An Overview of Energy-Efficient Ethernet

NGBASE-T Study Group

Michael J. Bennett
Lawrence Berkeley National Laboratory

Wael William Diab
Broadcom Corporation

Supporters

Alexander Umnov, Huawei

Dan Dove, Applied Micro

Alan Flatman, LAN Technologies

Oren Sela, Mellanox

Steve Carlson, High Speed Design

George Zimmerman, CME Consulting / Commscope

Topics

- Overview of Energy-Efficient Ethernet
  - Low Power Idle
  - Link Partner Communications
  - Things the study group should consider

What is Energy-Efficient Ethernet?

- Energy Efficient Ethernet (EEE) is a method to reduce energy used by an Ethernet device during periods of low link utilization

- Specified in IEEE 802.3az-2010™

- The premise for EEE is that Ethernet links have idle time and thus opportunity to save energy

- Specified for copper interfaces

  - "BASE-T' s'

  - Backplane (except 40G)

- The method is called Low Power Idle (LPI)

What is Low Power Idle?

- Concept: Transmit data as fast as possible, return to Low-Power Idle

- Saves energy by cycling between Active and Low Power Idle

  - Power reduced by turning off unused circuits during LPI

  - Energy use scales with bandwidth utilization
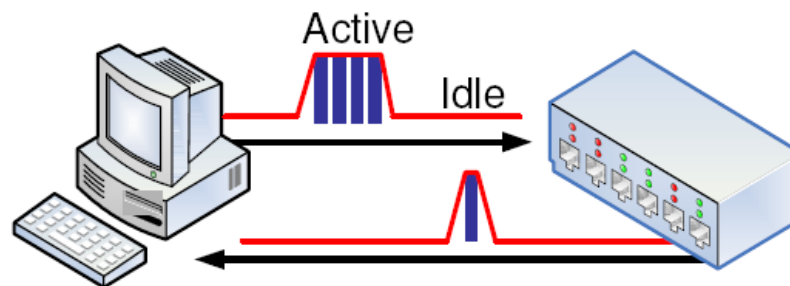
Example: link utilization

- Snapshot of a File Server with 1 Gb Ethernet link
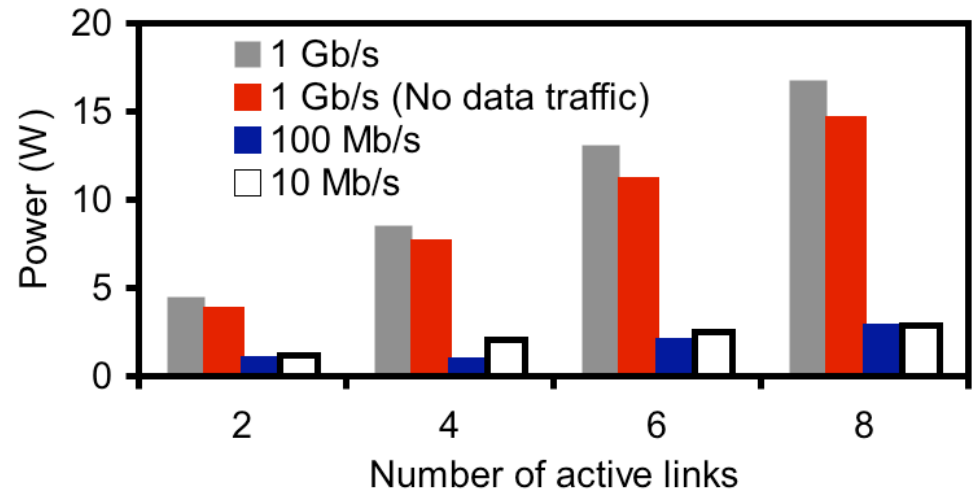 - Shows time versus utilization (trace from LBNL)

