I Wanted a New Network

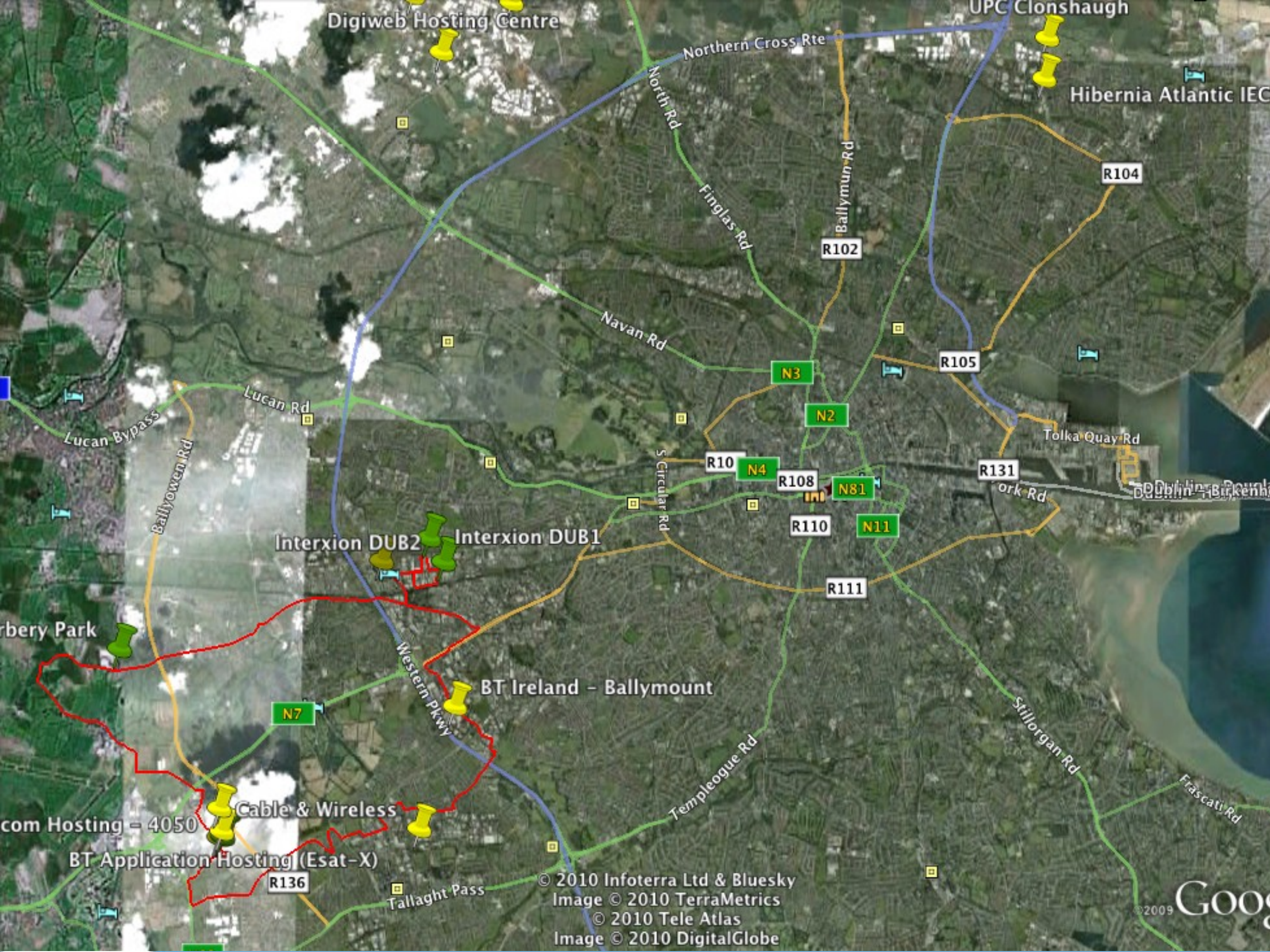
And a Pony, pls
And a Pony, pls
And a Pony, pls

