100G

Across the country, Across the Data Centre

Tim Rayner

Optical Engineer, AARNet

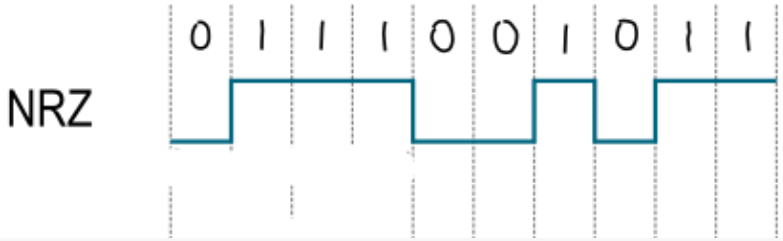
Tim.Rayner@AARNet.edu.au

Agenda

- Review of 10G, 40G &100G Standards
- 100G service roll-out across AARNet's Network
- 40G use in the data centre
- Newer 100G standards
- 100G around the data centre
- 400G, 200G and beyond

10Gbps Ethernet

- NRZ with 66b/64b coding -> line rate 10.3125 Gbps
- Electrical Lanes: 16 (300pin), 4 (XenPAK, X2), 1 (XFP, SFP+)
- Only 1 optical lane
- SR Multimode 300m@850nm OM3, LR singlemode 10km@1310nm
- ER SM 40km@ 1550, DWDM @ 1530-1560nm.. 80km or ~2000km with Amplifiers & Dispersion Compensation
- CWDM (Late arrival) – 1470 – 1610 @ 20nm spacing
- 10GHz bandwidth fits easily inside DWDM channels
  - 100GHz (0.8nm spacing) or 50GHz (0.4nm spacing)

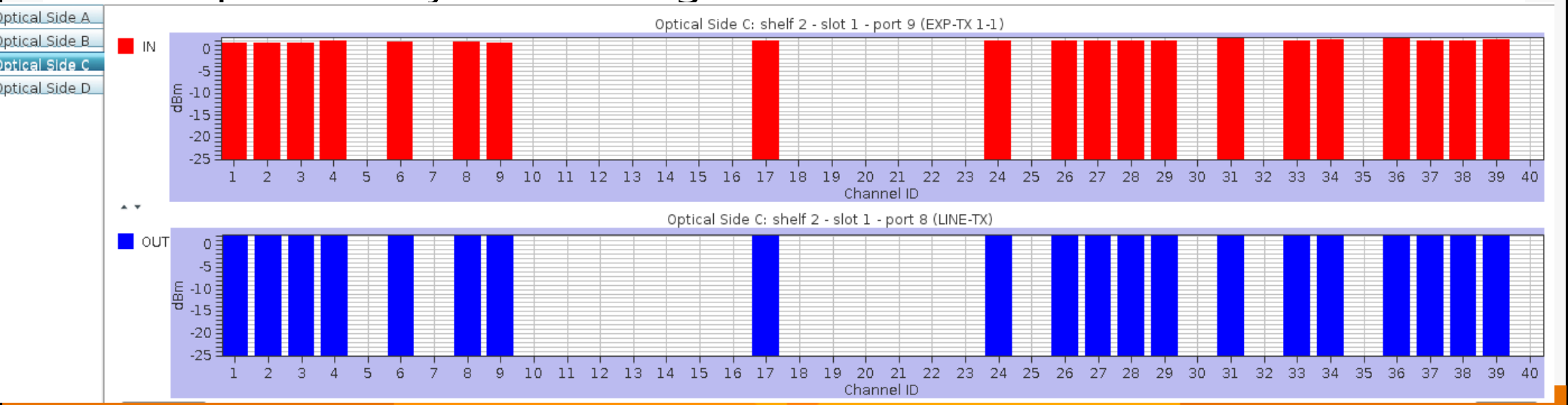


40G and 100G Ethernet

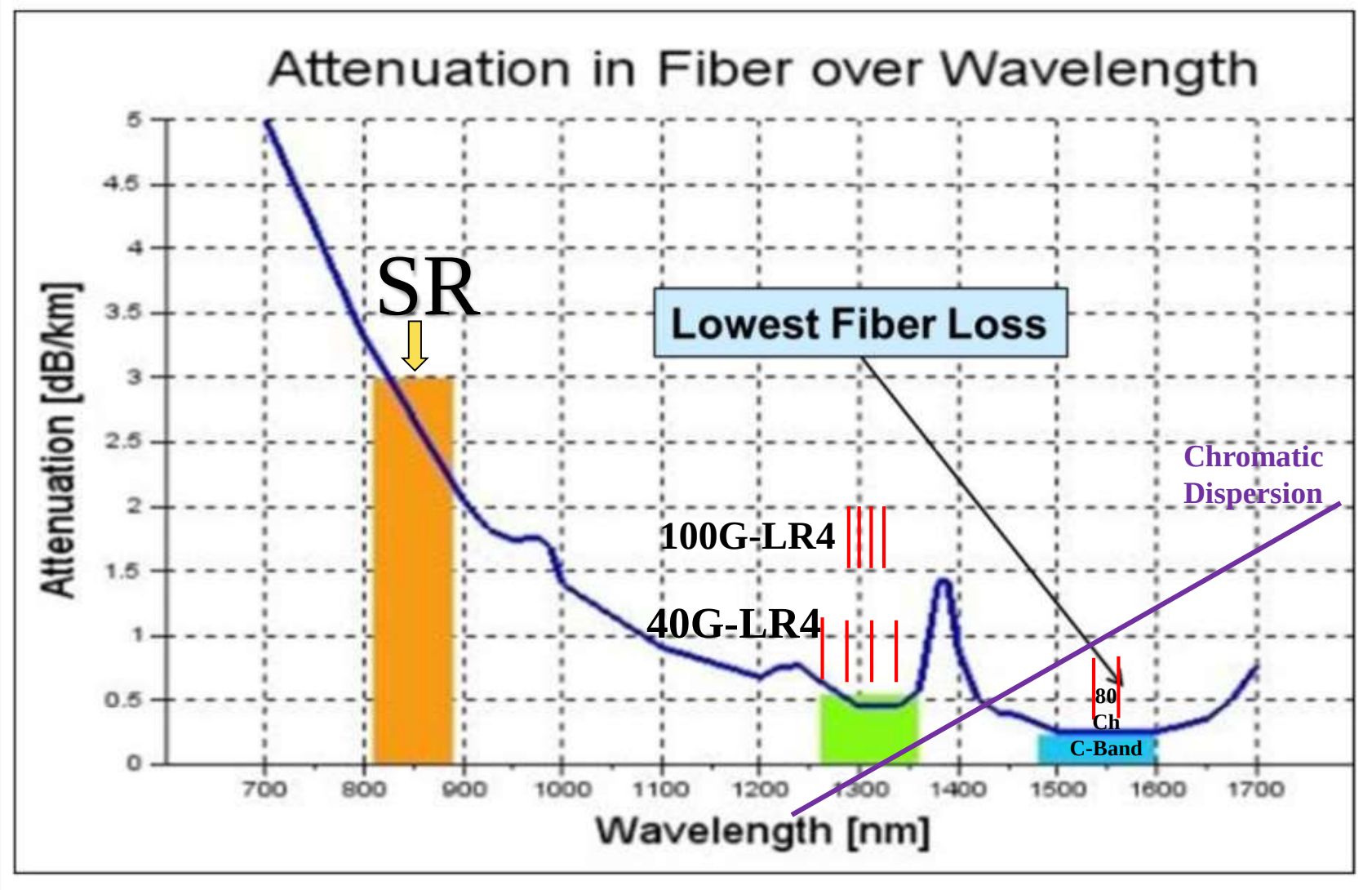
- Designed ~ 2008, standardized 2010: IEEE 802.3ba-2010
- Electrical lanes remain at 10Gbps NRZ
- Lanes at Layer1 (64/66 bit block level), no LAG hashes!
- 4 lanes for 40G, 10 lanes for 100G (CAUI-10)
- 40G-SR4, 100G-SR10 – parallel MultiMode OM3/4 ~ 100/150m
- matching electrical lanes – simple low power modules
- 40G-LR4 10km SM: matching electrical lanes to 20nm spaced CWDM
- 100G-LR4 10km SM: gearbox for 4 lanes @ 25G, closer spacing
- 100G-ER4 – longer range 40km – same wavelengths – available later.
- CFP module for 100G (LR4 & SR10), CXP SR10 from infiniband
- 40G CFP or smaller QSFP+
- Long range DWDM using DP-QPSK – external Transponder until recent CFP availability.

