Thoughts on 50Gb/s ASIC IO and backwards compatibility considerations for 50G, 100G and 200G

Gary Nicholl - Cisco

IEEE Atlanta, Jan 18-22, 2016

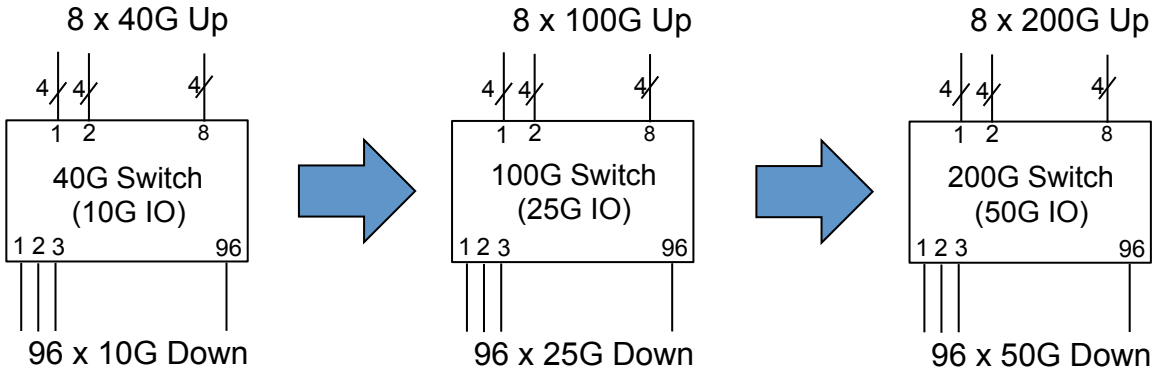
Topics

- Terminology
- Switch Chip evolution
  - 10Gb/s IO > 25Gb/s IO > 50Gb/s IO
- Switch/Network upgrade scenarios
  - 10Gb/s IO > 25Gb/s IO > 50Gb/s
- Higher density (port counts) for legacy (lower rate) interfaces
- Conclusions

Terminology

- Multiple terminology in use. No common understanding.
  - Backwards compatibility
  - Backwards commonality
  - Legacy support
  - Implementation commonality or reuse.
- Irrespective of the naming there are several key considerations/desires:
  - connecting new equipment to legacy equipment in field
    - support legacy interface rates/types in new equipment
  - support new interface types on legacy equipment (cheaper PMDs)
  - upgrading network speeds and feeds (e.g. 10G to 25G, etc)
  - efficient support for multi-rate ASIC/Chip designs (push towards maximizing re-use)