inex
internet neutral exchange

Nick Hilliard
Head of Operations
nick@inex.ie

Where it all started

- INEX had been looking to expand into a third node for some while
- Constrained primarily by financials
  - Revenue vs 5-year opex + capex TCO
- Interxion DUB1 provided opportunity for expansion
  - INEX had low cost option to break out fibre spur
  - Questions remained about cost of implementation
- Networking kit
  - Existing kit very reliable, good quality, very pleased, etc
  - For INEX's requirements, too expensive
  - Most ethernet switches are not suitable for IXPs
    - generally related to absence of advanced features that aren't present on cheaper switches


Digiweb Hosting Centre
UPC Clonshaugh
Hibernia Atlantic IEC
Northern Cross Rte
North Rd
Finglas Rd
Ballymun Rd
R104
R102
Navan Rd
N3
R105
N2
Lucan Rd
Lucan Bypass
Tolka Quay Rd
R10
N4
R108
N81
R131
S Circular Rd
R110
N11
Interxion DUB2
Interxion DUB1
R111
Western Pkwy
BT Ireland - Ballymount
N7
Templeogue Rd
Stillorgan Rd
Frascati Rd
Cable & Wireless
BT Application Hosting (Esat-X)
R136
Tallaght Pass
© 2010 Infoterra Ltd & Bluesky
Image © 2010 TerraMetrics
© 2010 Tele Atlas
Image © 2010 DigitalGlobe

InterXion Dublin
Hosting365
Ninth Lock Rd
Nangor Rd
R113
R110
Lock Rd
DEG - Kilcarbery Park
Boot Rd
BT Ireland - Bally
N82
Eircom Hosting Centre (4050 Kingswood Avenue)
BT Application Hosting (Esat-X)
Cable & Wireless
Telecity Dublin
Belgard Rd
Tallaght Pass