100G Singlemode

- Optical Lanes at 25 Gbps for LR4 & ER4. 1 lane NRZ signal per wavelength
- For DWDM DP-QPSK 25Gbaud signaling on 1 wavelength: 4 bits per symbol, rather than 4 lanes.
- Bandwidth ~ 33 GHz – fits within 50GHz DWDM channel
- DWDM systems up to 80 or 96 Channels @ 50GHz
- Flexspectrum systems can give 120 channels of 37.5 GHz

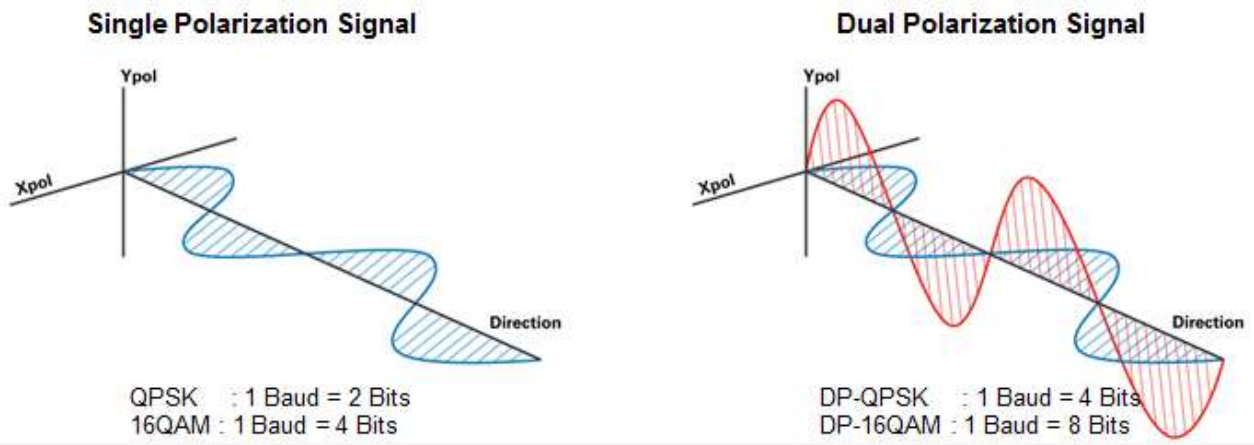


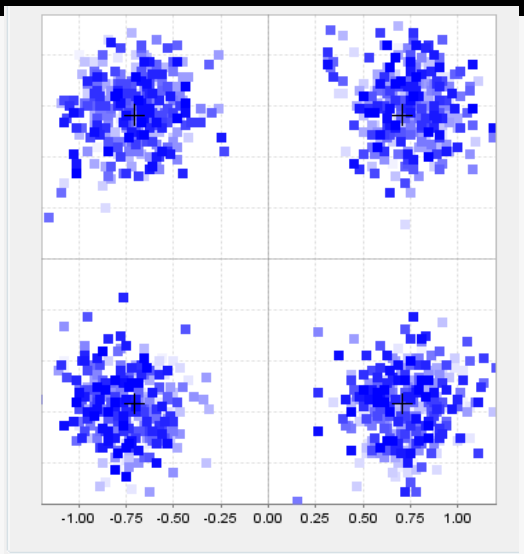
Service Wavelengths



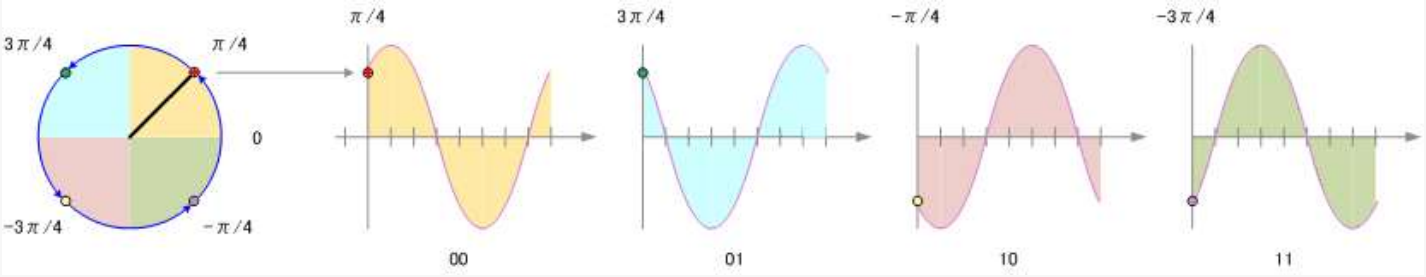
100G – DWDM Coherent – Long Range

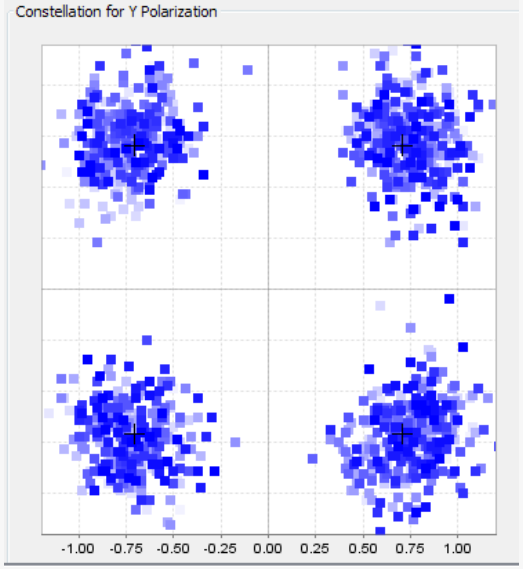
Dual Polarisation