File Server Bandwidth Utilization Profile

Start time 12:33 PM 2/8/2007 (30 min)

utilization <=1.0 %
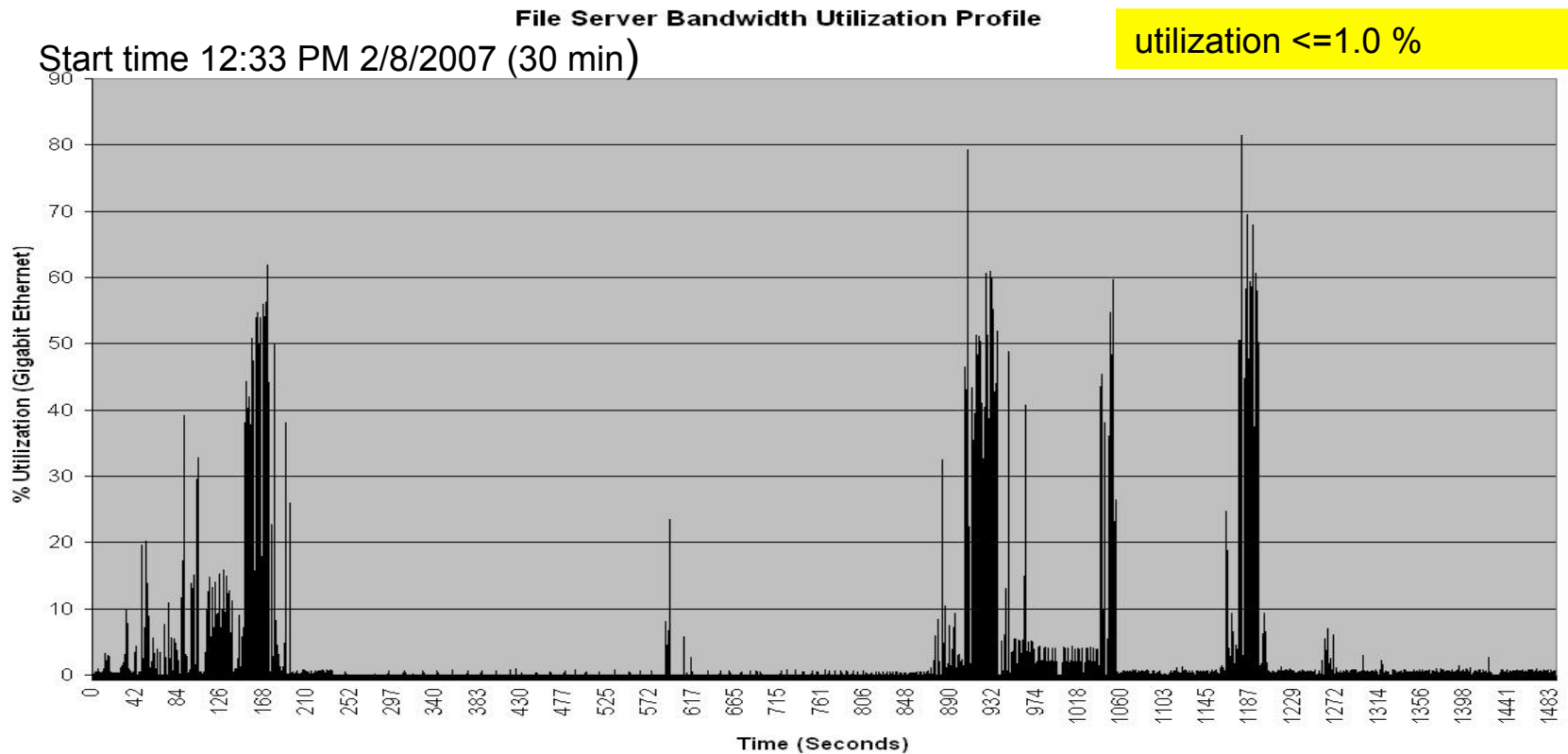