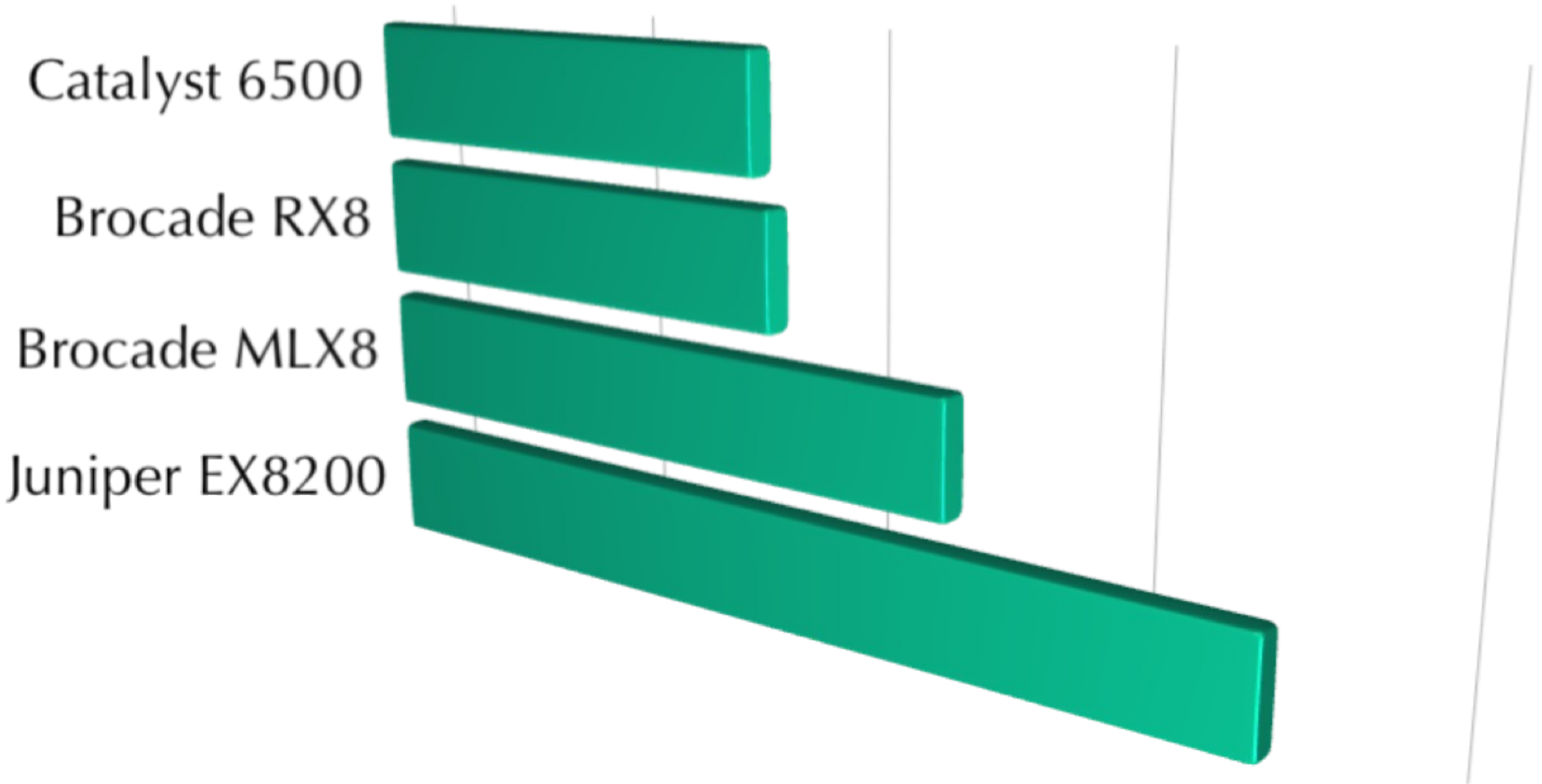
Looking Forward

- Where will INEX be in 3-5 years time?
  - Probably a lot more 10G
  - Probably less network growth due to stabilisation of Internet market
  - Probably significant reduction in port charges
  - Capex and op-ex costs very important
  - Decided that a medium density installation is best
- From technical point of view
  - How do we provision medium density 10G in future?
  - Where are things going with optics and networking kit?
  - How do we deal with lighting up metro links?

Procurement: Stage #1

- Naive approach:
  - Look at what other IXPs are doing and do the same thing
  - Guaranteed to work well
  - Potential cost problems
- Initial approach to vendors consisted of:
  - Specification
    - 10/100/1000 copper, 24-48 port
    - a bunch of SFP ports for GE over single-mode / multimode
    - A bunch of 10G ports - expandable
    - All ports non-contended (i.e. no over-subscribed 10G blades)
  - We duly got replies for
    - Cisco: Catalyst 6500
    - Brocade: RX8 / MLX8
    - Juniper: EX8200

Procurement: Stage #1 Costs



Procurement: Stage #1 - Stark Realisation

Procurement: Stage #1 - Budget

Procurement: Stage #2

- In-depth approach:
  - Analyse exactly what we need in terms of tech requirements
  - Look at suitable market products from known vendors
  - Slightly riskier, but who knows...?
- Starting out from the bottom, we needed
  - A whole bunch of technical features
  - Expandable support for 10G ports
  - A long term approach to 10G optics
  - Both SFP and 10/100/1000Base-TX support
  - Sane integration with the WDM network
  - Low power consumption
  - Critically: must be affordable

