Using 10GBASE-T Products for Energy Efficient Data Centers

George Zimmerman, Ph.D.
CTO & Founder, Solarflare Communications
February 26, 2010

ABSTRACT

10GBASE-T products provide data center managers and architects with unique capabilities to implement energy efficient data centers. Their use of structured cabling enables cooling and power distribution efficiencies through flexibility in equipment location, leveraging the hundreds of watts consumed by the server or switch. Solarflare® 10GBASE-T lowers the average power consumed by each port by using only the power each link length requires with its Dynamic Power Scaling™ feature. Additionally, 10GBASE-T enables Wake-on-LAN and saves power by turning off entire servers when they are not needed, saving hundreds of watts in each instance. In the future, 10GBASE-T will also provide IEEE 802.3az Energy Efficient Ethernet to further improve energy efficiency lowering consumption when the full capacity of the 10Gbps link is unused.

Overview

10GBASE-T offers 10Gbps Ethernet over structured cabling with a deployment model that mirrors traditional 1G Ethernet in enterprise and data center networks. However, getting 10Gbps speeds over twisted pair copper is a significant technical challenge and requires a great deal of signal processing, resulting in slightly higher power consumption than other 10G PHY options. However, 10GBASE-T offers substantial energy savings opportunities that go far beyond the PHY. This whitepaper details how 10GBASE-T saves energy in data center implementations today by allowing efficiency in power distribution and cooling, matching power consumption to cable lengths without requiring engineered links, and enabling server shutdown through wake on LAN technology. In addition, in the future, 10GBASE-T will support IEEE P802.3az Energy Efficient Ethernet, and allow further savings by minimizing power consumption during periods of low traffic. 10GBASE-T is the only 10Gbps technology to offer all this power saving potential to data center managers and architects, allowing savings on the hundreds of watts in server equipment and resulting in a more energy efficient choice.

TODAY: Flexibility in Data Center Arrangement

Since the advent of Fast Ethernet (100Mbps), IT managers and building planners have worked together to develop a set of infrastructure standards and practices to allow connectivity between equipment placed anywhere on a floor, commonly known as horizontal structured cabling. 10GBASE-T is the only 10 Gbps Ethernet technology which enables the use of structured horizontal cabling. As the industry focuses more on data center energy efficiency this fact becomes even more important, because it is widely held that cooling and power distribution account for around 1 watt of power consumption per 1 watt of power consumed [1]. As a result, groups such as the Green Grid (http://www.thegreengrid.org) recommend that the placement of power consuming electronics on the data center floor be engineered to optimize cooling and power dissipation efficiencies.

Reports have shown efficiencies of 15% or more are gained by proper placement [2]. Various techniques are used, largely based on the individual installation, including hot row/cold row arrangement, sealed, high-density arrangements, and use of blank panels (empty Rack Units) to

manage heat density in a range for optimum cooling efficiency. All of these are significantly constrained unless the data center architect has the ability to place equipment anywhere they choose on the floor, mixing equipment within the rack. 10GBASE-T enables such a placement strategy with 10Gbps links. Given that the savings is leveraged against the power of the entire end-point, e.g., a server, which might typically consume 200 watts or more, even a modest 2.5% cooling & distribution efficiency gain from flexible placement entirely offsets the entire 5 watt power of a 65nm 10GBASE-T device. Today, data center architects can improve efficiency of 10Gbps networks by using 10GBASE-T to enable optimization of cooling and power distribution arrangements. As the per port (PHY) power drops, or, as seen below, as the average power is lessened by other techniques, the economics of flexible placement become even more compelling.

