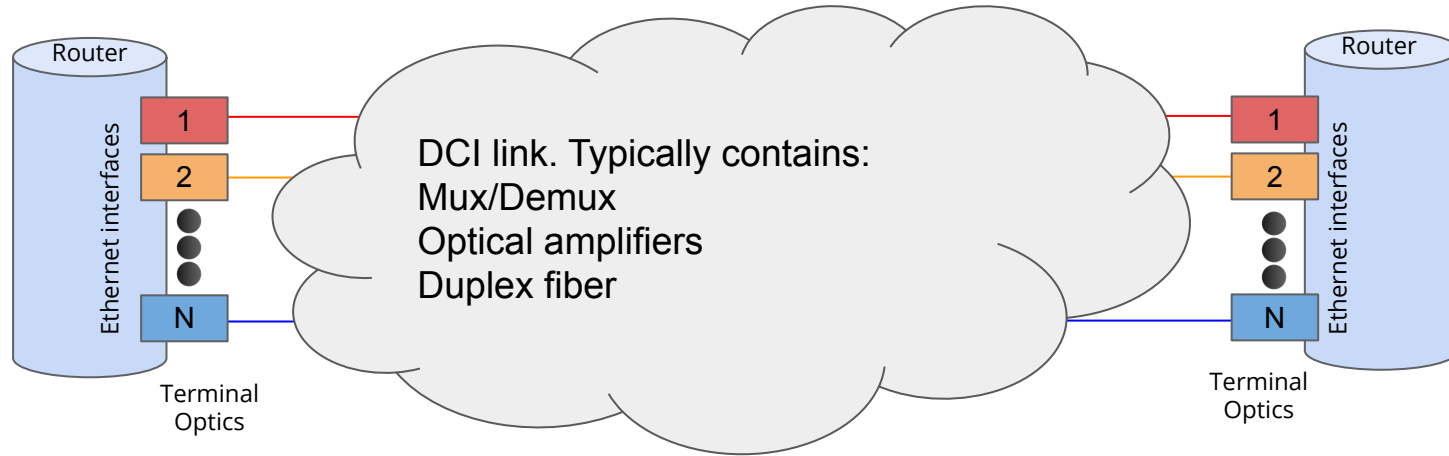Mark Nowell (Cisco)

Gary Nicholl (Cisco)

Winston Way (NeoPhotonics)

Tom Williams (Acacia)

Henk Bulthuis (Broadex Tech UK Ltd).

Ogawa Daisuke (NTT Electronics)

Typical DCI link

Router

Ethernet interfaces

1
2
●
●
●
N

Terminal Optics

DCI link. Typically contains:
Mux/Demux
Optical amplifiers
Duplex fiber

Router

Ethernet interfaces

1
2
●
●
●
N

Terminal Optics

- This is the most typical use case of 400 Gb/s DWDM
- Used for all links too long/expensive to use bulk dark fiber, where each fiber-pair carries a single 400 Gb/s stream
- All interfaces are natively Ethernet

3

Bandwidth growth

- Google's current traffic
  - Already using multiple DWDM pairs for Long-haul links
  - Anticipate needing multiple DWDM pairs for capacity intra-metro DCI links in 1-3 years
  - Traffic growth is not slowing down with compute. Traffic is now outpacing compute growth.
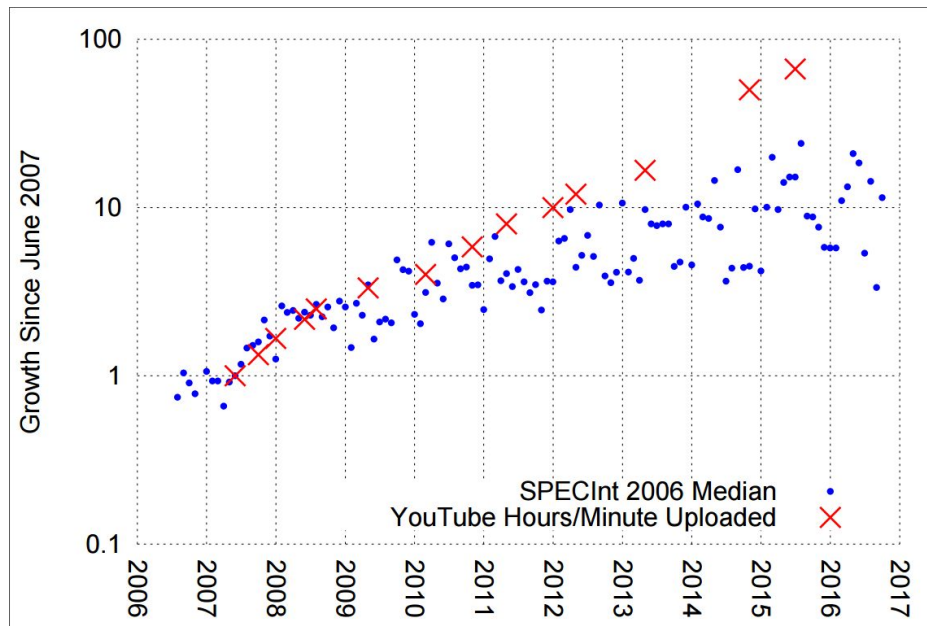
Amin Vahdat (Google Fellow) - 2018 ONF Keynote