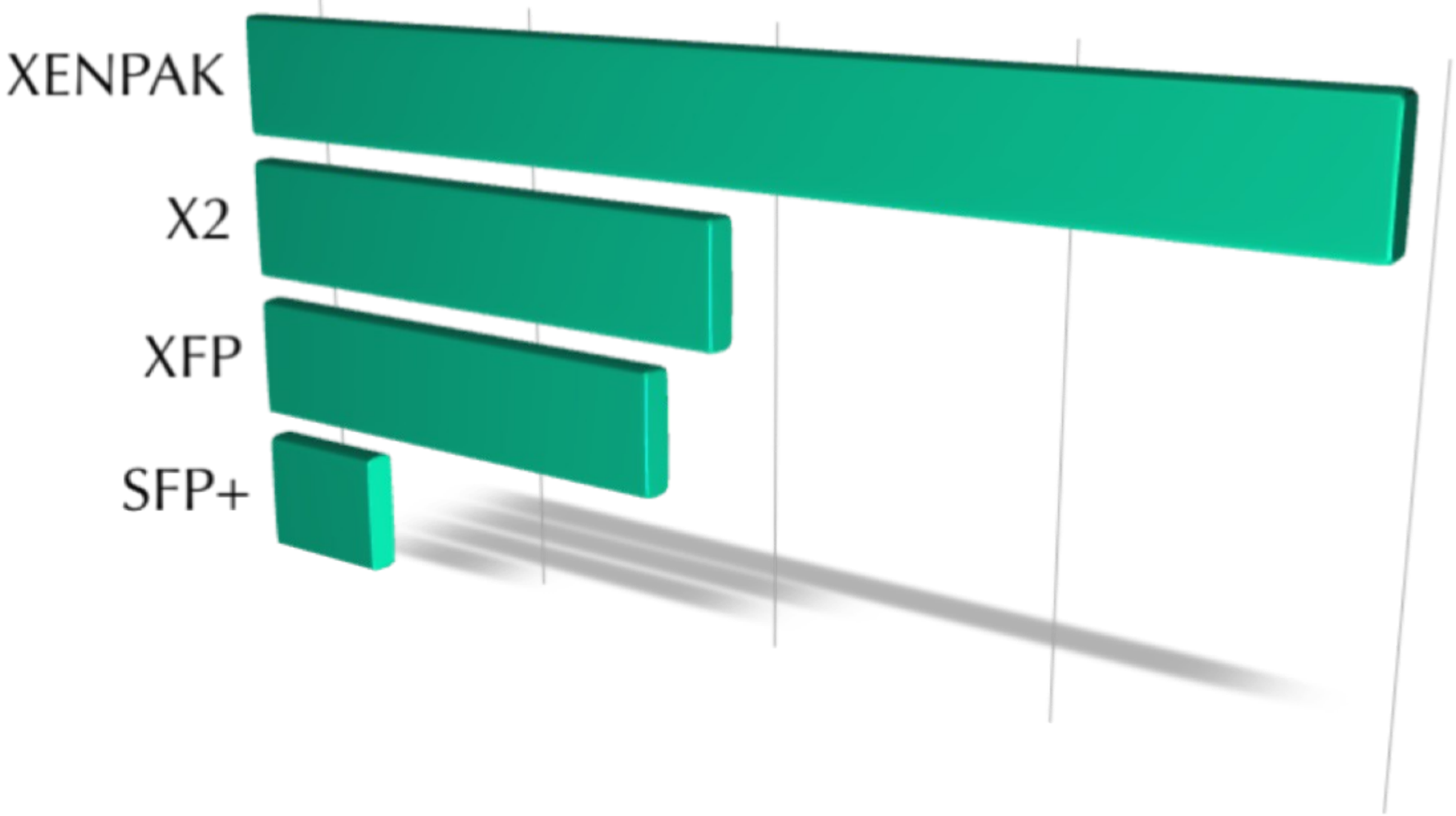
Procurement: Stage #2 - Wishlist

Wire Speed on all ports	Unicast flood control
Wire Speed L2, L3 filtering	OEM Optics
IPv6 ACLs on L2 interfaces	TDR support on TX ports
DHCP Snooping	Link aggregation with full features
IPv6 RA Guard	Port mirroring
PIM Snooping	Remote port mirroring
IGMP Snooping	Rapid spanning tree
MLD Snooping	BPDU guard
Dynamic ARP inspection	Bridge management other than than STP
Port security (mac address counting)	SSH CLI management
Sflow / Netflow	UDLD
Mac address accounting using ACL counters	Environmental monitoring
Broadcast / multicast storm control	Dual Hotswap PSU