TODAY: Averaging Power over Cable Lengths

Part of the value of structured cabling is that individual links do not have to be engineered. This is because the PHYs are capable of running 100m and the infrastructure is designed to be less than 100m. Early in the 10GBASE-T standard development, engineers recognized that PHY power consumption largely came from the receiver signal processing necessary to extract a signal at 100m reach. In order to motivate vendors to take advantage of the power savings possible when PHYs were running at shorter reaches, the IEEE Std. 802.3an-2006 (10GBASE-T) included a "short reach test mode", targeted at 30 meter distances. All compliant 10GBASE-T devices still needed to be able to go the full 100 meters, but would be able to save power when running on shorter lines. Multiple vendors have also announced "data center modes" running 10 meters or less, for use within racks, further saving power. Solarflare has innovated beyond the standard and implemented Dynamic Power Scaling (See Figure 1), which automatically reduces the receiver signal processing power on shorter lines. The resulting technology saves up to 20-30% of the PHY power on each line (depending on interfaces and technology node). Because the data center distances are mostly short, as shown in Figure 2, the average PHY power consumed today (65nm) will be closer to the 4 watts of a 30m link than the 5 watts of a full 100m link, further improving efficiency, ***all while maintaining the ability to connect on the occasional long link in a plug-and-play fashion (without engineering the link*).** Similarly, at 40nm, the PHY power scales down up to 25% for a 10m within-rack link. An important distinction is that Dynamic Power Scaling designed to interoperate with non-Solarflare link partners.

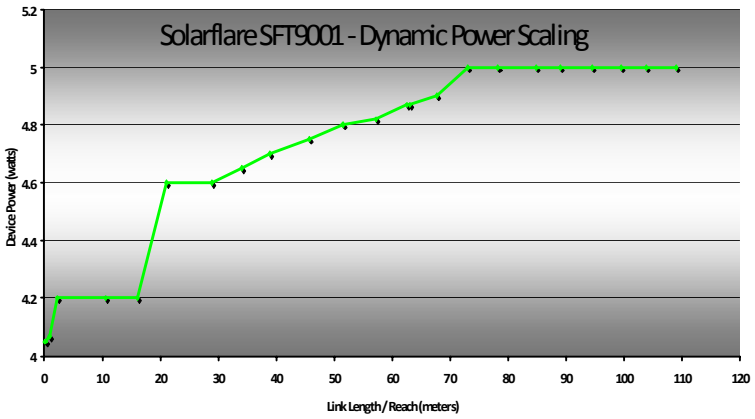

Figure 1: Solarflare Dynamic Power Scaling: Consumption vs. Line Length

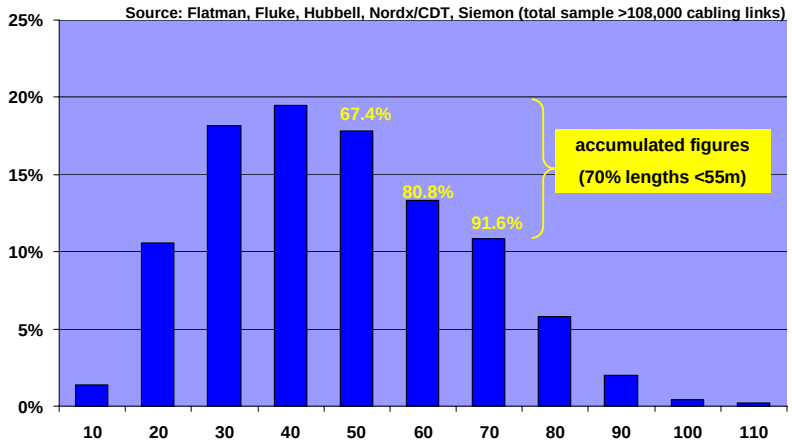

Figure 2: Cabling Length Distribution in Data Centers

TODAY: Turning off Unused Servers with Wake-On-LAN

To further improve efficiency, many 10GBASE-T offerings will include the ability to switch to a legacy 100BASE-TX state and implement the Wake-on-LAN protocol. This protocol, unique to BASE-T Ethernet PHYs allows the Server to go into a deep sleep state when not in use, and be brought back to life when a special packet is sent over the 100BASE-TX link. Through this mechanism, a 10GBASE-T PHY can drop in its power consumption to around 1 watt and, more importantly, the 200+ watts of server power consumption is saved when the server goes to sleep. Additionally the corresponding 200+ watts for cooling can be saved during periods when the server is not required, resulting in over 400 watts savings during periods when the server is idled. With the use of server pools and on-demand resource management through Vmotion and similar techniques, putting idle servers to sleep becomes an important power management tool. Generally, the capacity of data center servers is sized for peak loads which are seldom seen. Savings could easily be 20-30% of the entire server load, offsetting many times the 10GBASE-T PHY power consumption.