Quadrature Phase Shift Keying

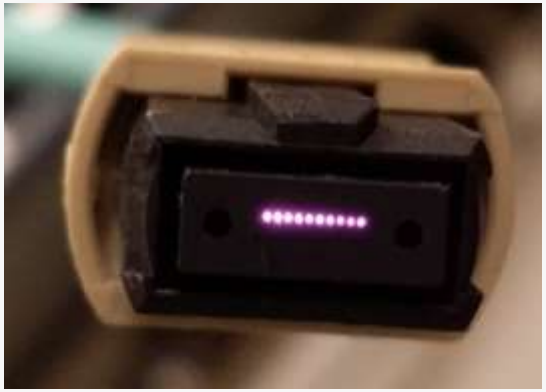




DP-QPSK

100G – First Gen modules & connectors

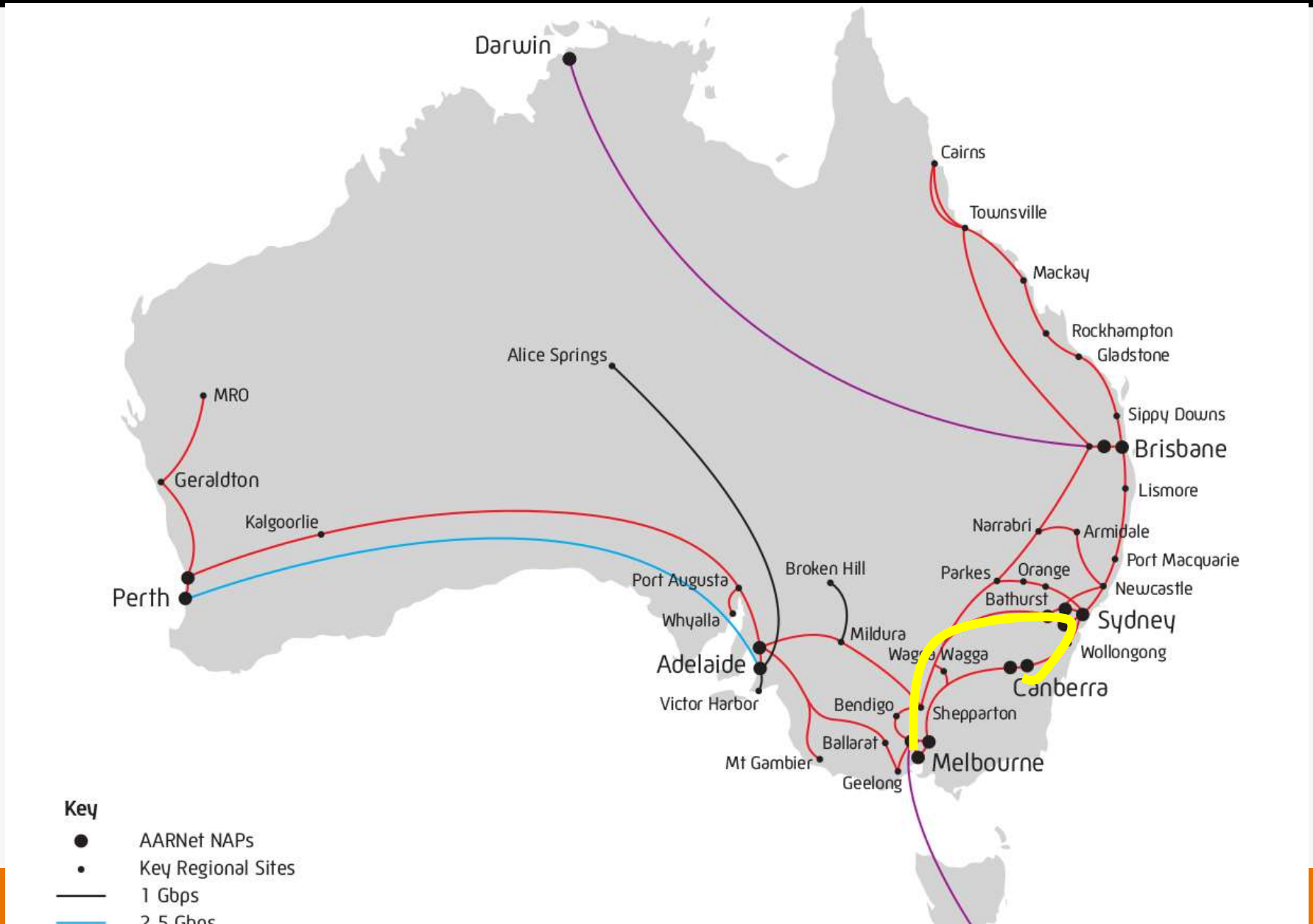
- CFP modules: LR4
- SR10 – 10 Tx & 10 Rx fibres
- MPO has 2 rows of 12 fibres
- Outside pairs unused for 100G-SR10, but CXP with infiniband can use 12 lanes



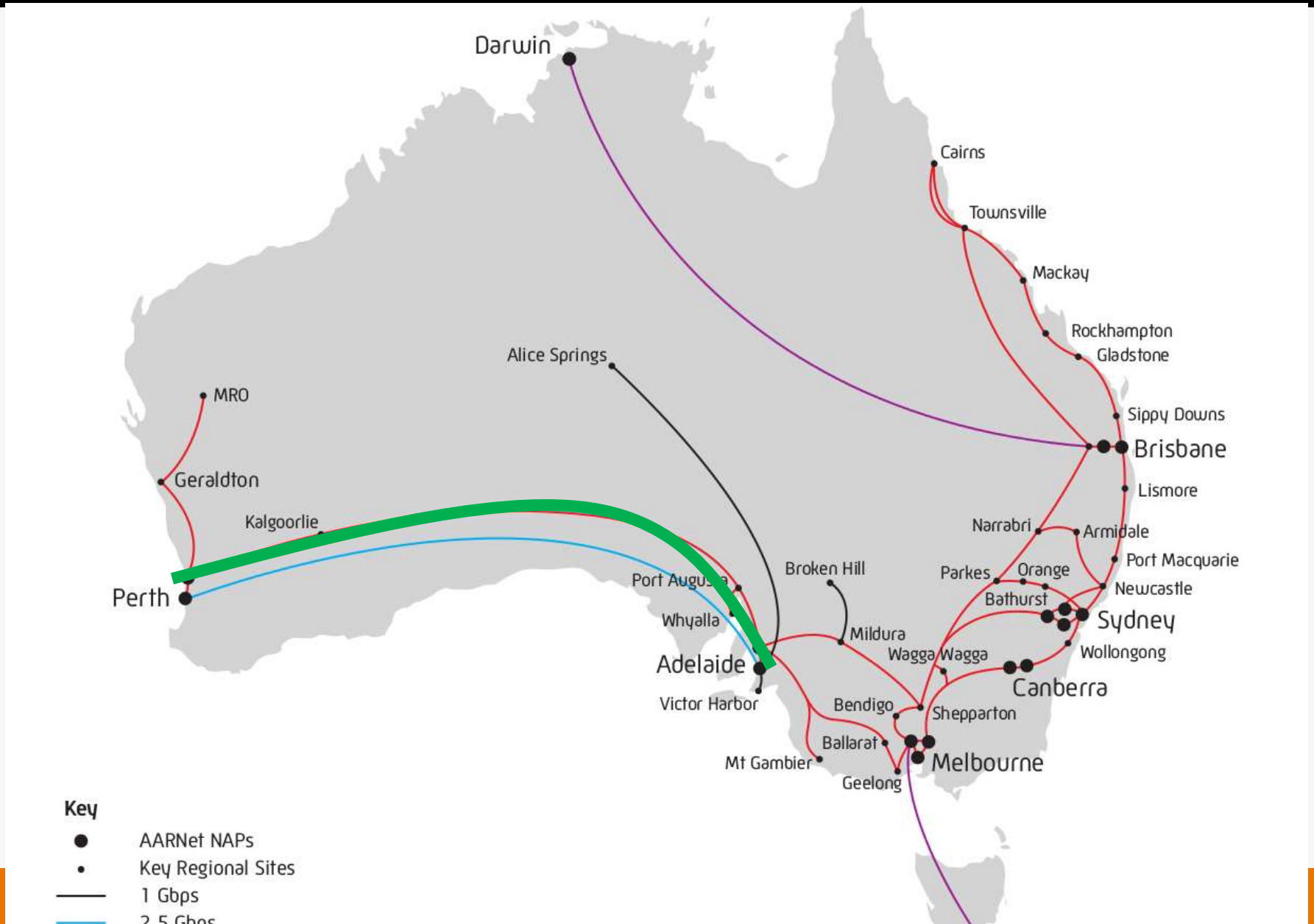
AARNet's 1st Gen 100G gear

- Cisco Transponder DP-QPSK
- With 10x10G card to make Muxponder
- Or 2 x 10G + 2 x 40G Clients
- With CXP client for 100GBASE-SR10
- LR4 in metro: up to 23km (no ER4 yet, or new ER4-Lite)