Link power

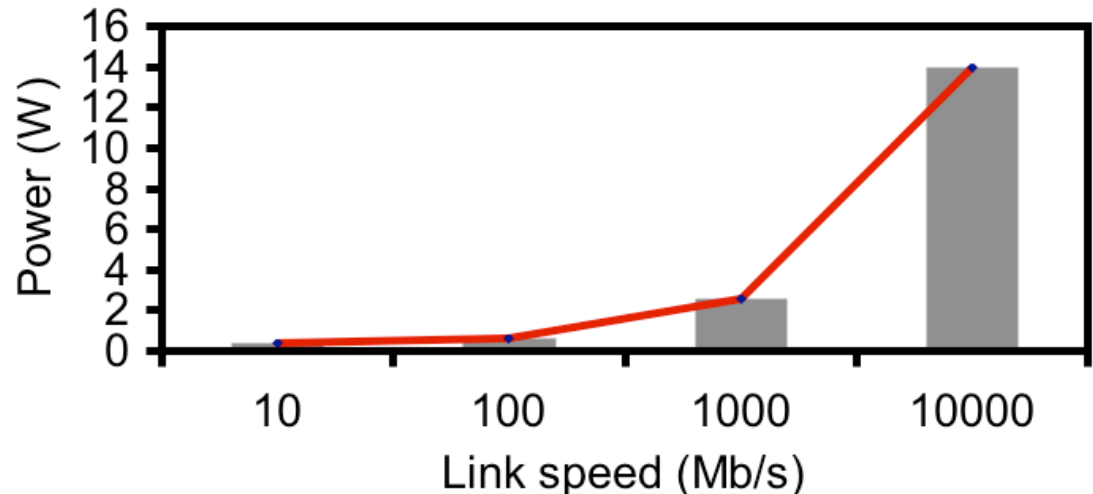## Results from (rough) measurements

— **all incremental AC power**

— **measuring 1st order**