Procurement: Stage #2 - The Optics Conundrum

- Let's recap on some 10G transceiver horror:
  - Many standards, the most common of which are:
    - XENPAK (almost obsolete)
    - X2 (used exclusively by Cisco and HP)
    - XFP (used by everyone else)
    - SFP+ (new kid on the block)
  - Grey: SR (multimode), LR (single mode), CX4 (copper)
  - Coloured: CWDM and DWDM
  - Vendor locking a rampant problem
- But it gets worse:
  - Coloured XENPAKs only made by Opnext (prev. Hitachi)
  - No competition to drive down X2 prices
  - Due to tiny size, SFP+ is low power and cannot drive long-distance fibre runs (i.e. no coloured SFP+)
  - SFP+ uses analog signalling out of transceiver

Transceiver Power Consumption

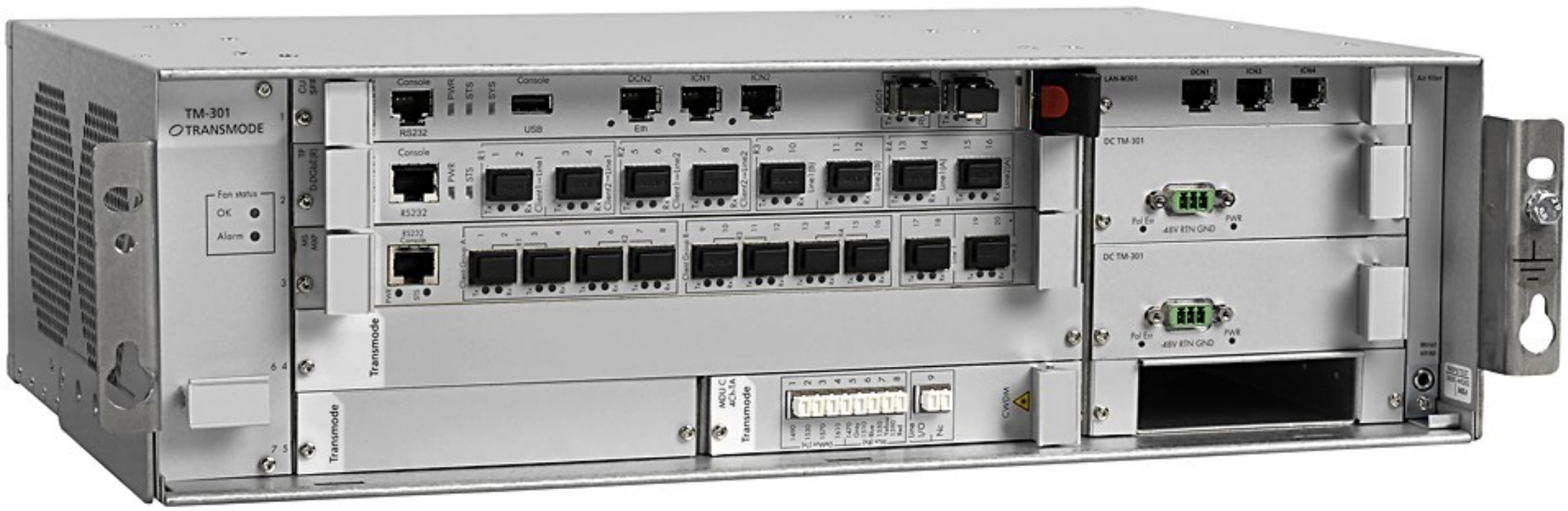


Procurement: Stage #2 - The Optics Conundrum

- On the other hand
  - XFP is widely used for coloured metro (i.e. good choice here)
  - SFP+ is really low cost and highly suitable for LANs
- So L2 choice boiled down to:
  - XFP-capable chassis with coloured optics
  - XFP to something (SFP+ or XFP) using WDM transponders

WDM at INEX

- INEX uses WDM because we have multiple LANs and only single fibre ring
- Each ring segment currently split into 4 channels
- 2x10G for production traffic, 1x1G for management, 1 free
- Didn't buy coloured XENPAKs due to cost / support / flexibility