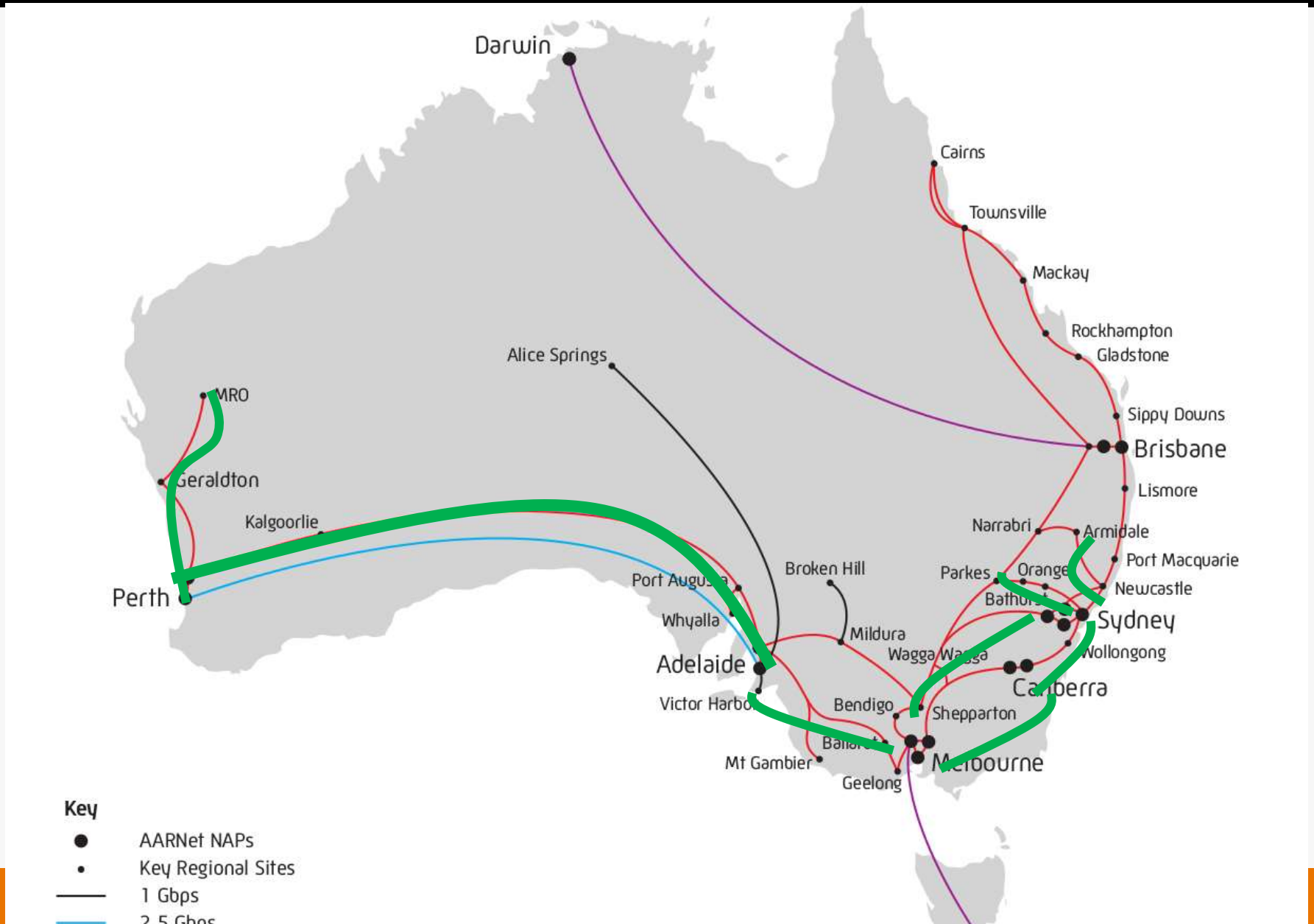
100G Field Trial – November 2011



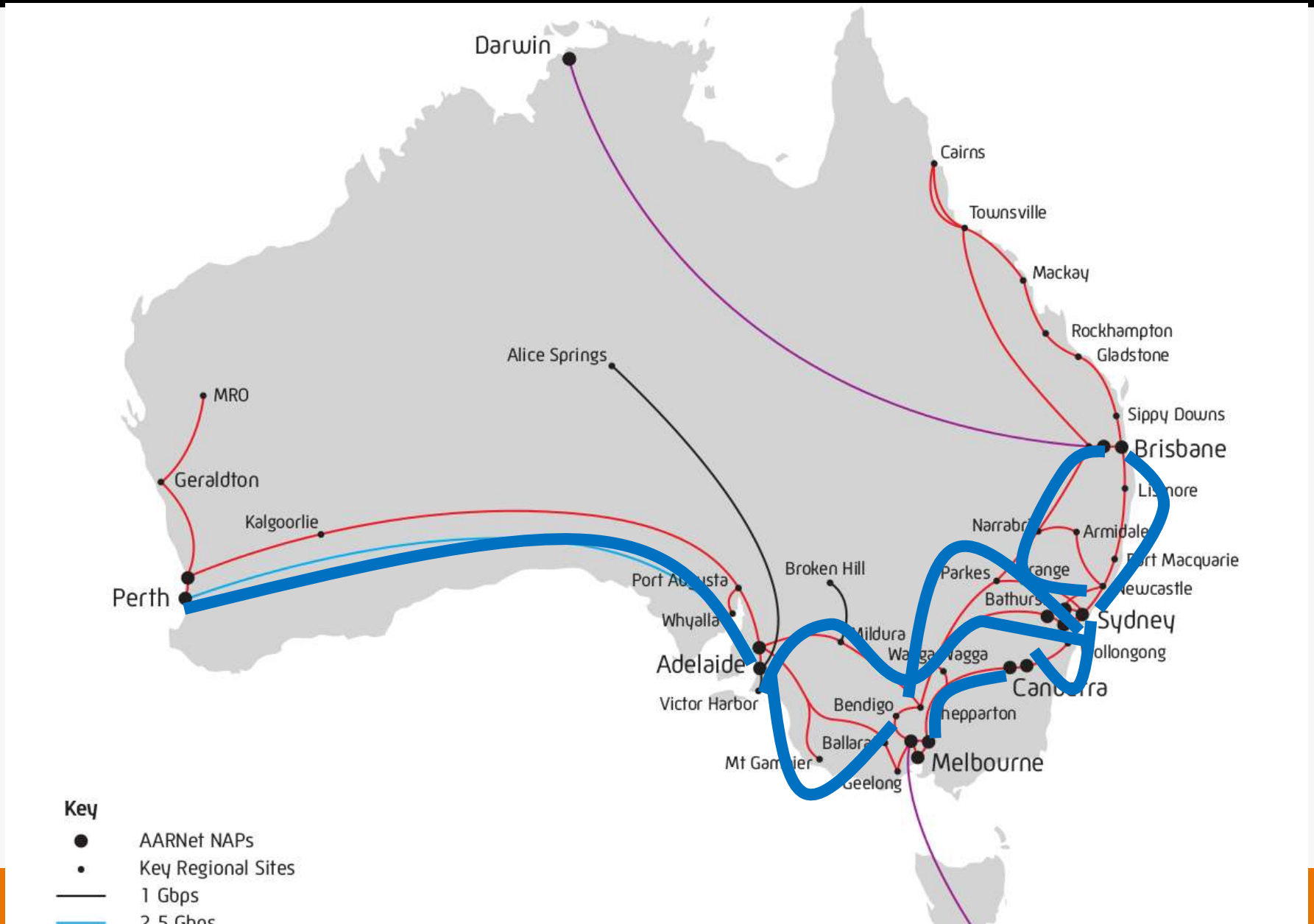
Adelaide to Perth: 10x10G mid 2013



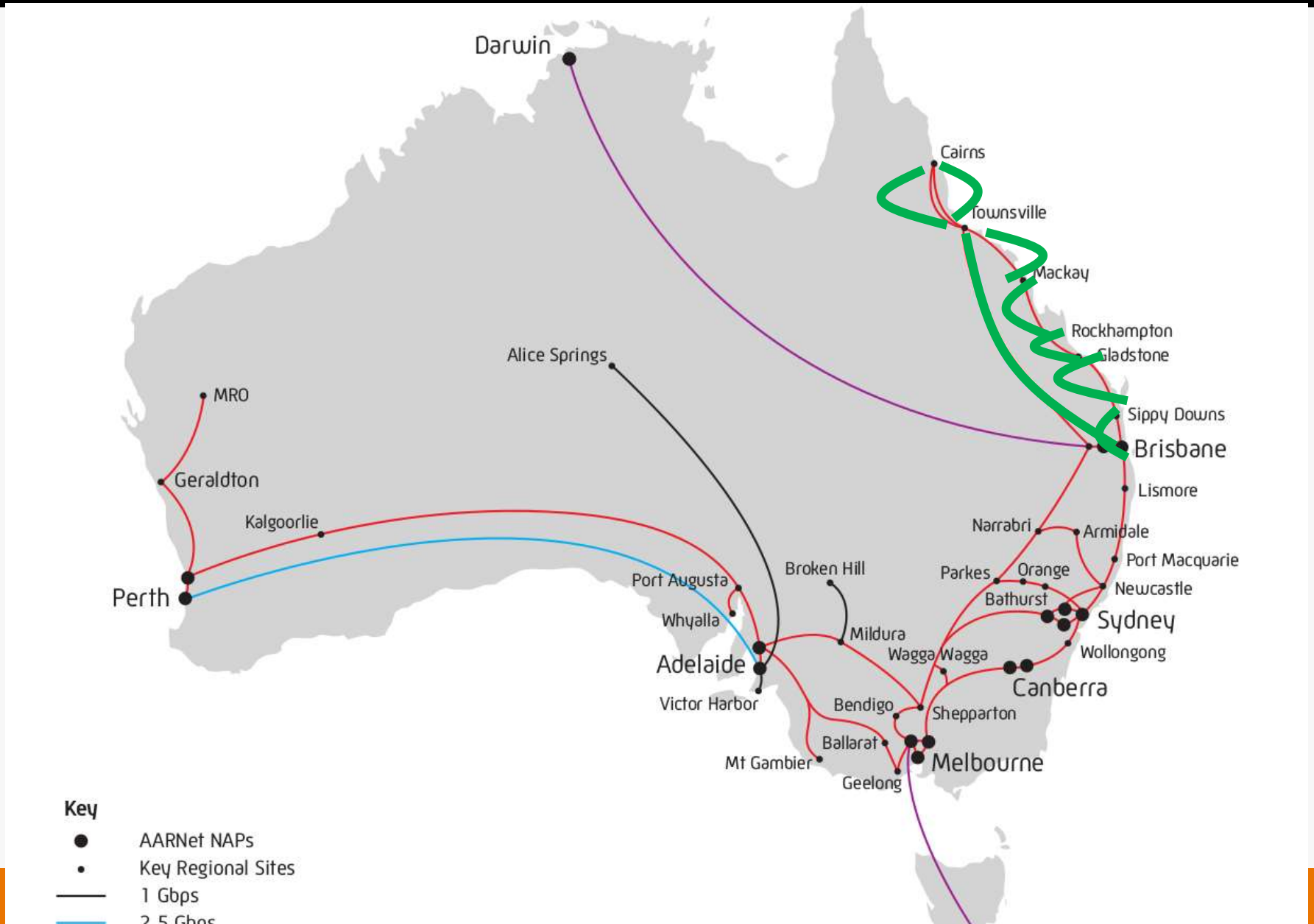