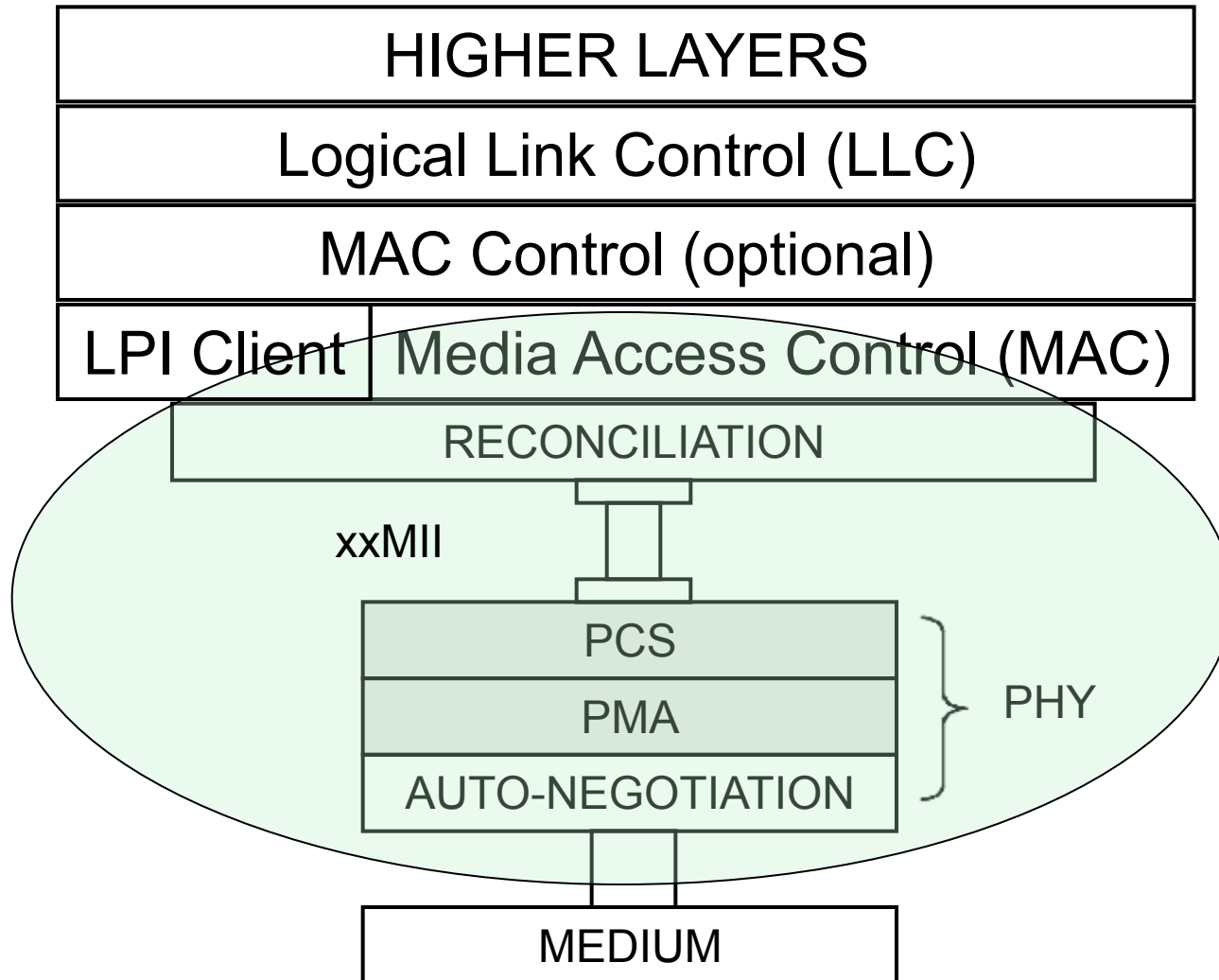- Typical switch with 24 ports 10/100/1000 Mb/s
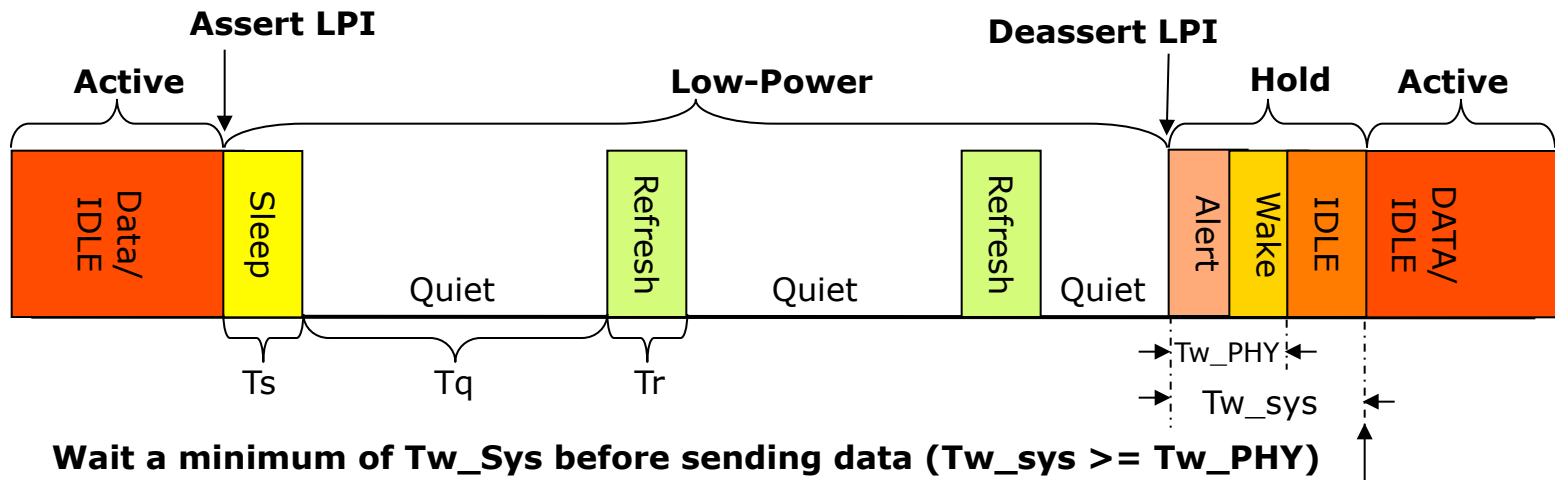