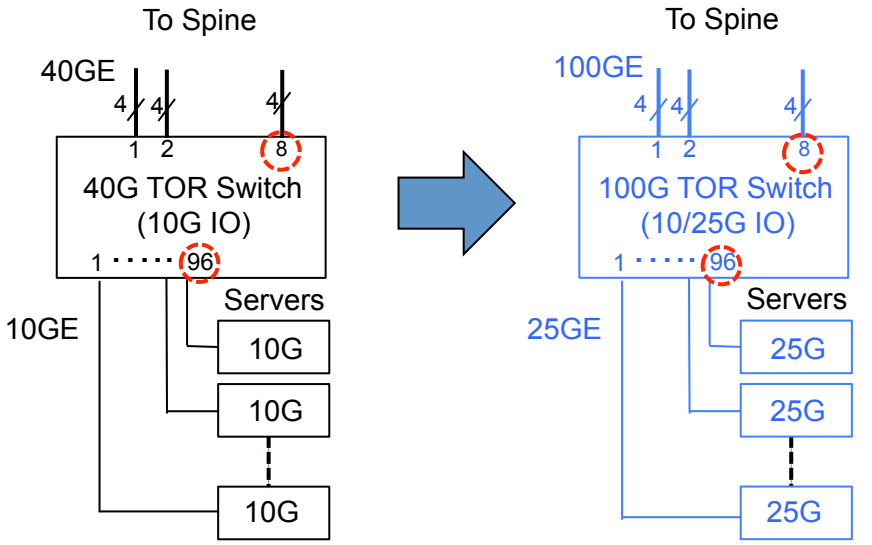
Switch Chip Evolution



Capacity	1.28T	3.2T	6.4T
Serdes rate	10Gb/s	25Gb/s	50Gb/s
Serdes count	128	128	128
QSFP ports	32	32	32
Uplink ports	8 x 40GE	8 x 100GE	8 x 200GE
Downlink ports	96(24x4) x 10GE	96(24x4) x 25GE	96(24x4) x 50GE
Oversub ratio	3:1	3:1	3:1
Uplink switch radix	8	8	8

Will the 25G>50G IO transition track the 10G>25G IO transition ?

Switch Upgrade Scenario 1 (Swap out)

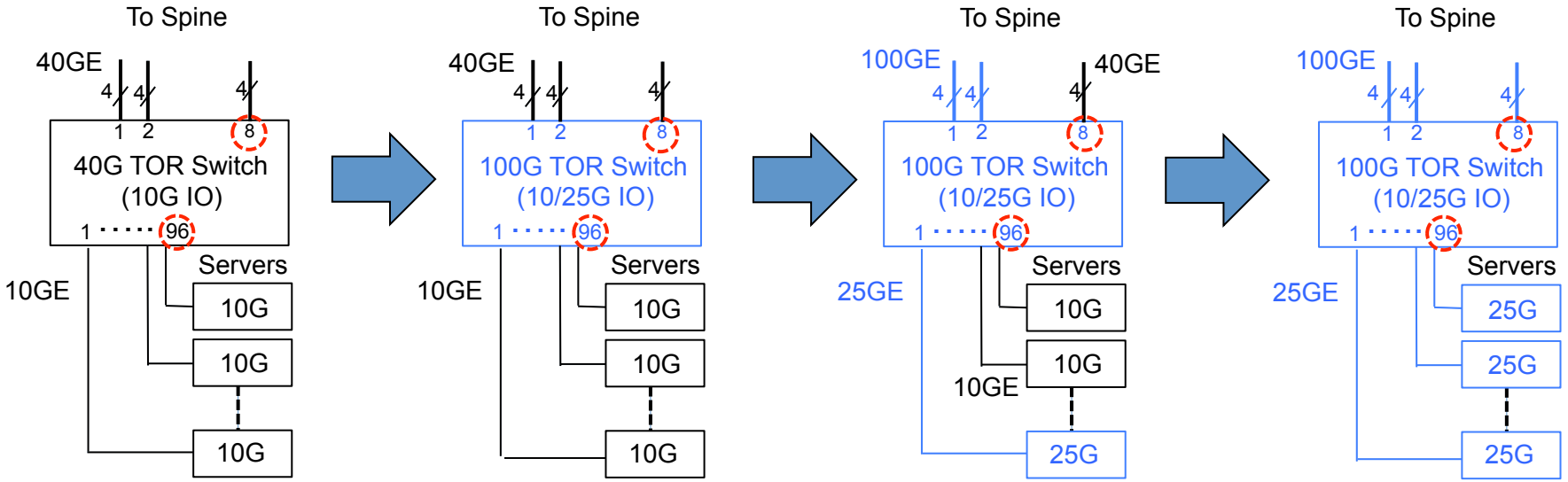


- Day 1 (starting point)
- 40GE/10GE config

- Swap out complete rack
- 100GE/25GE config
- Run all ports at new speeds/feeds Day 1
- no need to worry about support for legacy rates

Note: Same # of ports and links after the upgrade as before (8 up and 96 down), just running at double the rate.

Switch Upgrade Scenario 2 (Incremental)



- Day 1 (starting point)
- 8x40GE up
- 96x10GE down

- Install shiny new TOR !
- 8x40GE up
- 96x10GE down
- Switch IO in “downspeed”
- no network capacity loss !