Other 10x10G 2013 - 2014



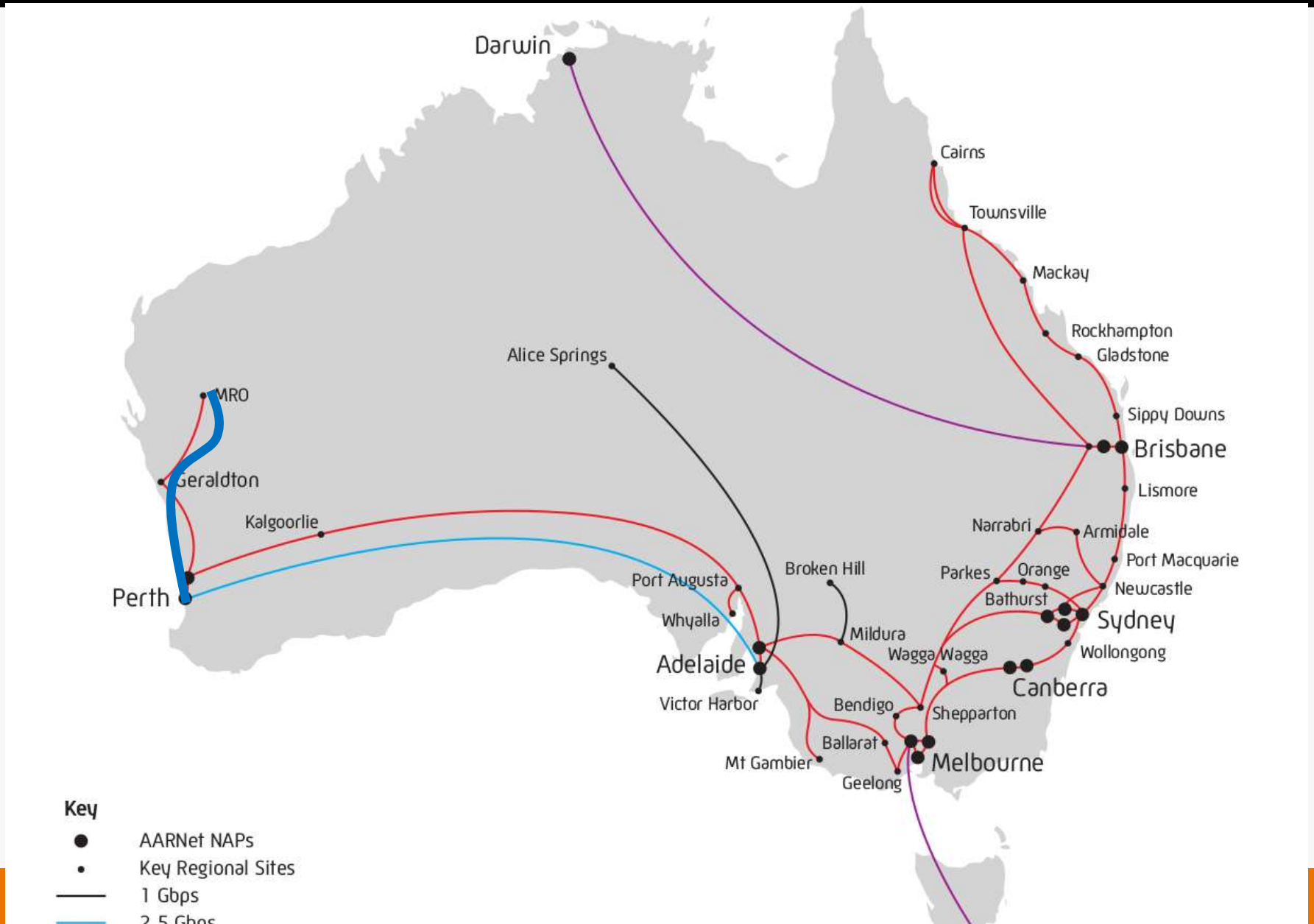
AARNet4 Backbone Routers - 2014



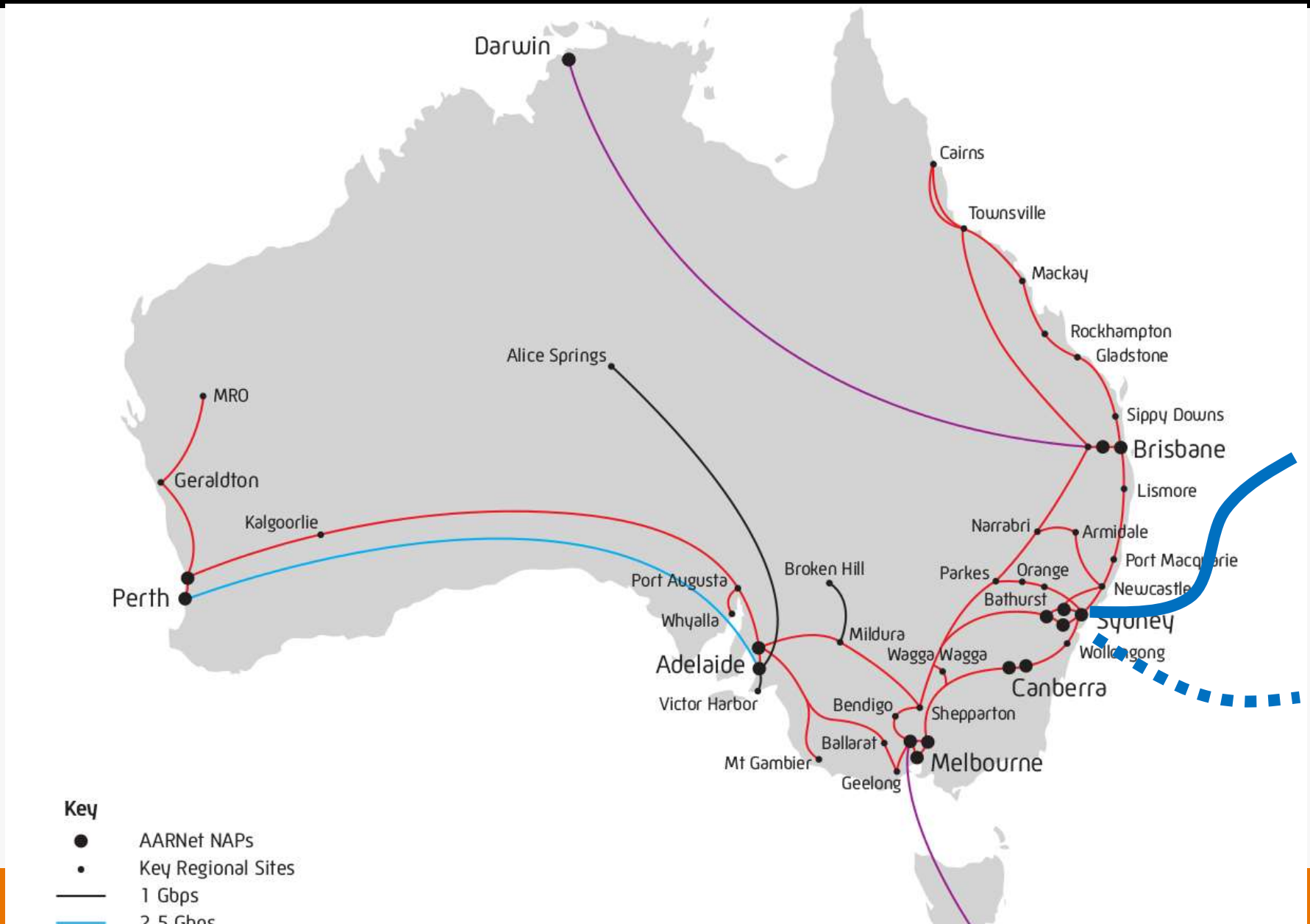
100G Sunshine Backbone – 2014 - 2016