4

75-GHz vs 100-GHz spacing

For DCI capacities of >15 Tb/s between A and B

- 100-GHz spaced - 48 channels: 400 Gb/s * 48 = 19.2 Tb/s
  - X fiber pairs needed per DCI
- 75-GHz spaced - 64 channels: 400 Gb/s * 64 = 25.6 Tb/s
  - 3/4X fiber pairs needed per DCI
- Advantages of using 75-GHz channel spacing
  - 6.4 Tb/s increase in capacity per system (64 vs 48 channels)
  - 33% increase in system capacity
- Broad market potential for 400 Gb/s over 75-GHz spaced DWDM
  - Experts affiliated with three hyperscalers (Google, Microsoft, Facebook) are aligned on moving to 400-Gb/s Ethernet AND on 75-GHz spaced, 64-channel DWDM systems for DCI
  - Enable metro-wide ubiquitous Ethernet datacenter networks

Cost of additional fiber-pairs

- More operational complexity
- Fiber leasing costs
  - Monthly recurring cost for each fiber pair ordered
  - Upfront splicing cost for each increase in the number of strands ordered
  - Carrier hotel patching costs: monthly recurring and upfront for each augmentation
  - Typically costs are by fiber mile so longer links cost more
- Line systems costs
  - EDFAs
  - AWGs
- Traffic growth is outpacing interface rate
  - Cost per bit in transceivers trends down with time
  - Cost per fiber lease and line system components have been more constant
 - Ratio of fiber/line system : transceiver is increasing

DWDM system and link costs

| Cost contributions	"DWDM system"	"DWDM link"
Mux/demux (AWG)	All	None
Optical amplifiers (EDFA)	All	None
Fiber plant setup (CapEx)	All	None
Fiber lease (recurring)	All	None
Intra-building patching	Smaller (system to ODN termination panel)	Larger (transceiver to line system)
Transceivers	None	All