TOMORROW: 802.3az Energy Efficient Ethernet

In the next few years, 10GBASE-T will utilize IEEE P802.3az Energy Efficient Ethernet (EEE) which will allow PHY power consumption to scale with link utilization. (Note that SFP+ transceivers are not specified by the IEEE, and are not included in this standard, and the standard as is does not include optical 10Gbps links). When links are utilized at less than the full 10Gbps capability, 10GBASE-T will be able to use IEEE P802.3az compliant Energy Efficient Ethernet to replace Idles with Low-power Idles, as shown in Figure 3. Because the large receiver signal processing load is not needed in 10GBASE-T PHYs during the low-power-idle period, savings in this mode can eventually be substantial (see
Figure 4). However, because achieving maximum power savings with 802.3az low-power-idle signaling will toggle major functions on and off, we expect it to stress many of the IC design processes, and be an area for much innovation over time. As a result, improvements in 802.3az PHY power savings and robustness are expected to be significant as later generations are released. It is important to note that EEE scales with link utilization and requires interoperable link partners while Dynamic Power Scaling runs at full 10G speeds, scales with reach, and is designed to work with non-Solarflare PHY devices.

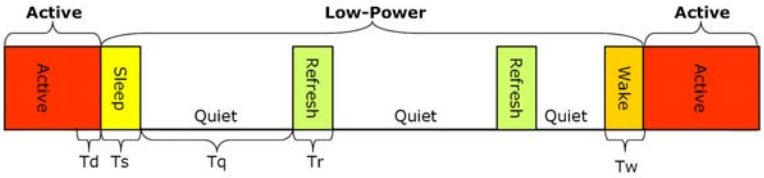

Figure 3: IEEE P802.3az Low Power Idle Concept

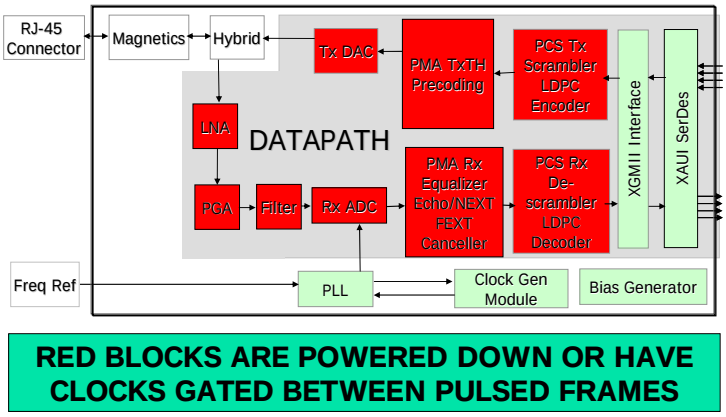

Figure 4: Signal Processing Savings Potential with 802.3az 10GBASE-T

Conclusion

10GBASE-T products provide data center managers and architects with unique capabilities to implement energy efficient data centers. Today's 10GBASE-T products, with their unique use of structured cabling enable cooling and power distribution efficiencies through flexibility in equipment location, leveraging the full power consumption of the server or switch. Second, Dynamic Power Scaling, unique to Solarflare 10GBASE-T lowers the average power used across the data center by using only the power each link length requires. Additionally, 10GBASE-T is unique among 10Gigabit technologies in leveraging Wake-on-LAN to save power by turning off entire servers when they are not needed. In the future, 10GBASE-T will also provide IEEE 802.3az Energy Efficient Ethernet to further improve energy efficiency by lowering consumption during periods of time when the full capacity of the 10Gbps link is not needed.

References

[1] U.S. Environmental Protection Agency, Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431, ENERGY STAR Program, August 2, 2007

[2] Hassen, Marcus; "Small Steps Lead To Big Data Center Efficiency Gains", Facilities Net: Data Center Master Improvement Program, May 2009, <http://www.facilitiesnet.com/datacenters/article/Small-Steps-Lead-to-Big-Data-Center-Efficiency-Gains--10827>