

10GBASE-T and Wake-On-LAN

Bill Woodruff
Aquantia

10GBASE-T

an Energy Efficient Technology

- 10GBASE-T is the lowest power interconnect for the Data Center
- 10GBASE-T? Lowest Power?

Let's examine how compute capacity is deployed in the Data Center, and how the choice of interconnect impacts systems power

The Fully Deployed Rack

- Data Centers view a fully deployed 42RU rack as a basic building block
- Top-of-Rack switches are used for first layer aggregation switching

**HP ProLiant SL
Extreme Scale-Out server family**



Prototypical Rack for Comparison

- 42RU Rack contains:
- Forty 1RU Servers
 - Power/server assumed at 300W each
 - Converged network, two 10GE ports per server
- Two Top of Rack Switches
 - Forty downlinks to servers, eight uplinks
 - Nominal power of 10W/port, plus PHY
- Total of 160 10GE ports for within-the-rack interconnect

10GBASE-T Design Objectives

- 10GBASE-T satisfies Enterprise topologies
 - Structured cabling
 - 100m total cable length
 - Four in-line connectors, supporting two patch panels
- Designing a 10GBASE-T PHY requires remarkable effort
 - Crosstalk & Echo cancellation covering 100m
 - Analog resolution to receive microvolts from within full-duplex PAM-16 transmit signal
- Full-Reach 10GBASE-T does this in Sub 4W

Evolution of Direct-Attach

- Top of Rack Switch topology drives new class of interconnect
- Direct Attach copper first addressed this need with twinax copper with affixed SFP+ termination
 - Not able to be field terminated
 - Non-trivial connection technology
 - Full 10.3Gb/s signaling rate
 - 5m to 7m max distance, depends on EDC used



10GBASE-T Data Center Mode™

- “Within the Rack” will utilize a fraction of the capability within a 10GBASE-T PHY
 - At a fraction of the power!
- Data Center Mode™ :
 - Up to 10m of Cat6a cable
 - No patch panels
- Aquantia’s 40nm PHYs are 2W/port in Data Center Mode™
 - Greater than 40% power savings over full-reach

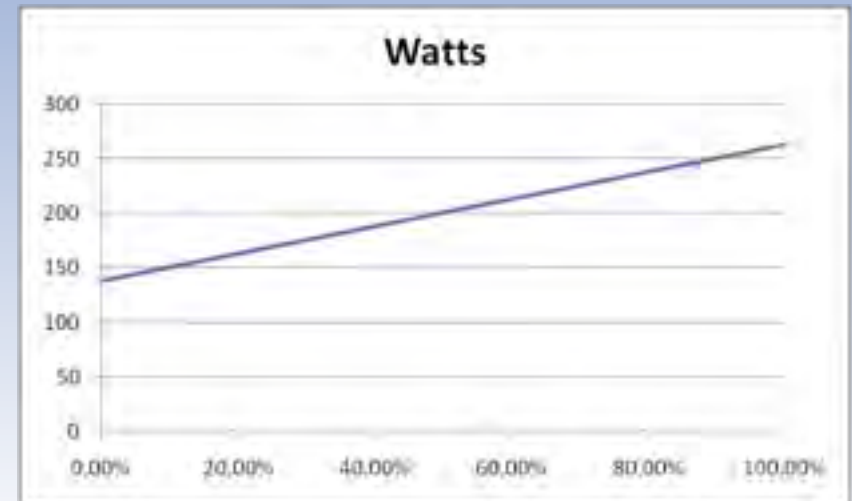
- Power with 10GBASE-T
 - Based on 2W/PHY for 10GBASE-T
 - 12,000W for Servers, 960W for Switches
- Power with Direct Attach Copper
 - Based on 0.75W/PHY for Electronic Dispersion Compensation
 - 11,900W for Servers, 860W for Switches
- 10GBASE-T is 1.5% higher power than Direct Attach copper (200W out of 13KW)
 - Not greener yet...