DWDM systems required
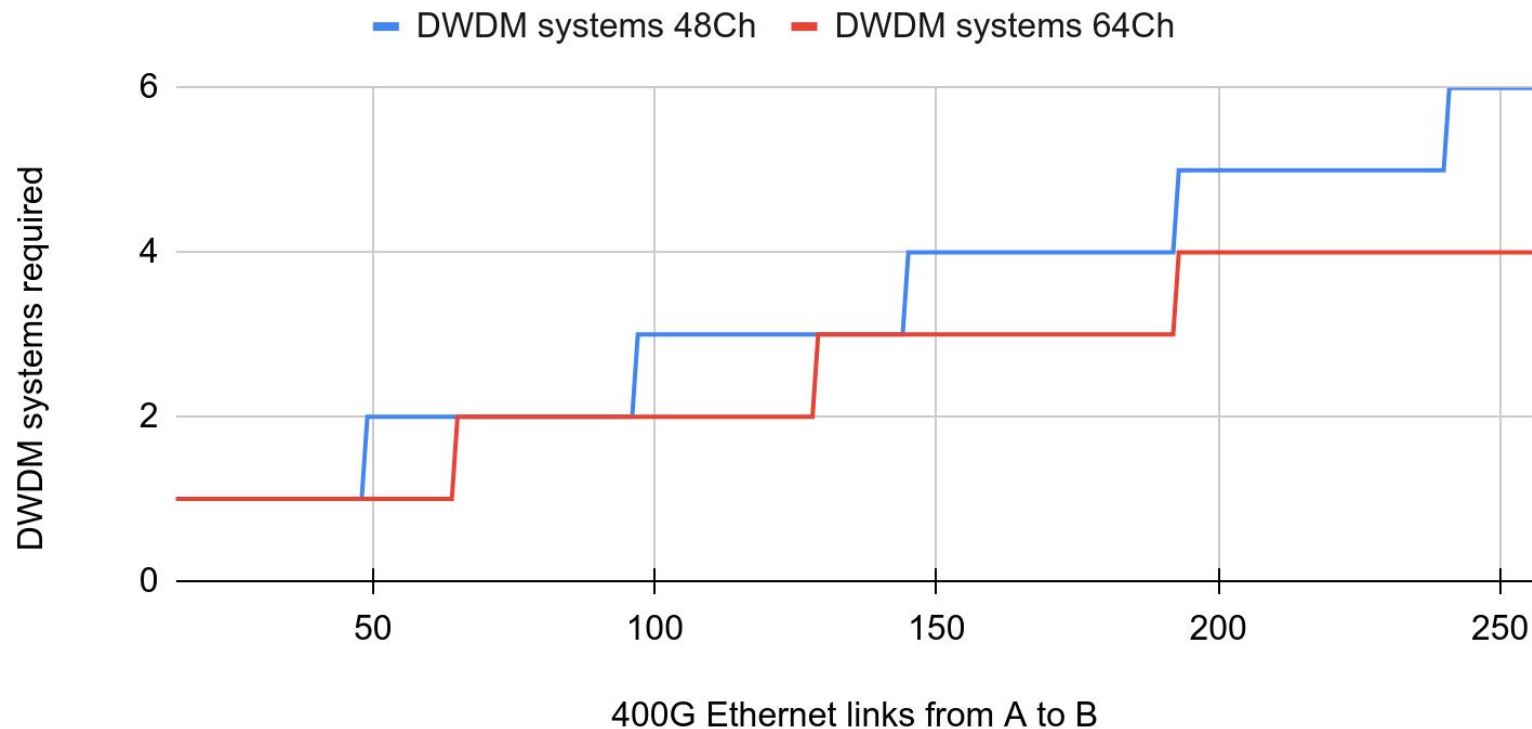
DWDM systems per 400G Ethernet link
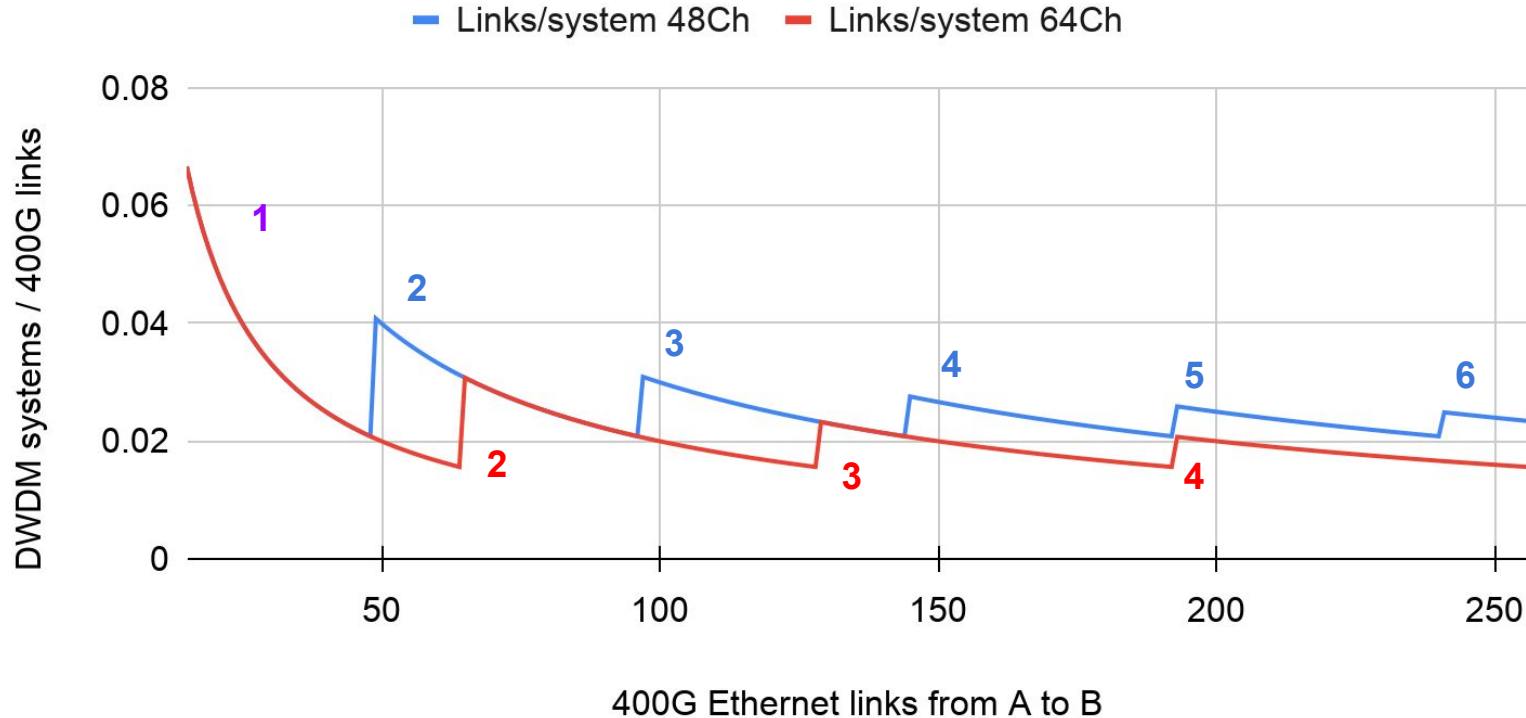
64-channels is a good number
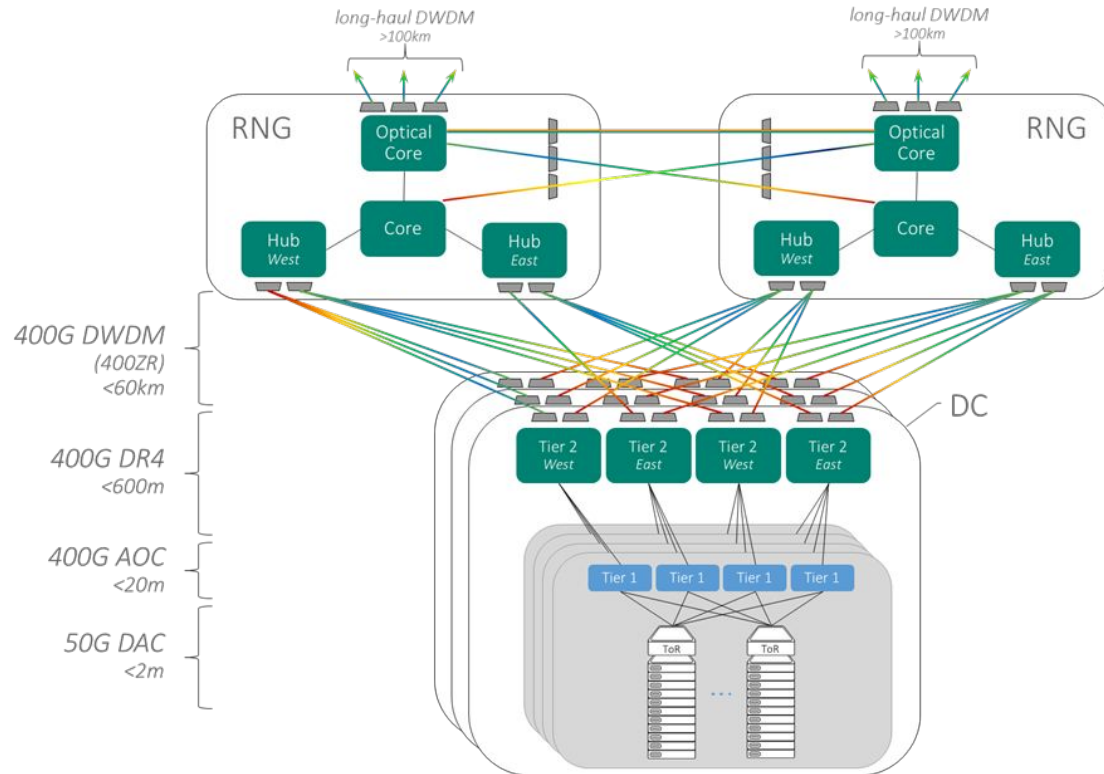
- Microsoft builds metro areas using a single Clos fabric [baca_3ct_01_190328]

- This requires groups of 2^n 400G links between any two datacenter locations

- Per-link savings (64Ch vs 48Ch):
  50% for 64 channels,
  33% for 128 channels

- Ethernet has typically used links with 2^n lanes. There is alignment with IEEE practices to allow for 64 channels

Timing of 400G DWDM Ethernet links

- DCI is the currently the main driver for 400 Gb/s DWDM Ethernet.
    - DCI needs 400G DWDM Ethernet soon
    - DCI needs the extra capacity offered by 64-channels enabled by 75-GHz channel spacing

- DCI has very different traffic patterns to Access networks eg. MSOs
    - Different system architectures
    - Different volume growth

Conclusions

- The demand for 400-Gb/s Ethernet inter-datacenter links is very real

- Optical experts affiliated with three data center operators (Google, Microsoft, Facebook) believe 75-GHz spaced 400-Gb/s DWDM Ethernet links will be the most cost effective DCI solution

- Discussions with AWG, transceiver module, and DSP experts thus far suggest 75-GHz spacing capable modules and 100-GHz spacing capable modules will have identical BOM and DSP algorithms