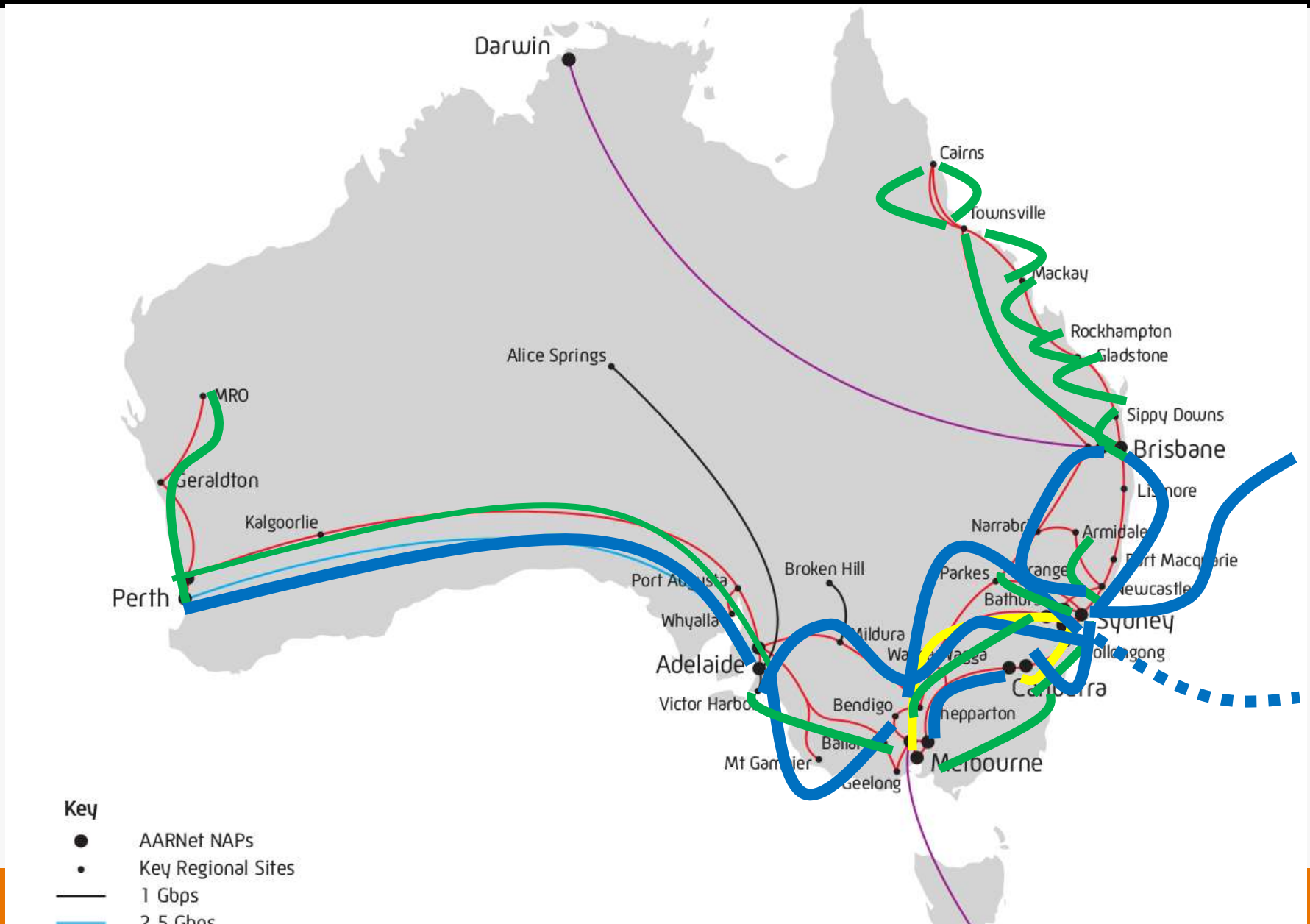
Customer 100G Optical Circuit - 2016



100G Across the Pacific- 2016



Combined – 100G Across the Country



100G: "Across the Country"