Impact of Virtualization

- Original focus of virtualization was the reduction in the number of servers
 - Increase flexibility while decreasing cost and power
- Dynamic Consolidation offered by Virtualization enhances concept
 - By consolidating tasks energy efficiency is increased, idled servers can be place in a sleep state
 - Should use for extended sleep times given finite time required to return to “wake”

Power as a function of Load

- www.spec.org contains logs of 130 machines collected over the past two years
- Average power at full load = 263W
- Average power at active idle = 138W
- Servers doing no useful work still consume 53% of full-load power

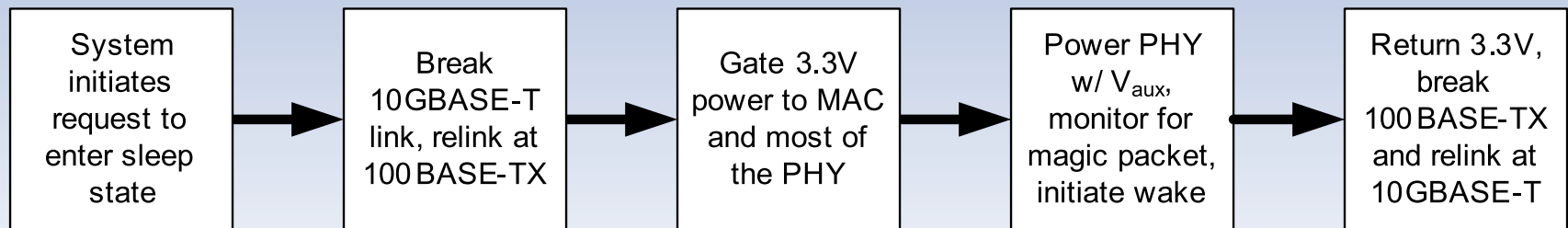


Wake on LAN

- WoL originally used by Enterprise client applications for 3rd shift maintenance
- WoL lets server enter sleep state while monitoring for “magic packet” to exit sleep
- V_{aux} provides just over 1W in sleep
- 10GBASE-T can use V_{aux} in ways that Direct Attach copper does not
- 10GBASE-T supports WoL and Dynamic Consolidation

10GBASE-T and the Magic Packet

- V_{aux} is the key! How to do network monitoring under one watt?
- Implement magic packet filter in the PHY!
 - 10GBASE-T PHYs include MAC monitoring logic
 - Simple to add the WoL magic packet filter!



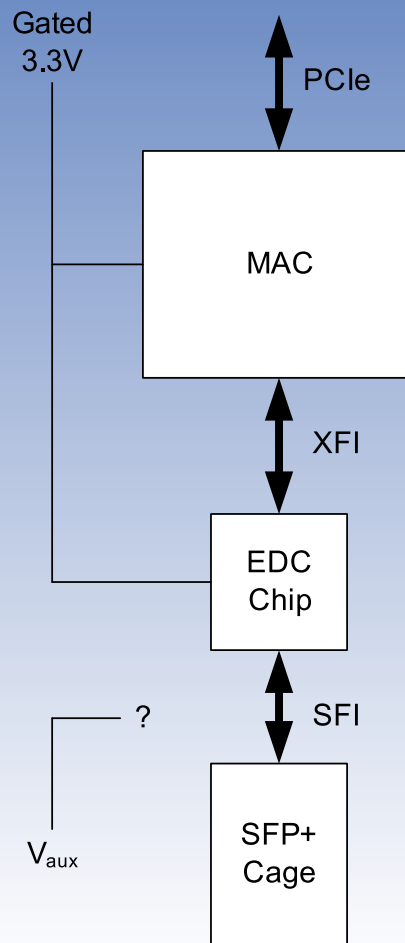
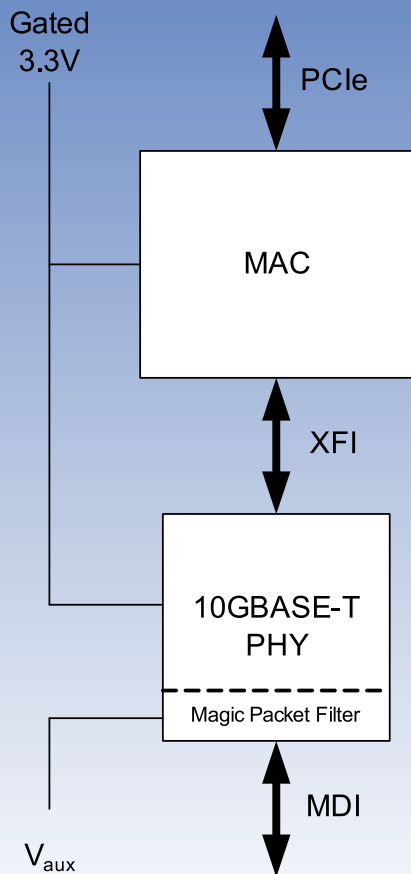
WoL under One Watt!

