

HSE

David Ofelt, Jeffrey Maki, Brad Turner : Juniper Networks

Mark Nowell, Gary Nicholl : Cisco

2012-07-18



BANDWIDTH GROWTH

BWA shows the exponential growth of network bandwidth

The longer between new interface speeds:

- the faster the new speed needs to be
- the larger the technical jump is required
- the more uncomfortable the customer is before the next transition

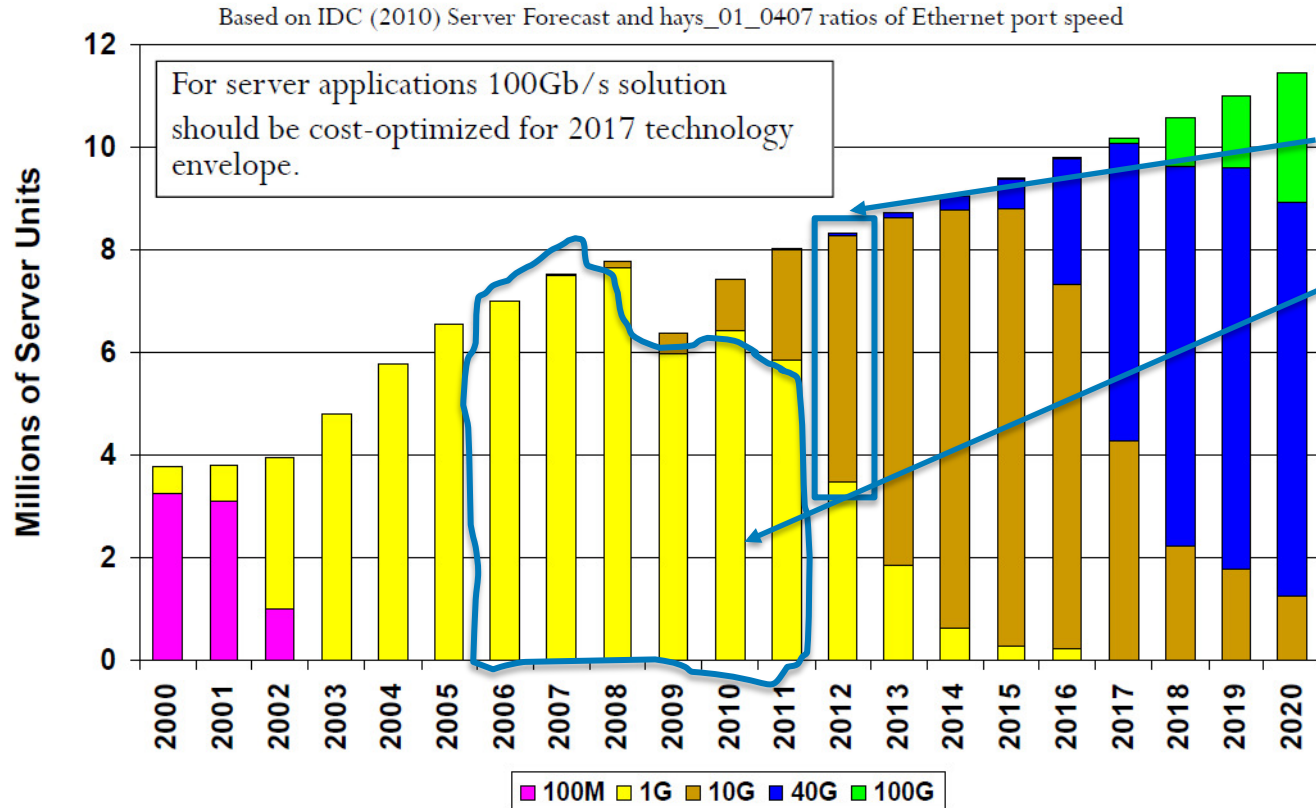
Iterating more quickly:

- Provides the customers with better tools more predictably
- Allows for smaller increments in technology
- Increases the odds that the later generations of the previous speed provide technology for the new speed.

EXPONENTIALS ARE TRICKY

x86 servers by Ethernet connection speed (2010 forecast)