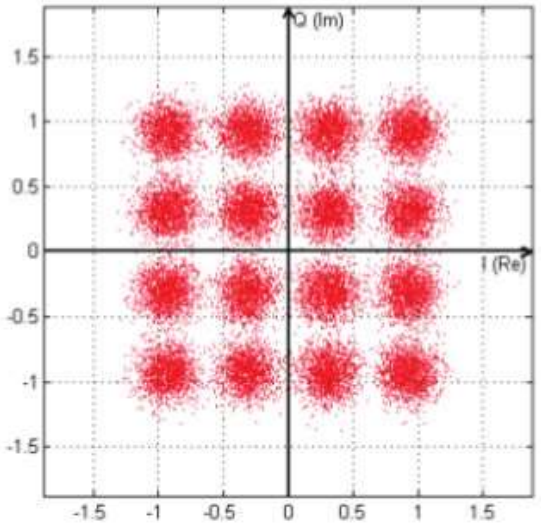
- Maps show Regional connections only
- Many more metro connections @ 100G
- Around 50 point-point links @ 100G DP-QPSK
- 100G performance: 3500km non-regen ADL-MRO
- 2700km ADL-PER in Production
- Tested 1 direction MEL-PER.
- Tested BPSK worked BNE to MRO (>5000km) 1-dir
  - Telstra "record" > 10,000km BPSK 100G, but 2 x 50G bearers.
- AARNet's 200G wavelengths – using 16QAM
  - Trialled up to 500km, Canberra – Sydney comfortably
  - in Production between Sydney & Brookvale
  - Hoping for 800km MRO to Perth. With 2dB improvement

Beyond 100G – 16QAM for 200G, 256QAM for 400G

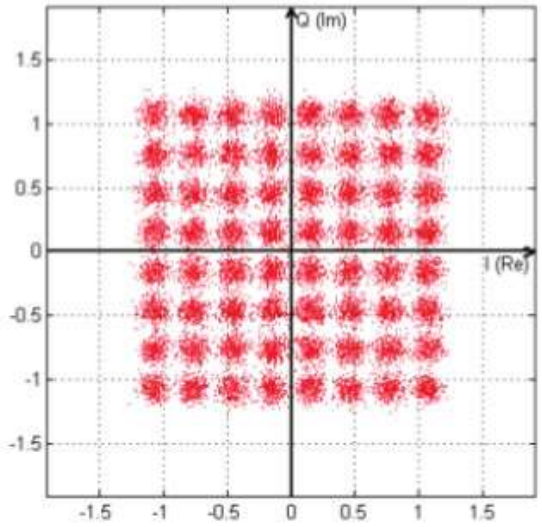
- Trade-offs OSNR vs bits-per-second vs bandwidth
- 200G may reach 1000km, not 3500
- 400G may only work in metro

448	2	56	100	4	DP-16QAM	22.4
448	2	42	75	6	DP-64QAM	26.6
448	2	28	50	8	DP-256QAM	31.9

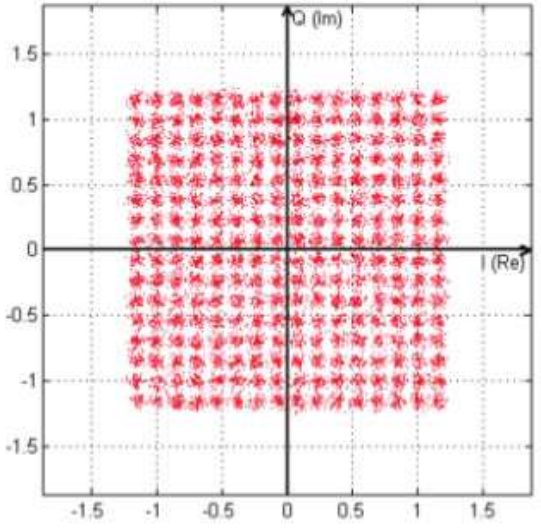
16 QAM



64 QAM



256 QAM



Carrier Space

- 100G dominated quickly
- 100G DP-QPSK performed as well as 10G wavelengths
  - 50G BPSK for ultra-long-haul. Submarine
- Capacity increase: 80 x 10G to 80 x 100G = 8Tbps
- Optical network: ROADMs & Amplifiers unchanged
- Just buy Transponders if you own Optical Network & have available wavelengths.
- “we’re just skipping 40G and going straight to 100G”
- Why would anyone bother with 40G ?

40G – in the Data Centre Space

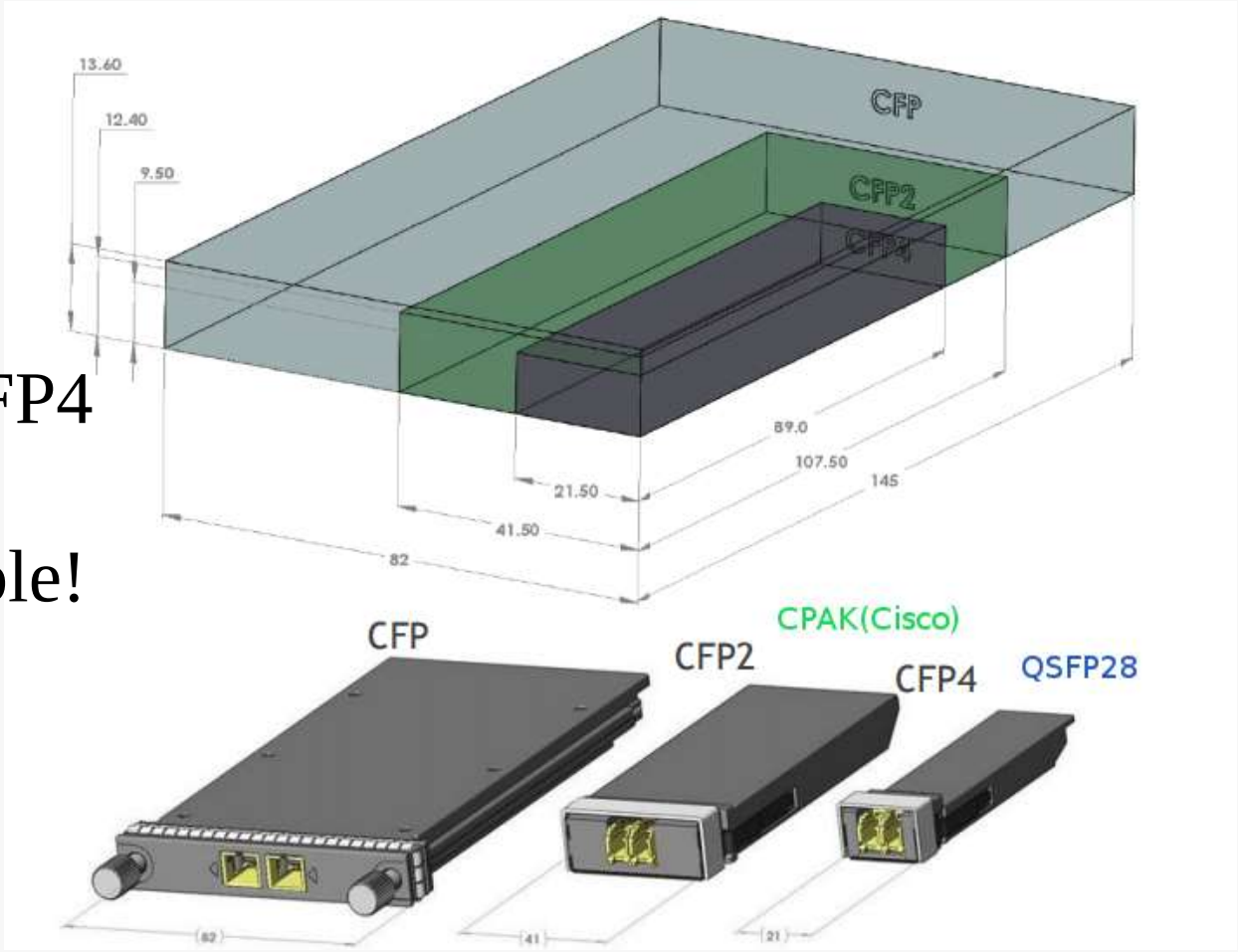
- QSFP+ form factor (alternate to CFP for 40G)
- 4 lanes at 10G – each same spec. as popular SFP+
- 40G for inter-switch, 10G for servers
- Fan-out QSFP+ to 4 x SFP+ cables for high density server connections
- Twin-ax connectors inside rack
- SR4 links between switches. MPO/MPT OM3/OM4 – gives 100m/150m distances
- Few servers have 40G connectors… We say 40G, but its largely 4 x 10G fan-out for servers.
- Next Step ? : QSFP28 with 4 x 25G lanes… ie. 100G in the Data Centre.