Using 10GBASE-T for Interconnect

Using SFP+ for Interconnect

When entering Sleep), turn off the PCIe 3.3V supply to the NIC, eliminating all standby and leakage power

Create power islands in the 10GBASE-T PHY to support 100BASE-TX link operation and the magic packet filter within the V_{aux} ~1W limit



3.3V needs to be switched off to enter S3 or S4, however MAC functionality only exists in MAC

Even just leakage in MAC would exceed the power limit of V_{aux}

Impact of Sleep – per Server

- Standby power provided by 5VSB supply
 - 5VSB not gated by power switch
 - Drives multiple PCIe slots and motherboard
 - Assume 9W to supply all standby requirements

Power in active idle 150W

Power in Sleep 9W

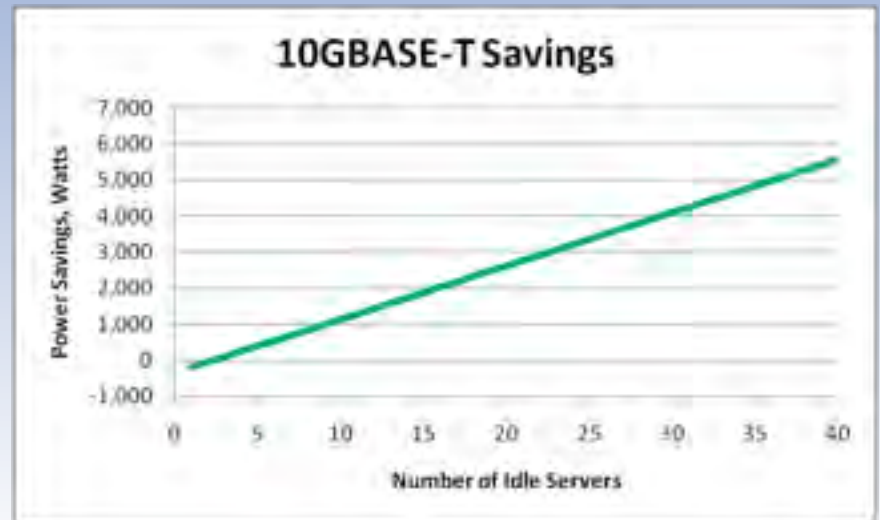
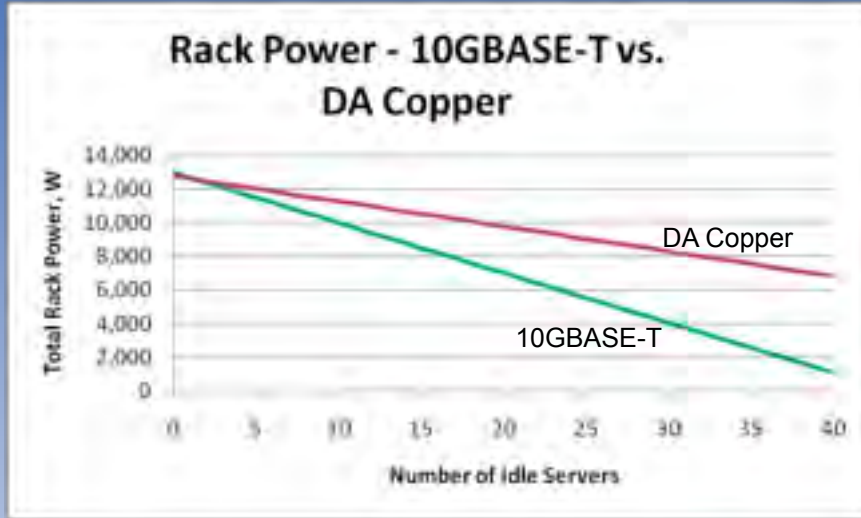
Power savings 141W

10GBASE-T is Getting Greener....

- Dynamic Consolidation now matters
- Put one server in Sleep – Save 141W
- Put second server in Sleep – Save 282W

Once dynamic consolidation puts the 2nd server in Sleep, 10GBASE-T becomes greener!

And greener, and greener



10GBASE-T is Greener!

- 10GBASE-T enables partitioning that creates opportunities
- System-wide power management can work with dynamic consolidation
- 10GBASE-T provides lowest overall system power!

10GBASE-T Wrapup

- Lowest Power interconnect
 - Dynamic Consolidation creates a new reality
- Flexible, User Friendly
 - Familiar RJ45, field terminate-able
 - Supports structured cabling, up to 100m
- Native networking connector on Servers
 - LAN on Motherboard within planning horizon
- Multi-rate
 - Simple 100/GE/10GE Autonegotiation
 - Easily supports WoL