Procurement: Stage #2 - Top-of-Rack

- Started evaluation of “Top-Of-Rack” switches to be used in combination with 10/100/1000 access switch.
  - All arrived on the networking scene in 2009
  - 10G cut-through
  - 20-40 ports in 1U or 2U
  - Mostly SFP+ form factor - supports SFP too!
- Started detailed beauty contest with:
  - Cisco (Nexus N5K)
  - Brocade (TurboIron TI24X)
  - Arista Networks 7124
- Due to lack of features, did not evaluate:
  - Blade Network Technologies - RackSwitch 8124 (prev. Nortel)
  - Juniper EX2500 (rebadged BNT 8124)
  - Force10 S2410P
  - Extreme Networks Summit X650

Procurement: Stage #2 - Arista 7124

- Arista Networks set up in 2003 to produce low cost switches for cloud computing
- 5 switches available in different configurations
- Became clear relatively shortly that they didn't have the features we needed, in particular
  - Port security
  - MAC ACLs
  - Sflow / netflow
- On the other hand
  - Ridiculously low latency
  - ASIC programming possible from linux shell
  - Serious hacking possibility

Procurement: Stage #2 - Cisco Nexus N5K

- Nexus 5000 is interesting new TOR switch from Cisco
  - Runs NX-OS, not IOS
  - can use N2K as a port extender for copper ports
  - Unfortunately, this configuration had limitations
    - problems with link aggregation
    - only supported 1G - no support for 10/100
    - otherwise, makes quite a compelling combination
- Combination of N5K + Catalyst 3560
  - Neither device supported netflow / sflow
  - Both switches support OEM optics using special command, although DOM is hobbled
  - C3560 is feature rich, very stable, high quality switch, but had certain limitations:
    - No internal support for dual PSUs
    - External RPS still not 100% reliable
    - link aggregation limitations
  - N5K:
    - High nominal power consumption, no MLD snooping

Procurement: Stage #2 - Brocade

- Combination of Brocade TurboIron TI24X and FES-X624
- TI24X is a completely new switch - Started “shipping” in August 2009
- FES-X624 is the current generation SP access switch
- Very good compliance with our tech wish-list
- Some minor niggles:
  - TI24X doesn’t support IPv6 ACLs on L2 ports
  - Neither device supports remote port mirroring
  - Like Cisco, can operate with OEM optics but won’t do DOM
  - TI24X doesn’t support MRP (yet)
  - Will take a little work to integrate with our provisioning system
- On the positive side
  - Nominal power consumption significantly lower for TI24X than for Cisco N5K.
  - Both switches support internal multiple PSUs
  - MRP provides much better bridging management than RPVSTP+