From the 100G Cu Backplane & Twinax CFI



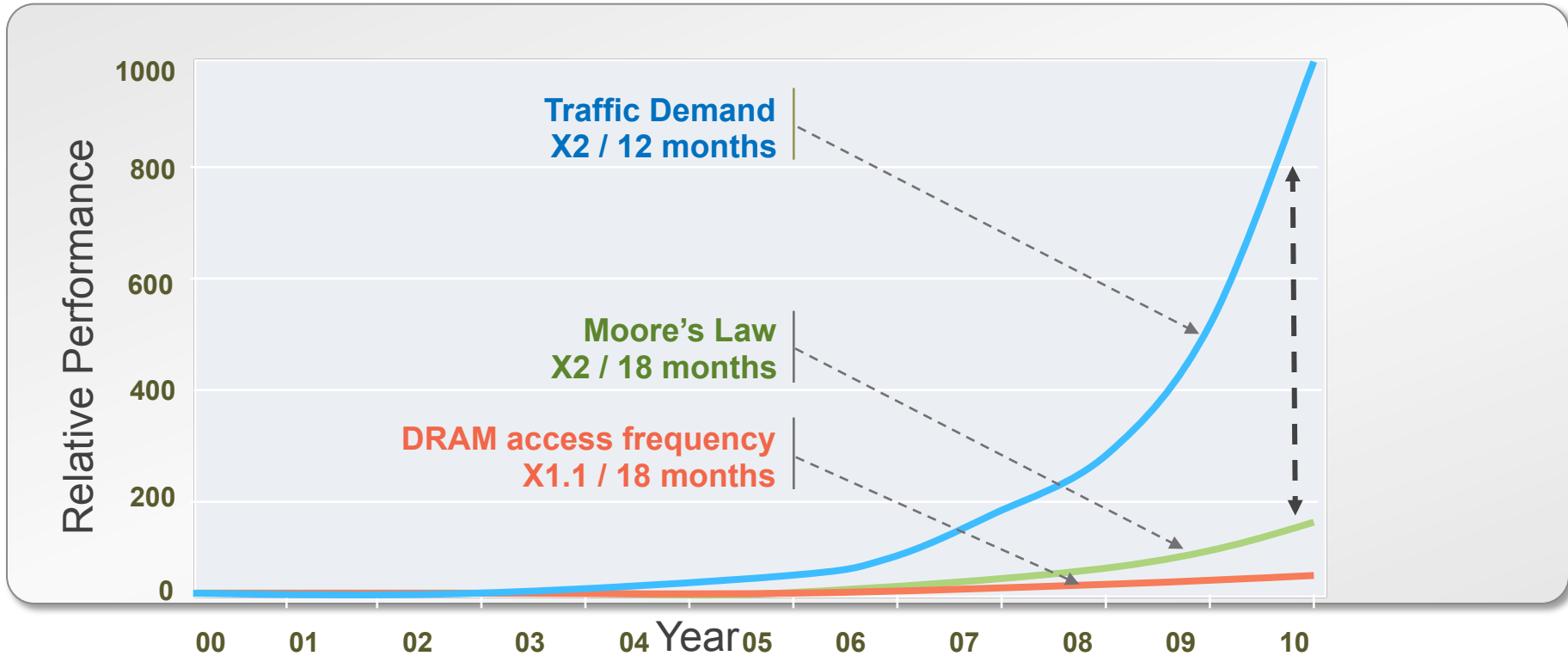
2012 10GE total Port bandwidth is Same as previous 6 years of 1GE!

100GbE Backplane/Cu Cable CFI
IEEE 802 Plenary, Dallas, TX, Nov 2010

November 9, 2010

MORE EXPONENTIAL FUN

Networking and semiconductor industry trends



Note: Based on ISSCC 2010

Our problem is getting harder faster than our tools are improving

CONTINUUM –VS- BIG JUMPS

Systems evolve faster than interface speeds

At any point in time, there may be:

- a sweet-spot interface speed
 - System/line cards designed around these
- a previous-generation interface speed
 - May not be able to do full density due to face-plate issues
- a new interface speed
 - Density should be proportional to the sweet-spot speed

A newer-generation interface should leverage current-generation building blocks.

IMMEDIATE APPLICATIONS FOR HSE

Core-to-transport handoff

- 400Gb/s transport gear has been demonstrated and is shipping

Core-to-core interconnect

Datacenter-to-datacenter connections

Datacenter-to-internet handoff

Upper layers of datacenter interconnect

- Not going to argue about datacenter architecture
 - Server layer is definitely not HSE
 - Top-layer can definitely use HSE
 - Middle bit can be argued a number of ways

WHY 400GE INSTEAD OF 1TE?

Customers want parity in j/bit, \$/bit, and bits/system

The faster the interface, the more exotic the implementation

- Not competitive per W, per \$, or on density.
- More research and investment required – can only cause delay

400GE can reuse 100GE building blocks

400GE fits in the dense 100GE system roadmap

Next rate will follow the same pattern using 400GE as starting point

LARGER PIPES –VS- LAGS

Traffic is often trunked into large tunneled flows

- Harder to find enough entropy to do hashing well
- LAGs are inefficient due to limitations of hashing
- Flows are hashed without taking into account bandwidth
 - This leads to imbalances
- A faster interface provides predictable performance

There are sources of large flows:

- Content distribution
- Secure traffic

Fewer things to manage can provide operational efficiency

- Bandwidth is growing exponentially
- Without faster links, link count grows exponentially
 - Therefore management pain grows exponentially

FOCUS

Tackle the parts of the market that need first-generation 400GE

Use building blocks that are based on commodity 100GE

Build a MAC/PCS and electrical interface that will support fullness of 400GE ecosystem

- 803.3ba is great base
- 16x25G PCS lanes & CDAUI lanes
- Probably build in FEC cleanly from the start

Focus on the PMDs, media and reaches that will be successful.

- Ex: SMF - 2km-type reach

Longer term solution will:

- probably use the 3rd generation 100GE building blocks
- Can happen as potential volume is ramping
- Significant investment at that point is an easier sell



everywhere