New 100G Standards

- 100GBASE-SR4 & 100GBASE-CR4 twinax cables
- IEEE 802.3bm-2015
- 4 lanes of 25Gbps CAUI-4
- QSFP28 module seems dominant
- All the benefits of 40G, but faster 25G lanes (2.5 x faster)
- Backwards compatibility with 10/40G possible.
- OM3/4 distance reduced to 70/100m – won't affect many
- New servers at 25G SFP28 (¼ of QSFP28)
- 25G just standardized: IEEE 802.3by 30th June 2016.
- 50G not yet standardised – 1 lane or 2 ?, perhaps both ?
- Switch vendors & Mega Data centres are keen.

100G Module Evolution

- CFP2: 4 or 10 lane @ 10,25G
  - 200 & 250G
  - DWDM-ACO
- Cisco CPAK
- QSFP28
  - Similar size to CFP4
  - Lower power
  - QSFP+ Compatible!



A 100G NIC – Mellanox ConnectX-4

- QSFP28 interface
- Finisar QSFP28-LR4 optic
- 16 lane PCIe3.0: 16x8Gbps = 128Gbps
- ConnectX-5 now announced

Connecting old & new

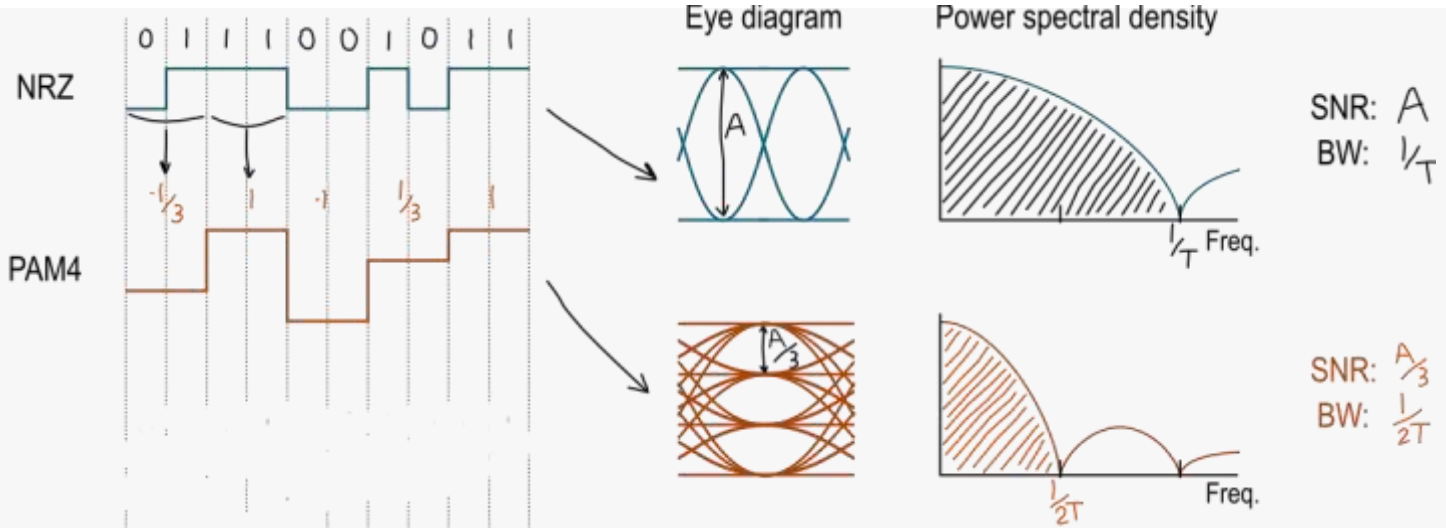
- Connecting the 10 lane CFP to 4 lane QSFP28
- 100GBASE-SR10 doesn't talk to 100GBASE-SR4
- 100GBASE-LR4 available in CFP & QSFP28
  - Should "just work"
  - 100G NIC to Juniper CFP
  - We'll be doing more tests ☺
- What QSFP28 module type to use ?
  - Twinax within racks (cheaper, or SR4 if more reliable)
  - SR4 between racks - 8/12 Fibre MPO limited to 70/100m range
  - Should still suit most Uni's, corporate data centres
  - SR4 optics still significantly cheaper
  - LR4 singlemode for longer links

Beyond SR4 – Megascale Data Centres

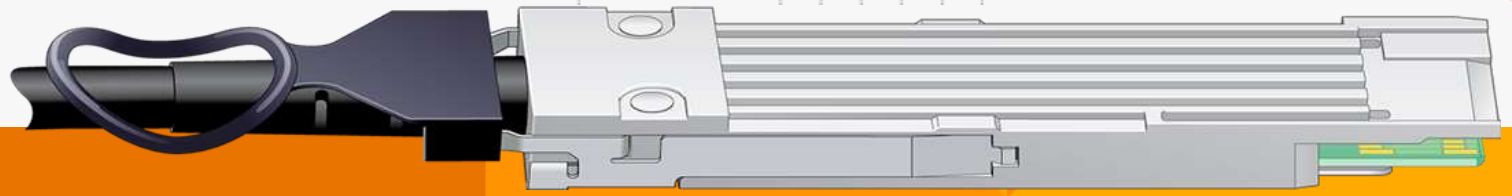
- Big operators are going singlemode
- LR4 optics too expensive ~ $5k ?, SR4 ~ $1.5k
- New competing SM standards cost <= ~ $2k eventually
  - Target range of 500m or 2km
  - CWDM4 20nm spaced CWDM – just like 40GBASE-LR4
  - CLR4, similar, interoperable?, FEC is optional -> lower latency
  - PSM4 – Parallel singlemode
  - Openoptics MSA: 4 – 32 wavelengths around 1550nm
    - Backed by Mellanox
    - Future roadmap to 1.6Tbps
  - The winner is uncertain
- Recommend: stick with SR4 & LR4 while the battle rages!

Beyond QSFP28

- Smaller uQSFP – more 100G ports (72) in 1RU faceplate
- QSFP-DD – Higher power dissipation
  - Backwards compatible with QSFP28 & QSFP
  - Extra row of pins enables 8 lanes
  - Electrical lanes to PAM4 -> 50Gbps x 8 : 400G client optic



QSFP-DD



Future Modules

- A CDFP has been defined
  - 400G – 16 lanes of 25Gbps NRZ
  - Needs more fibres than an MPO-24 – maybe 32,36 or 48 ??
- 400G Ethernet is within reach, 200G will be defined like 50G
- Terabit Ethernet – was predicted for 2015
- must wait for 100Gbps Electrical Lanes
- won't be defined with 400G, like 40/100G
- Not feasible until after 2020

Summary

- 100G dominant in carrier space
- 100G widely deployed within AARNet
- 40G has been popular in the Date Centre
- 100G with 4-lane QSFP28 will be next in the Data Centre
  - Backwards compatibility with 40G
  - 25G lanes backwards compatible with 10G lanes
- AARNet can help:
  - We're experienced with 100G & trying to stay across it
  - Chat to us about what's possible with 100G in data centre, across campus or across town.