- Various computer NICs averaged

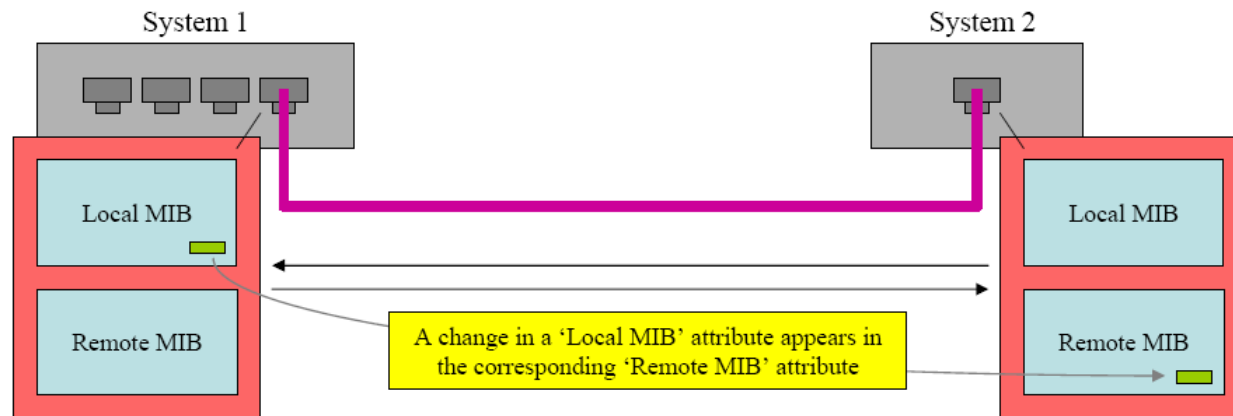Note: Measurement data from 2006

Where EEE Fits

LPI Overview

Wait a minimum of Tw_Sys before sending data (Tw_sys >= Tw_PHY)

- LPI – PHY non-essential circuits shut down during idle periods

- During power-down, maintain coefficients and sync to allow rapid return to Active state

- Wake times (Tw_PHY) for Twisted-Pair PHYs:
  - 100BASE-TX: <= 30 usec
  - 1000BASE-T: <= 16.5 usec
  - 10GBASE-T: <= ~8 usec (2 modes)

Link partner communication

- Uses auto-negotiation to notify link partner of EEE capabilities

- Uses Link Layer Discovery Protocol (LLDP) to notify link partner of parameter changes

  - E.g. control policy
    - User can choose energy savings preferred over performance or vice versa

System 1

System 2

Local MIB

Remote MIB

Local MIB

Remote MIB

A change in a 'Local MIB' attribute appears in the corresponding 'Remote MIB' attribute

Things to consider

- If the study group decides to do EEE for NGBASE-T
  - it is much better to do so earlier
  - Hard to retrofit

- Latency
  - It takes time to wake up the PHY
    - For example, 10GBASE-T has a wake time of ~ 8 usec
  - Fast wake being developed in P802.3bj (sub usec)
  - EEE is optional, so for the set of applications that can not tolerate increased latency (even in the order of a few hundred nano-seconds) it can be disabled

Things to consider

- An objective for EEE as an optional feature has been added for:
  - P802.3bj (100G Backplane and Copper Cable)
  - Reduced Twisted-Pair Gigabit Ethernet

- EEE is being considered as an optional feature for:
  - P802.3bm (40 Gb/s and 100 Gb/s Operation Over Fiber Optic Cables)