- Incrementally upgrade uplinks to 100GE and downlinks to 25GE
- Hybrid configuration

- Upgrade complete !
- 8x100GE up
- 96x25GE down

Note: This example also illustrates the power of the SFP/QSFP eco-system (new TOR switch accepts legacy SFP/QSFP modules)

Observations on 10G to 25G IO Transition

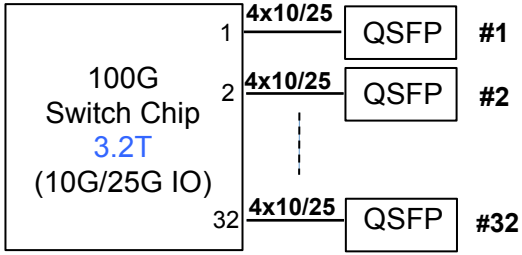
- New switch chips must support a ‘down speed’ serdes mode (to connect to legacy interfaces/equipment)
- During all the upgrade scenarios considered the number of ports and links in the network didn’t change (they just migrated to running 2x faster)
- Number of serdes on the switch chip didn’t change (stayed at 128)
- Number of QSFP ports on the TOR switch didn’t change (stayed at 32)
- No value (in the examples considered) in supporting double density of the lower speed ports (40GE in this case) on the switch chip

Support higher density of lower speed ports?

- Is there ever a need for a new switch chip to support higher density of lower speed (legacy) ports, than on the previous generation of switch chip?
- Not in the case of the scenarios shown in the previous slides
- But are there potential other applications where this may be needed/ desired ?
- Let’s again look at how this was dealt with during the transition from 10G IO based switch chips to 25G IO based switch chips.

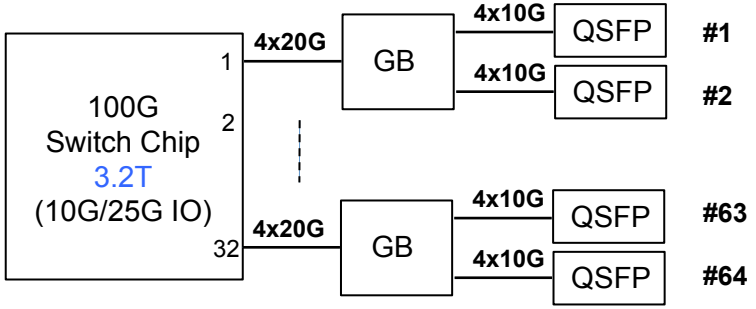
Dense 40GE from a 100GE Switch chip ?

Dense (32 port) 100G Application



- "Direct Connect" to QSFP
- 32 ports in 1RU
- 32 ports of 100GE/4x25GE b/o (3.2T)
- 32 ports of 40GE/4x10GE b/o (1.28T)
- Legacy ports run in 'downspeed' mode
- Broad application space

Dense (64 port) 40G application

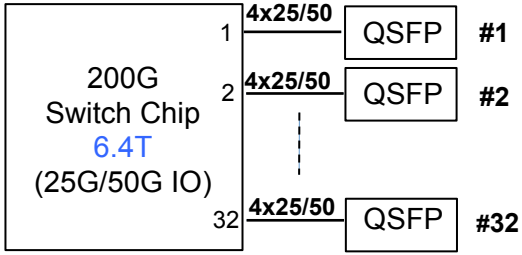


- Switch + 1:2 external gearbox breakout + QSFP
- Need new 2x20G 40GE AUI
- 64 ports in 2RU (from single switch chip)
- 64 ports of 40GE (2.56T)
- Only 50% density for 100GE/10GE (i.e. 32/128 ports in 2RU)
- Cannot support 16x40G up + 192x10G down (2 x 40G chip)
- Very Limited application space (dense 40GE only) ?

How useful/common is the product configuration on the right ?