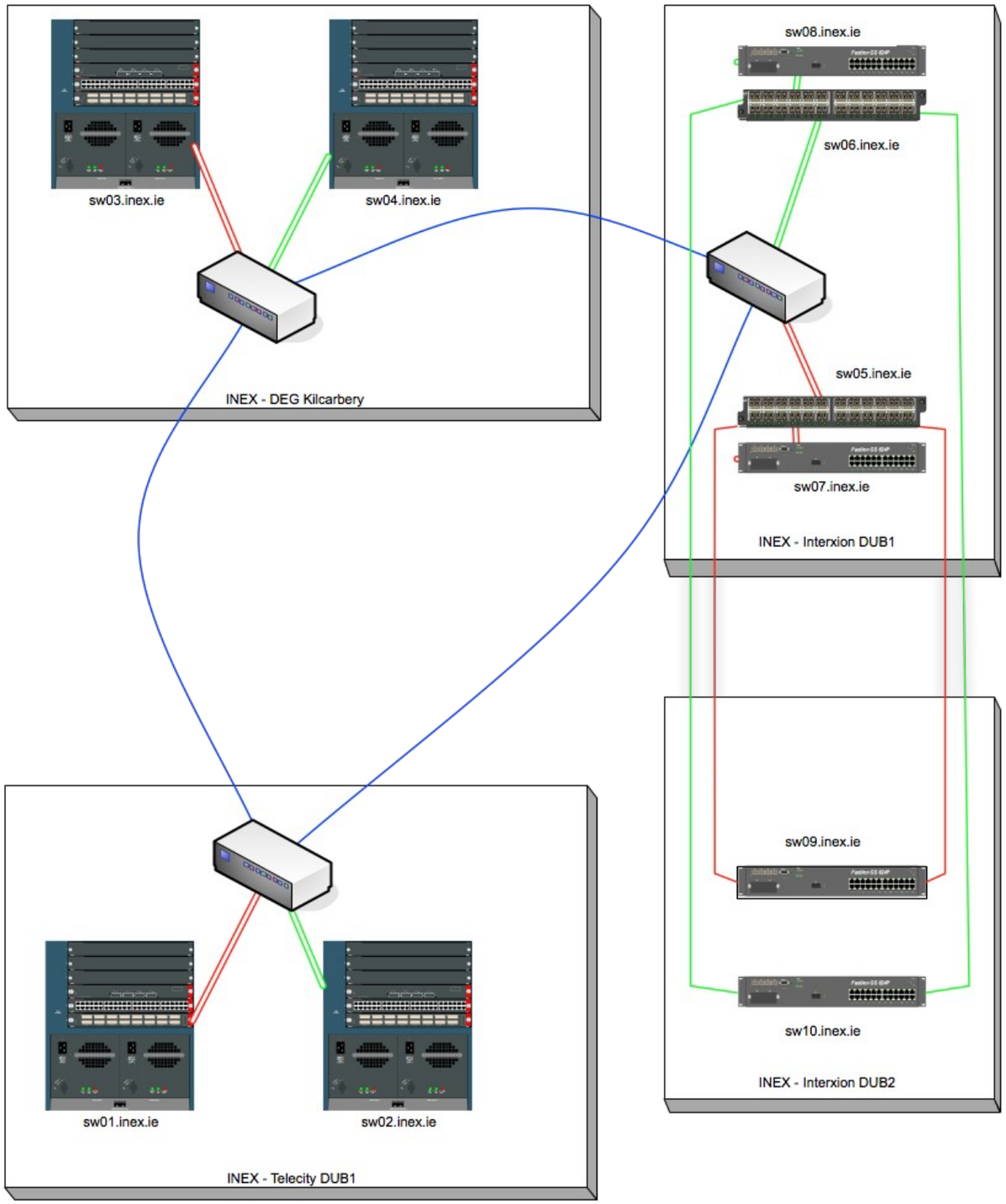
Procurement Results

- INEX made a decision to go ahead with the Brocade
- Best of all, we managed to get two PoPs:
  - Interxion DUB1 and DUB2
  - LAN1 and LAN2
  - At 45% of the capex of kitting out IX-DUB1 with big chassis boxes
    - Boxes dovetail into each other
      - Turboiron 24X provides 24 x 10G + 4 x 10/100/1000 + Mgmt port
      - FES-X624 provides 24 x 10/100/1000 + 2 x 10G
- Go-live for Interxion-DUB1 was end of October, 2009
  - “Project Slightly Behind Due to Unforeseen Circumstances”
- Go-live for Interxion-DUB2 was end of April, 2010
  - “Project Slightly Behind Due to Unforeseen Circumstances”

Subsequent Testing

- Did some testing on this networking gear
  - Extremely grateful to LINX for use of kit + driver
  - RFC-2544 and Latency
- Confirms the following
  - TI24X handles linerate 80G full duplex, 64 - 1518 byte packets
  - Didn’t have any more transceivers to test higher speeds
  - Latency is ~5µs (typical store-n-forward is > 16µs)
  - FES-X624 10G ports will handle linerate at 64-1518 bytes, 9µs
  - SNMP Port counters appear to work correctly at these speeds
- Didn’t get to test
  - port contention at scary speeds and 1G latency / throughput
- Oops, some features missing
  - Port security, L2 ACLs on TI24X
  - FESX-624 limitations with L2 ACLs, snooping + sflow

This Is What We Got



inex
internet neutral exchange

No Pony

That's all folks...

...any questions?