- The NGBASE-T Study Group should consider including EEE as part of the *non-controversial* objective set
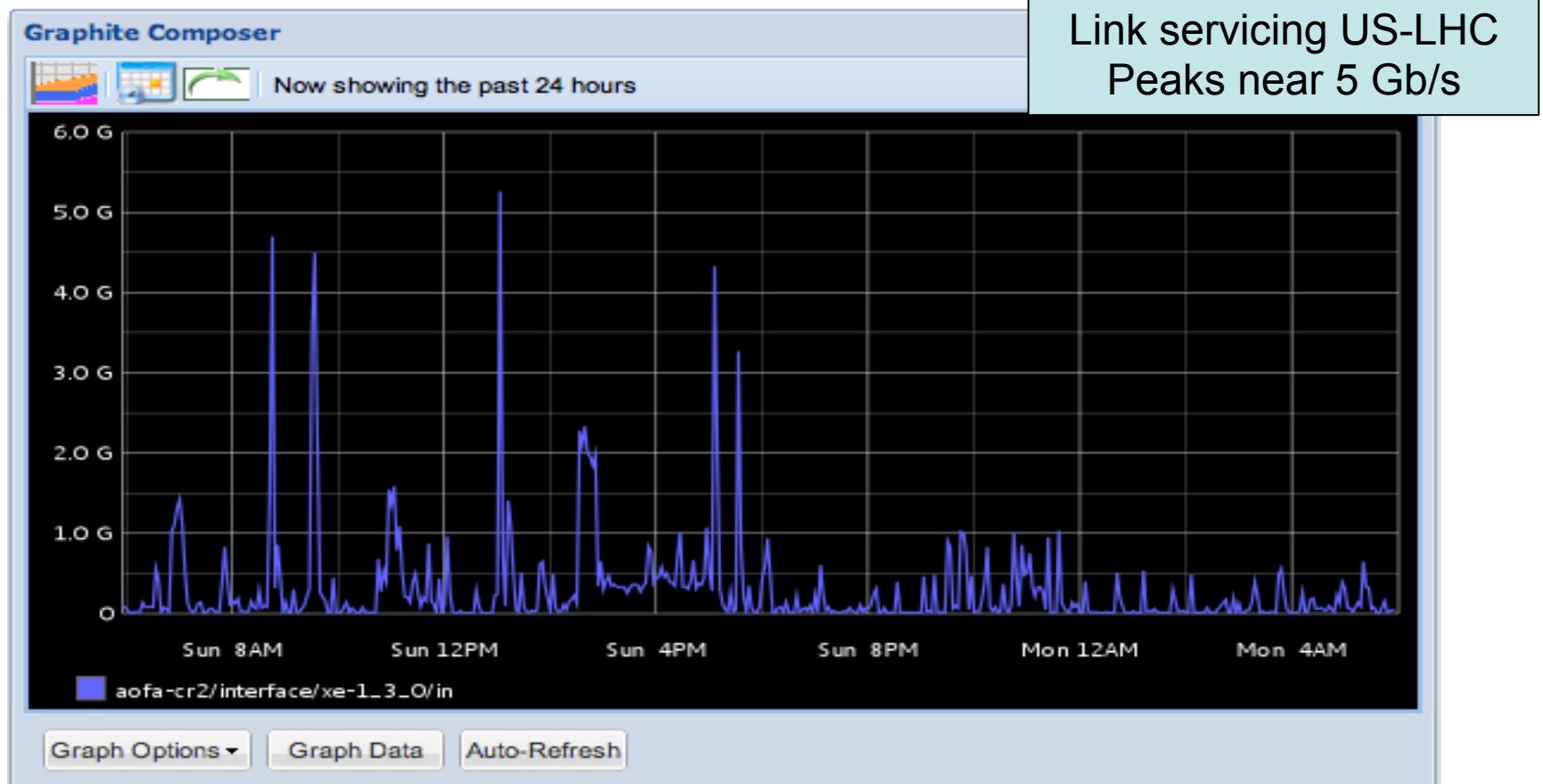 - *Proposed objectives for Next-Gen BASE-T*, by George Zimmerman
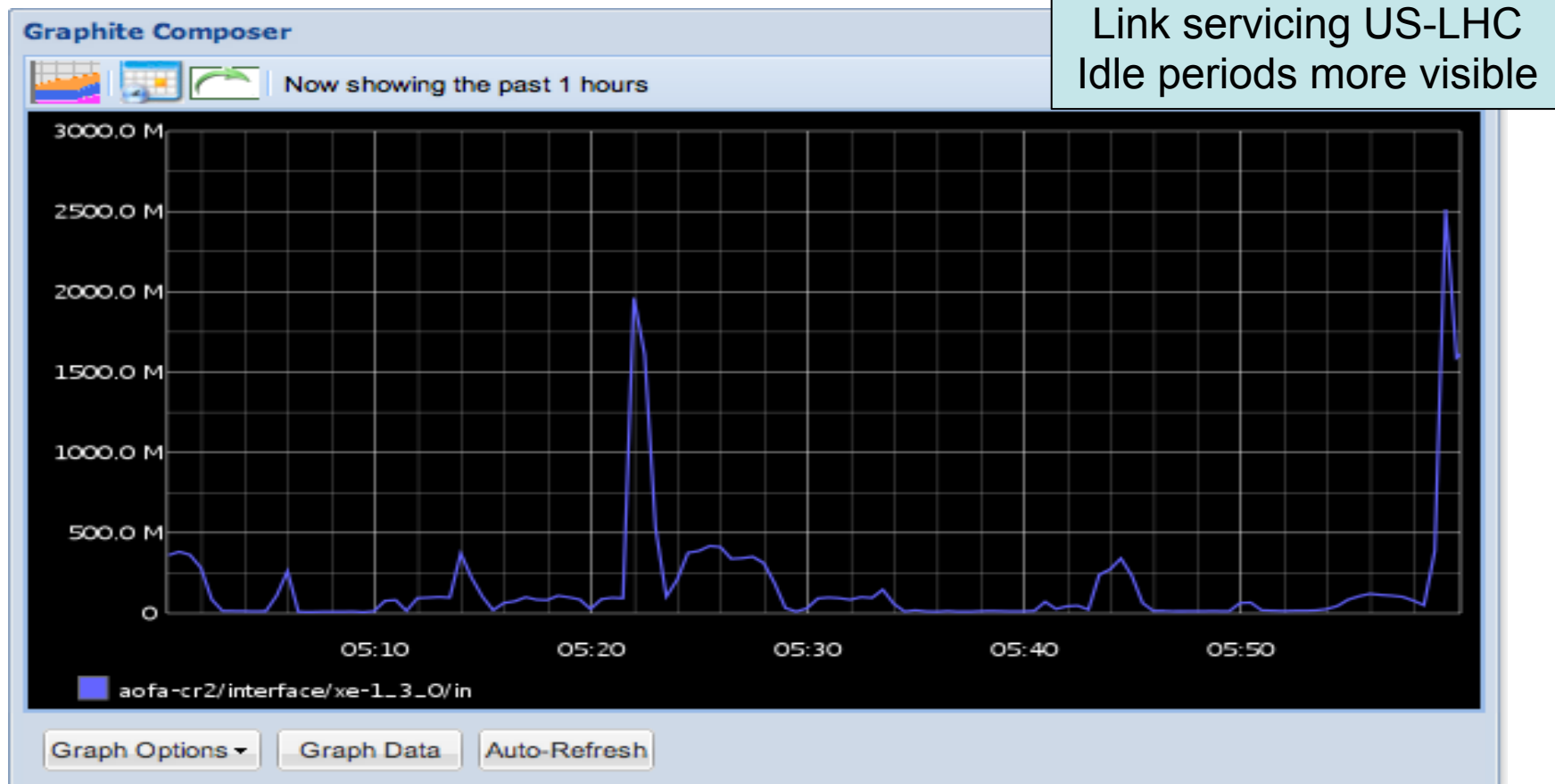
Questions?

mjbennett@lbl.gov

Thank You!

Extras

Example: link utilization

- Snapshot of a File Server with 10 Gb Ethernet link
  - Shows time versus utilization (trace from ESnet)

Link servicing US-LHC
Peaks near 5 Gb/s

Example: link utilization

- ## Snapshot of a File Server with 10 Gb Ethernet link
  - Shows time versus utilization (trace from ESnet)

Link servicing US-LHC
Idle periods more visible

Example: link utilization

- Snapshot of a File Server with 10 Gb Ethernet link
  - Shows time versus utilization (trace from ESnet)