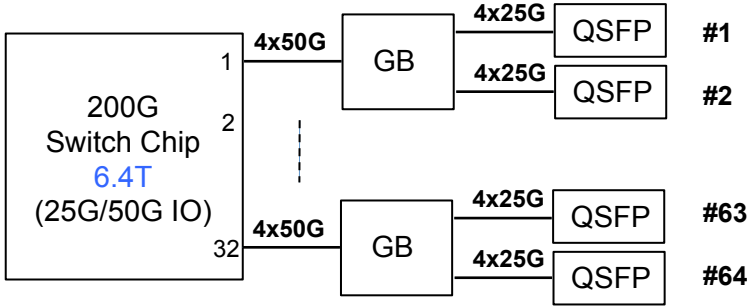
Dense 100GE from a 200GE Switch chip ?

Dense (32 port) 200G Application



- "Direct Connect" to QSFP
- 32 ports in 1RU
- 32 ports of 200GE/4x50GE b/o (6.4T)
- 32 ports of 100GE/4x25GE b/o (3.2T)
- Legacy ports run in 'downspeed' mode
- Broad application space

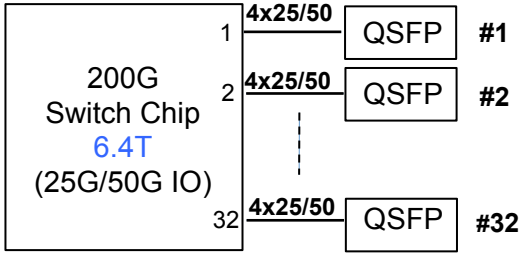
Dense (64 port) 100G application



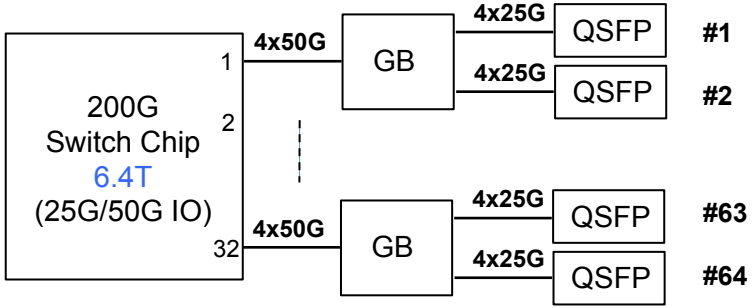
- Switch + 1:2 external gearbox breakout + QSFP
- Need new 2x25G 100GE CAUI-2 (FEC issues ?)
- 64 ports in 2RU (from single switch chip)
- 64 ports of 100GE (6.4T)
- Only 50% density for 200GE/50GE (i.e. 32/128 ports in 2RU)
- Cannot support 16x100G up + 192x25G down (2 x 100G chip)
- Very Limited application space (dense 100GE only) ?

Dense 100GE from a 200GE Switch chip ?

Dense (32 port) 200G Application



Dense (64 port) 100G application



- Main stream volume application appears to be 32x200G (32x100G /w downspeed for legacy)
- Dense (64 port) 100G configuration appears to have limited applications
  - linecard faceplate is limited to 32 x QSFP anyway
  - doesn't support sufficient 25GE density (cannot run 2x25GE over single 50G IO)
  - FEC for 2x50G CAUI-2 further complicates things
- When making design decisions and tradeoffs going forward (and related to 50Gb/s IO), we should probably be optimizing for the configuration on the left and not on the right.

Conclusions

- If the 25G-to-50G ASIC IO transition mirrors what happened during the 10G-to-25G ASIC IO transition, then:
  - A ‘downspeed’ mode will be required (to support legacy 100GE and 25GE PMDs)
  - Running at reduced chip capacity , but with the same switch port density, for legacy interfaces is acceptable (and likely the primary application)
  - The application for a higher density mode for legacy interfaces remains unclear
- The above comments appear to have held true for previous Ethernet rate transitions, i.e. from 100M>1G>10G>40G>100G
- The FEC choice should be optimized based on the signaling rate, rather than the Ethernet MAC rate.